Curve Fitting Toolbox™
User's Guide

R2012b

MATLAB®

MathWorks®

How to Contact MathWorks

www.mathworks.com	Web
comp.soft-sys.matlab	Newsgroup
www.mathworks.com/contact_TS.html	Technical Support

suggest@mathworks.com	Product enhancement suggestions
bugs@mathworks.com	Bug reports
doc@mathworks.com	Documentation error reports
service@mathworks.com	Order status, license renewals, passcodes
info@mathworks.com	Sales, pricing, and general information

508-647-7000 (Phone)

508-647-7001 (Fax)

The MathWorks, Inc.
3 Apple Hill Drive
Natick, MA 01760-2098

For contact information about worldwide offices, see the MathWorks Web site.

Curve Fitting Toolbox™ User's Guide

Trademarks

MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.

Patents

MathWorks products are protected by one or more U.S. patents. Please see www.mathworks.com/patents for more information.

Contents

Getting Started

1

Interactive Fitting

2

Programmatic Curve and Surface Fitting

3

Linear and Nonlinear Regression

4

Custom Linear and Nonlinear Regression

5

Interpolation and Smoothing

6

Fit Postprocessing

7

Spline Fitting

About Splines

8

Simple Spline Examples

9

Types of Splines

10

Advanced Spline Examples

11

Function Reference

12

1

Getting Started

- "Product Description" on page 1-2
- "Curve Fitting Tools" on page 1-3
- "Curve Fitting" on page 1-4
- "Surface Fitting" on page 1-6
- "Spline Fitting" on page 1-8

Product Description

Fit curves and surfaces to data using regression, interpolation, and smoothing

Curve Fitting Toolbox™ provides graphical tools and command-line functions for fitting curves and surfaces to data. The toolbox lets you perform exploratory data analysis, preprocess and post-process data, compare candidate models, and remove outliers. You can conduct regression analysis using the library of linear and nonlinear models provided or specify your own custom equations. The library provides optimized solver parameters and starting conditions to improve the quality of your fits. The toolbox also supports nonparametric modeling techniques, such as splines, interpolation, and smoothing.

After creating a fit, you can apply a variety of post-processing methods for plotting, interpolation, and extrapolation; estimating confidence intervals; and calculating integrals and derivatives.

Key Features

- Graphical tools for curve and surface fitting

- Linear and nonlinear regression with custom equations

- Library of regression models with optimized starting points and solver parameters

- Interpolation methods, including B-splines, thin plate splines, and tensor-product splines

- Smoothing techniques, including smoothing splines, localized regression, Savitzky-Golay filters, and moving averages

- Preprocessing routines, including outlier removal and sectioning, scaling, and weighting data

- Post-processing routines, including interpolation, extrapolation, confidence intervals, integrals and derivatives

Curve Fitting Tools

Curve Fitting Toolbox software allows you to work in two different environments:

- An interactive environment, with Curve Fitting Tool and Spline Tool graphical user interfaces
- A programmatic environment that allows you to write object-oriented MATLAB® code using curve and surface fitting methods

To open Curve Fitting Tool or Spline Tool, enter one of the following:

- cftool
- splinetool

To list the Curve Fitting Toolbox functions for use in MATLAB programming, type

```
help curvefit
```

The code for any function can be opened in the MATLAB Editor by typing

```
edit function_name
```

Brief, command line help for any function is available by typing

```
help function_name
```

Complete documentation for any function is available by typing

```
doc function_name
```

You can change the way any toolbox function works by copying and renaming its file, examining your copy in the editor, and then modifying it.

You can also extend the toolbox by adding your own files, or by using your code in combination with functions from other toolboxes, such as Statistics Toolbox™ or Optimization Toolbox™ software.

Curve Fitting

Interactive Curve Fitting

To interactively fit a curve, follow the steps in this simple example:

1 Load some data at the MATLAB command line.

```
load hahn1
```

2 Open Curve Fitting Tool. Enter:

```
cftool
```

3 In Curve Fitting Tool, select **X Data** and **Y Data**.

Curve Fitting Tool creates a default interpolation fit to the data.

4 Choose a different model type using the fit category drop-down list, e.g., select **Polynomial**.

5 Try different fit options for your chosen model type.

6 Select **File > Generate Code**.

Curve Fitting Tool creates a file in the Editor containing MATLAB code to recreate all fits and plots in your interactive session.

For more information about fitting curves in Curve Fitting Tool, see "Interactive Curve and Surface Fitting" on page 2-2.

For details and examples of specific model types and fit analysis, see the following sections:

1 "Linear and Nonlinear Regression"

2 "Interpolation"

3 "Smoothing"

4 "Fit Postprocessing"

Programmatic Curve Fitting

To programmatically fit a curve, follow the steps in this simple example:

1 Load some data.

```
load hahn1
```

Create a fit using the `fit` function, specifying the variables and a model type (in this case `rat23` is the model type).

```
f = fit( temp, thermex, 'rat23' )
```

Plot your fit and the data.

```
plot( f, temp, thermex )
f( 600 )
```

See these sections:

1 "Curve and Surface Fitting" on page 3-2

2 "Curve and Surface Fitting Objects and Methods" on page 3-6

For details and examples of specific model types and fit analysis, see the following sections:

1 "Linear and Nonlinear Regression"

2 "Interpolation"

3 "Smoothing"

4 "Fit Postprocessing"

Surface Fitting

Interactive Surface Fitting

To interactively fit a surface, follow the steps in this simple example:

1 Load some data at the MATLAB command line.

```
load franke
```

2 Open Curve Fitting Tool. Enter:

```
cftool
```

3 In Curve Fitting Tool, select **X Data**, **Y Data** and **Z Data**.

Curve Fitting Tool creates a default interpolation fit to the data.

4 Choose a different model type using the fit category drop-down list, e.g., select **Polynomial**.

5 Try different fit options for your chosen model type.

6 Select **File > Generate Code**.

Curve Fitting Tool creates a file in the Editor containing MATLAB code to recreate all fits and plots in your interactive session.

For more information about fitting surfaces in Curve Fitting Tool, see "Interactive Curve and Surface Fitting" on page 2-2.

For details and examples of specific model types and fit analysis, see the following sections:

1 "Linear and Nonlinear Regression"

2 "Interpolation"

3 "Smoothing"

4 "Fit Postprocessing"

Programmatic Surface Fitting

To programmatically fit a surface, follow the steps in this simple example:

1 Load some data.

```
load franke
```

2 Create a fit using the `fit` function, specifying the variables and a model type (in this case `poly23` is the model type).

```
f = fit( [x, y], z, 'poly23' )
```

3 Plot your fit and the data.

```
plot(f, [x,y], z)
```

To programmatically fit surfaces, see the following topics:

1 "Curve and Surface Fitting" on page 3-2

2 "Curve and Surface Fitting Objects and Methods" on page 3-6

For details and examples of specific model types and fit analysis, see the following sections:

1 "Linear and Nonlinear Regression"

2 "Interpolation"

3 "Smoothing"

4 "Fit Postprocessing"

Spline Fitting

In this section...

About Splines in Curve Fitting Toolbox

You can work with splines in Curve Fitting Toolbox in several ways.

Using Curve Fitting Tool or the `fit` function you can:

- Fit cubic spline interpolants to curves or surfaces

- Fit smoothing splines and shape-preserving cubic spline interpolants to curves (but not surfaces)

The toolbox also contains specific splines functions to allow greater control over what you can create. For example, you can use the `csapi` function for cubic spline interpolation. Why would you use `csapi` instead of the `fit` function `'cubicinterp'` option? You might require greater flexibility to work with splines for the following reasons:

- You want to combine the results with other splines, e.g., by addition.

- You want vector-valued splines. You can use `csapi` with scalars, vectors, matrices, and ND-arrays. The `fit` function only allows scalar-valued splines.

- You want other types of splines such as ppform, B-form, tensor-product, rational, and stform thin-plate splines.

- You want to create splines without data.

- You want to specify breaks, optimize knot placement, and use specialized functions for spline manipulation such as differentiation and integration.

If you require specialized spline functions, see the following sections for interactive and programmatic spline fitting.

Interactive Spline Fitting

You can access all spline functions from the `splinetool` GUI.

See "Introducing Spline Fitting" on page 8-2.

Programmatic Spline Fitting

To programmatically fit splines, see "Construction" for descriptions of types of splines and numerous examples.

2

Interactive Fitting

Interactive Curve and Surface Fitting

In this section...
"Introducing the Curve Fitting Tool" on page 2-2
"Fitting a Curve" on page 2-2
"Fitting a Surface" on page 2-4
"Model Types for Curves and Surfaces" on page 2-6

Introducing the Curve Fitting Tool

You can fit curves and surfaces to data and view plots with the Curve Fitting Tool GUI. This tool lets you:

- Create, plot, and compare multiple fits.

- Use linear or nonlinear regression, interpolation, smoothing, and custom equations.

- View goodness-of-fit statistics, display confidence intervals and residuals, remove outliers and assess fits with validation data.

- Automatically generate code to fit and plot curves and surfaces, or export fits to the workspace for further analysis.

Fitting a Curve

1 Load some example data at the MATLAB command line:
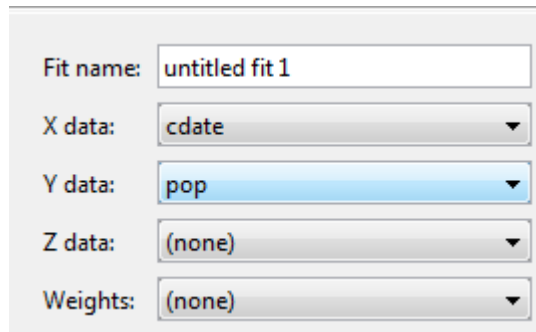
```
load census
```

2 Open the Curve Fitting Tool by entering:

```
cftool
```

Alternatively, select **Curve Fitting** on the **Apps** tab.

3 Select **X data** and **Y data**. For details, see "Data Selection" on page 2-8.

The Curve Fitting Tool creates a default polynomial fit to the data.

4 Try different fit options. For example, change the polynomial **Degree** to 3 to fit a cubic polynomial.

5 Select a different model type from the fit category list, e.g., **Smoothing Spline**. For information about models you can fit, see "Model Types for Curves and Surfaces" on page 2-6.
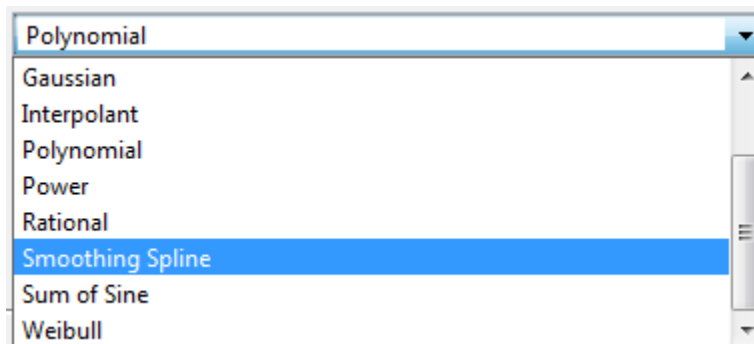
6 Select **File > Generate Code**.

The Curve Fitting Tool creates a file in the Editor containing MATLAB code to recreate all fits and plots in your interactive session.

Tip For a detailed workflow example, see "Curve Fitting to Census Data" on page 2-22.

To create multiple fits and compare them, see "Interactive Fit Comparison" on page 2-13.

Fitting a Surface
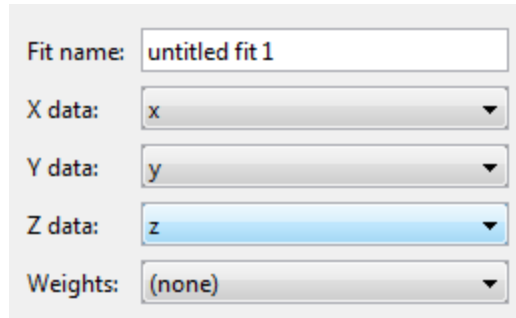
1 Load some example data at the MATLAB command line:

```
load franke
```

2 Open the Curve Fitting Tool:

```
cftool
```

3 Select **X data**, **Y data** and **Z data**. For more information, see "Data Selection" on page 2-8.
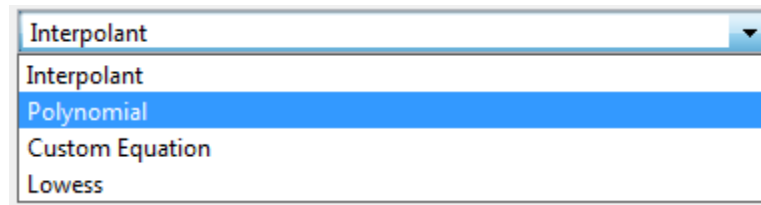
The Curve Fitting Tool creates a default interpolation fit to the data.

4 Select a different model type from the fit category list, e.g., **Polynomial**.

For information about models you can fit, see "Model Types for Curves and Surfaces" on page 2-6.

5 Try different fit options for your chosen model type.

6 Select **File > Generate Code**.

The Curve Fitting Tool creates a file in the Editor containing MATLAB code to recreate all fits and plots in your interactive session.

Tip For a detailed example, see "Surface Fitting to Franke Data" on page 2-36.

To create multiple fits and compare them, see "Interactive Fit Comparison" on page 2-13.

Model Types for Curves and Surfaces

Based on your selected data, the fit category list shows either curve or surface fit categories. The following table describes the options for curves and surfaces.

Fit Category	Curves	Surfaces
Regression Models		
Polynomial	Yes (up to degree 9)	Yes (up to degree 5)
Exponential	Yes	
Fourier	Yes	
Gaussian	Yes	
Power	Yes	
Rational	Yes	
Sum of Sine	Yes	
Weibull	Yes	
Interpolation		
Interpolant	Yes Methods: Nearest neighbor Linear Cubic Shape-preserving (PCHIP)	Yes Methods: Nearest neighbor Linear Cubic Biharmonic (v4)
Smoothing		
Smoothing Spline	Yes	
Lowess		Yes
Custom		
Custom Equation	Yes	Yes

For information about these fit types, see:

- "Linear and Nonlinear Regression"

- "Custom Models" on page 5-2
- "Interpolation"
- "Smoothing"

Data Selection

In this section...
"Opening the Curve Fitting Tool" on page 2-8
"Selecting Data to Fit" on page 2-8
"Selecting Compatible Size Surface Data" on page 2-10
"Troubleshooting Data Problems" on page 2-11

Opening the Curve Fitting Tool

Open the Curve Fitting Tool using one of these methods:

- Enter cftool at the MATLAB command line.

- Select from the MATLAB **Start** menu:

  Start > Toolboxes > Curve Fitting > Curve Fitting Tool (cftool)

Next, you are ready to select the data you want to fit.
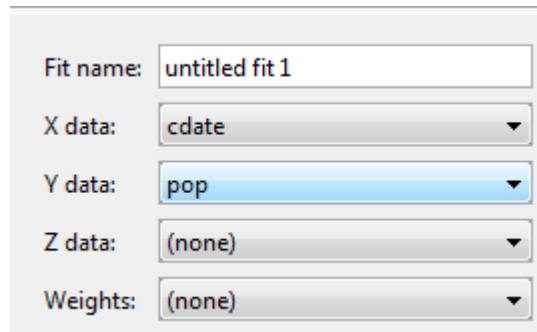
Selecting Data to Fit

To select data to fit, use the drop-down lists in the Curve Fitting Tool to select variables in your MATLAB workspace.

- To fit curves:

  - Select **X data** and **Y data**.

  - Select only **Y data** to plot Y against index (x=1:length(y)).

- To fit surfaces, select **X data**, **Y data** and **Z data**.

You can use the Curve Fitting Tool drop-down lists to select any numeric variables (with more than one element) in your MATLAB workspace.

Similarly, you can select any numeric data in your workspace to use as **Weights**.

For curves, X, Y, and Weights must be matrices with the same number of elements.

For surfaces, X, Y, and Z must be either:

- Matrices with the same number of elements
- Data in the form of a table

For surfaces, weights must have the same number of elements as Z.

For more information see "Selecting Compatible Size Surface Data" on page 2-10.

When you select variables, the Curve Fitting Tool immediately creates a curve or surface fit with the default settings. If you want to avoid time-consuming refitting for large data sets, you can turn off **Auto fit** by clearing the check box.

Note The Curve Fitting Tool uses a snapshot of the data you select. Subsequent workspace changes to the data have no effect on your fits. To update your fit data from the workspace, first change the variable selection, and then reselect the variable with the drop-down controls.

Selecting Compatible Size Surface Data

For surface data, you can select either "Matrices of the Same Size" on page 2-10 or "Table Data" on page 2-10.

Matrices of the Same Size

Curve Fitting Tool expects inputs to be the same size. If the sizes are different but the number of elements are the same, then the tool reshapes the inputs to create a fit and displays a warning in the **Results** pane. The warning indicates a possible problem with your selected data.

Table Data

Table data means that X and Y represent the row and column headers of a table (sometimes called *breakpoints*) and the values in the table are the values of the Z output.

Sizes are compatible if:

- **X** is a vector of length n.
- **Y** is a vector of length m.
- **Z** is a 2D matrix of size [m,n].

The following table shows an example of data in the form of a table with n = 4 and m = 3.

	x(1)	x(2)	x(3)	x(4)
y(1)	z(1,1)	z(1,2)	z(1,3)	z(1,4)
y(2)	z(2,1)	z(2,2)	z(2,3)	z(2,4)
y(3)	z(3,1)	z(3,2)	z(3,3)	z(3,4)

Like the surf function, the tool expects inputs where length(X) = n, length(Y) = m and size(Z) = [m,n]. If the size of Z is [n,m], the tool creates a fit but first transposes Z and warns about transforming your data. You see a warning in the **Results** pane like the following example:

```
Using X Input for rows and Y Input for columns
to match Z Output matrix.
```

For suitable example table data, run the following code:

```
x = linspace( 0, 1, 7 );
y = linspace( 0, 1, 9 ).';
z = bsxfun( @franke, x, y );
```

Weights

If you specify surface Weights, assign an input the same size as Z. If the sizes are different but the number of elements is the same, Curve Fitting Tool reshapes the weights and displays a warning.

Troubleshooting Data Problems

If there are problems with the data you select, you see messages in the **Results** pane. For example, the Curve Fitting Tool ignores Infs, NaNs, and imaginary components of complex numbers in the data, and you see messages in the **Results** pane in these cases.

If you see warnings about reshaping your data or incompatible sizes, read "Selecting Compatible Size Surface Data" on page 2-10 for information.

If you see the following warning: `Duplicate x-y data points detected: using average of the z values.`, this means that there are two or more data points where the input values (x, y) are the same or very close together. The default interpolant fit type needs to calculate a unique value at that point. You do not need do anything to fix the problem, this warning is just for your information. The Curve Fitting Tool automatically takes the average z value of any group of points with the same x-y values.

Other problems with your selected data can produce the following error:

`Error computing Delaunay triangulation. Please try again with different data.`

Some arrangements of data make it impossible for Curve Fitting Tool to compute a Delaunay triangulation. Three out of the four surface interpolation methods (linear, cubic, and nearest) require a Delaunay triangulation of the data. An example of data that can cause this error is a case where all the data

lies on a straight line in x-y. In this case, Curve Fitting Tool cannot fit a surface to the data. You need to provide more data in order to fit a surface.

Note Data selection is disabled if you are in debug mode. Exit debug mode to change data selections.

Interactive Fit Comparison

In this section...
"Refining Your Fit" on page 2-13
"Creating Multiple Fits" on page 2-13
"Duplicating a Fit" on page 2-14
"Deleting a Fit" on page 2-14
"Displaying Multiple Fits Simultaneously" on page 2-14
"Using the Statistics in the Table of Fits" on page 2-18

Refining Your Fit

After you create a single fit, you can refine your fit, using any of the following optional steps:

- Change fit type and settings. Select GUI settings to use the Curve Fitting Tool built-in fit types or create custom equations. For fit settings for each model type, see "Linear and Nonlinear Regression", "Interpolation", and "Smoothing".

- Exclude data by removing outliers in the Curve Fitting Tool. See "Removing Outliers" on page 7-10.

- Select weights. See "Data Selection" on page 2-8.

- Select validation data. See "Selecting Validation Data" on page 7-11

- Create multiple fits and you can compare different fit types and settings side by side in the Curve Fitting Tool. See "Creating Multiple Fits" on page 2-13.

Creating Multiple Fits

After you create a single fit, it can be useful to create multiple fits to compare. When you create multiple fits you can compare different fit types and settings side-by-side in the Curve Fitting Tool.

After creating a fit, you can add an additional fit using any of these methods:

- Click the New Fit button next to your fit figure tabs in the Document Bar.
- Right-click the Document Bar and select **New Fit**.
- Select **Fit > New Fit**.

Each additional fit appears as a new tab in the Curve Fitting Tool and a new row in the **Table of Fits**. See "Interactive Fit Comparison" on page 2-13 for information about displaying and analyzing multiple fits.

Optionally, after you create an additional fit, you can copy your data selections from a previous fit by selecting **Fit > Use Data From >** *Other Fit Name*. This copies your selections for x, y, and z from the previous fit, and any selected validation data. No fit options are changed.

Use sessions to save and reload your fits. See "Sessions" on page 2-20.

Duplicating a Fit

To create a copy of the current fit tab, select **Fit > Duplicate "***Current Fit Name***"**. You also can right-click a fit in the **Table of Fits** and select **Duplicate**

Each additional fit appears as a new tab in the Curve Fitting Tool.

Deleting a Fit

Delete a fit from your session using one of these methods:

- Select the fit tab display and select **Fit > Delete** *Current Fit Name*.
- Select the fit in the **Table of Fits** and press **Delete**.
- Right-click the fit in the table and select **Delete** *Current Fit Name*.

Displaying Multiple Fits Simultaneously

When you have created multiple fits you can compare different fit types and settings side by side in the Curve Fitting Tool. You can view plots simultaneously and you can examine the goodness-of-fit statistics to compare your fits. This section describes how to compare multiple fits.

To compare plots and see multiple fits simultaneously, use the layout controls at the top right of the Curve Fitting Tool. Alternatively, you can click **Window** on the menu bar to select the number and position of tiles you want to display. A *fit figure* displays the fit settings, results pane and plots for a single fit. The following example shows two fit figures displayed side by side. You can see multiple fits in the session listed in the **Table of Fits**.
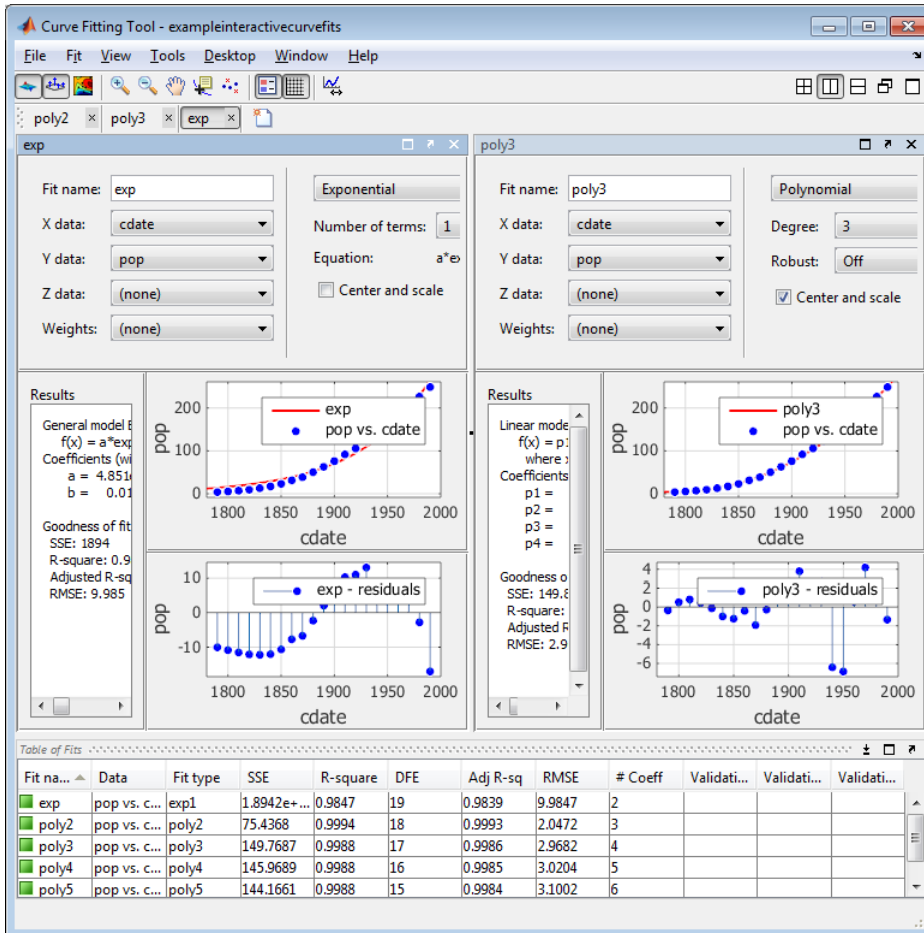
You can close fit figures displays (with the Close button, **Fit** menu, or context menu), but they remain in your session. The **Table of Fits** displays all your

fits (open and closed). Double-click a fit in the **Table of Fits** to open (or focus if already open) the fit figure. To remove a fit, see "Deleting a Fit" on page 2-14

Tip If you want more space to view and compare plots, as shown next, use the **View** menu to hide or show the **Fit Settings**, **Fit Results**, or **Table of Fits** panes.
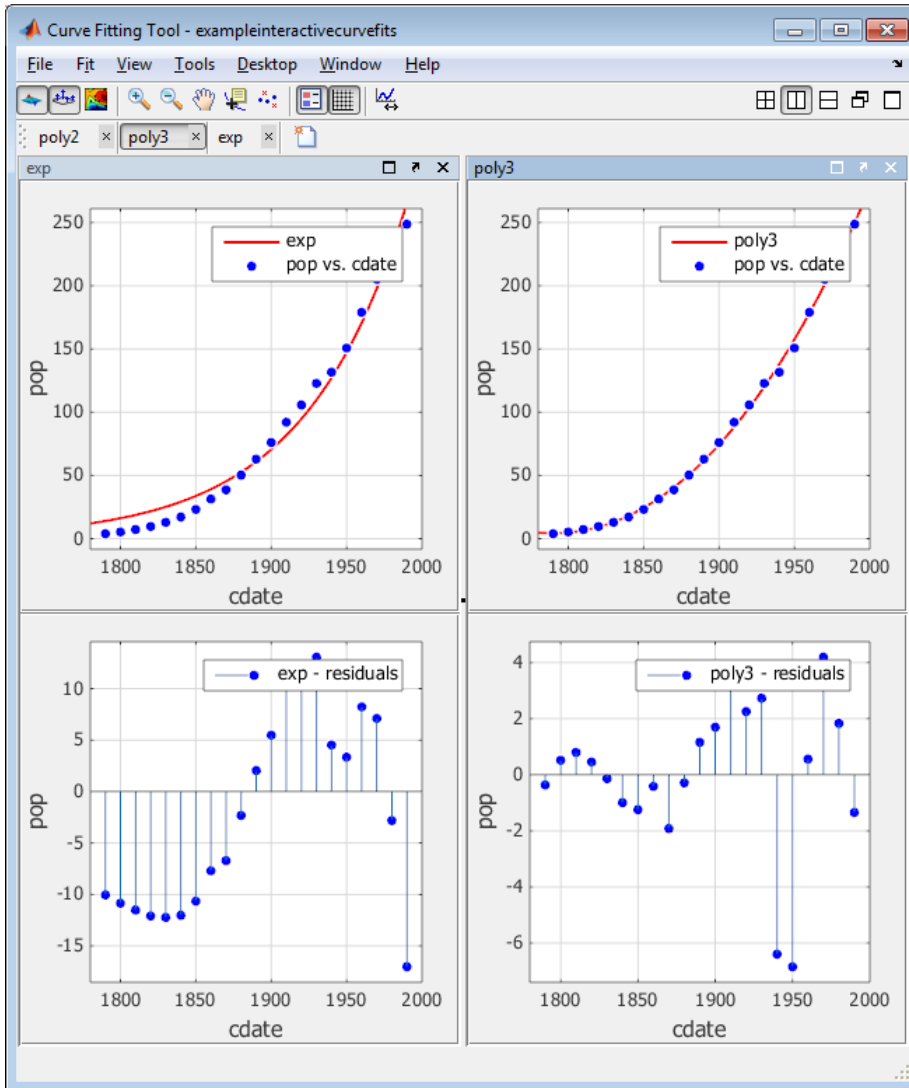
You can dock and undock individual fits and navigate between them using the standard MATLAB **Desktop** and **Window** menus in the Curve Fitting Tool. For more information, see "Optimize Desktop Layout for Limited Screen Space" in the MATLAB Desktop Tools and Development Environment documentation.

Using the Statistics in the Table of Fits

The **Table of Fits** list pane shows all fits in the current session.

Fit name	Data	Fit type	SSE ▲	R-square	DFE	Adj R-sq	RMSE	# Coeff
🟩 poly6	pop vs. cd...	poly6	106.9276	0.9991	14	0.9988	2.7636	7
🟩 poly5	pop vs. cd...	poly5	144.1661	0.9988	15	0.9984	3.1002	6
🟩 poly4	pop vs. cd...	poly4	145.9689	0.9988	16	0.9985	3.0204	5
🟩 poly3	pop vs. cd...	poly3	149.7687	0.9988	17	0.9986	2.9682	4
🟩 poly2	pop vs. cd...	poly2	159.0293	0.9987	18	0.9986	2.9724	3
🟩 exp	pop vs. cd...	exp1	1.8942e+03	0.9847	19	0.9839	9.9847	2

After using graphical methods to evaluate the goodness of fit, you can examine the goodness-of-fit statistics shown in the table to compare your fits. The goodness-of-fit statistics help you determine how well the model fits the data. Click the table column headers to sort by statistics, name, fit type, and so on.

The following guidelines help you use the statistics to determine the best fit:

- **SSE** is the sum of squares due to error of the fit. A value closer to zero indicates a fit that is more useful for prediction.

- **R-square** is the square of the correlation between the response values and the predicted response values. A value closer to 1 indicates that a greater proportion of variance is accounted for by the model.

- **DFE** is the degree of freedom in the error.

- **Adj R-sq** is the degrees of freedom adjusted R-square. A value closer to 1 indicates a better fit.

- **RMSE** is the root mean squared error or standard error. A value closer to 0 indicates a fit that is more useful for prediction.

- **# Coeff** is the number of coefficients in the model. When you have several fits with similar goodness-of-fit statistics, look for the smallest number of coefficients to help decide which fit is best. You must trade off the number of coefficients against the goodness of fit indicated by the statistics to avoid overfitting.

For a more detailed explanation of the Curve Fitting Toolbox statistics, see "Goodness-of-Fit Statistics" on page 7-17.

To compare the statistics for different fits and decide which fit is the best tradeoff between over- and under-fitting, use a similar process to that described in "Curve Fitting to Census Data" on page 2-22.

Sessions

In this section...
"Overview" on page 2-20
"Saving Sessions" on page 2-20
"Reloading Sessions" on page 2-20
"Removing Sessions" on page 2-20

Overview

You can save and reload sessions for easy access to multiple fits. The session file contains all the fits and variables in your session and remembers your layout.

Saving Sessions

To save your session, first select **File > Save Session** to open your file browser. Next, select a name and location for your session file (with file extension `.sfit`).

After you save your session once, you can use **File > Save *MySessionName*** to overwrite that session for subsequent saves.

To save the current session under a different name, select **File > Save Session As** .

Reloading Sessions

Use **File > Load Session** to open a file browser where you can select a saved curve fitting session file to load.

Removing Sessions

Use **File > Clear Session** to remove all fits from the current Curve Fitting Tool session.

Generating MATLAB Code and Exporting Fits

Interactive Code Generation and Programmatic Fitting

Curve Fitting Tool makes it easy to plot and analyze fits at the command line. You can export individual fits to the workspace for further analysis, or you can generate MATLAB code to recreate all fits and plots in your session. By generating code you can use your interactive curve fitting session to quickly assemble code for curve and surface fits and plots into useful programs.

1 Select **File > Generate Code**.

The Curve Fitting Tool generates code from your session and displays the file in the MATLAB Editor. The file includes all fits and plots in your current session. The file captures the following information:

- Names of fits and their variables
- Fit settings and options
- Plots
- Curve or surface fitting objects and methods used to create the fits:
  - A cell-array of `cfit or sfit` objects representing the fits
  - A structure array with goodness-of fit information.

2 Save the file.

For more information on working with your generated code, exporting fits to the workspace, and recreating your fits and plots at the command line, see:

- "Generating Code from the Curve Fitting Tool" on page 7-12
- "Exporting a Fit to the Workspace" on page 7-13

Curve Fitting to Census Data

Interactive Curve Fitting Workflow

The next topics fit some census data using polynomial equations up to the sixth degree, and a single-term exponential equation. The steps demonstrate how to:

- Load data and explore various fits using different library models.

- Search for the best fit by:
    - Comparing graphical fit results
    - Comparing numerical fit results including the fitted coefficients and goodness-of-fit statistics

- Export your best fit results to the MATLAB workspace to analyze the model at the command line.

- Save the session and generate MATLAB code for all fits and plots.

Loading Data and Creating Fits

You must load the data variables into the MATLAB workspace before you can fit data using the Curve Fitting Tool. For this example, the data is stored in the MATLAB file census.mat.

1 Load the data:

```
load census
```

The workspace contains two new variables:

- `cdate` is a column vector containing the years 1790 to 1990 in 10-year increments.

- `pop` is a column vector with the U.S. population figures that correspond to the years in `cdate`.

2 Open the Curve Fitting Tool:

`cftool`

3 Select the variable names `cdate` and `pop` from the **X data** and **Y data** lists.

The Curve Fitting Tool creates and plots a default fit to X input (or predictor data) and Y output (or response data). The default fit is a linear polynomial fit type. Observe the fit settings display `Polynomial`, of **Degree** 1.

4 Change the fit to a second degree polynomial by selecting 2 from the **Degree** list.

The Curve Fitting Tool plots the new fit. The Curve Fitting Tool calculates a new fit when you change fit settings because **Auto fit** is selected by default. If refitting is time consuming, e.g., for large data sets, you can turn off **Auto fit** by clearing the check box.

The Curve Fitting Tool displays results of fitting the census data with a quadratic polynomial in the **Results** pane, where you can view the library model, fitted coefficients, and goodness-of-fit statistics.

5 Change the **Fit name** to `poly2`.

6 Display the residuals by selecting **View > Residuals Plot**.

The residuals indicate that a better fit might be possible. Therefore, continue exploring various fits to the census data set.
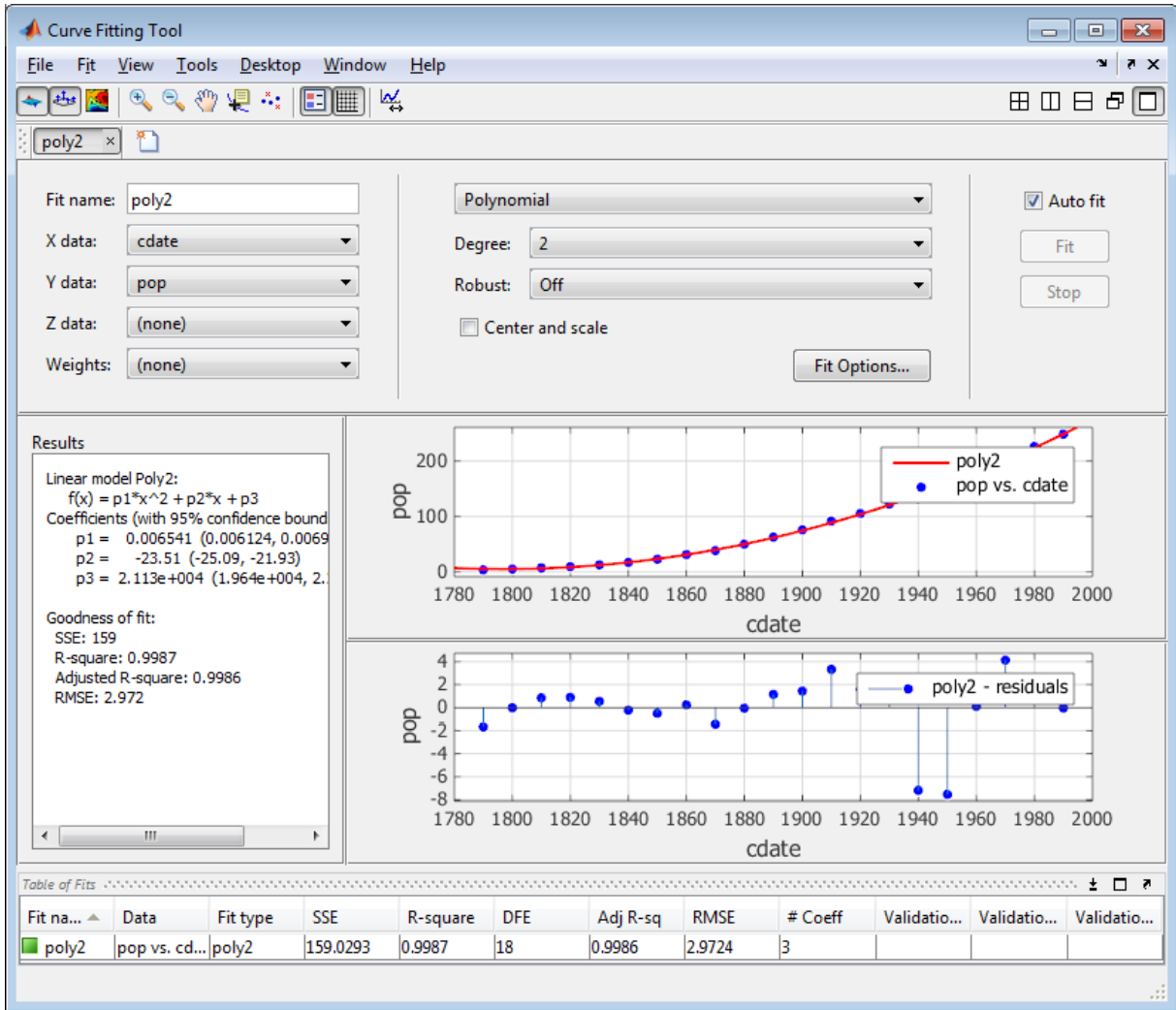
7 Add new fits to try the other library equations.

> **a** Right-click the fit in the **Table of Fits** and select **Duplicate "*poly2*"** (or use the **Fit** menu).

> **Tip** For fits of a given type (for example, polynomials), use **Duplicate** "*fitname*" instead of a new fit because copying a fit requires fewer steps. The duplicated fit contains the same data selections and fit settings.

b Change the polynomial **Degree** to 3 and rename the fit `poly3`.

c When you fit higher degree polynomials, the **Results** pane displays this warning:

```
Equation is badly conditioned. Remove repeated data points
or try centering and scaling.
```

Normalize the data by selecting the **Center and scale** check box.

d Repeat steps a and b to add polynomial fits up to the sixth degree, and then add an exponential fit.

e For each new fit, look at the **Results** pane information, and the residuals plot in the Curve Fitting Tool.

The residuals from a good fit should look random with no apparent pattern. A pattern, such as a tendency for consecutive residuals to have the same sign, can be an indication that a better model exists.

About Scaling

The warning about scaling arises because the fitting procedure uses the `cdate` values as the basis for a matrix with very large values. The spread of the `cdate` values results in a scaling problem. To address this problem, you can normalize the `cdate` data. Normalization scales the predictor data to improve the accuracy of the subsequent numeric computations. A way to normalize `cdate` is to center it at zero mean and scale it to unit standard deviation. The equivalent code is:

```
(cdate - mean(cdate))./std(cdate)
```

Note Because the predictor data changes after normalizing, the values of the fitted coefficients also change when compared to the original data. However, the functional form of the data and the resulting goodness-of-fit statistics do not change. Additionally, the data is displayed in the Curve Fitting Tool plots using the original scale.

Determining the Best Fit

To determine the best fit, you should examine both the graphical and numerical fit results.

Examine the Graphical Fit Results

1 Determine the best fit by examining the graphs of the fits and residuals. To view plots for each fit in turn, double-click the fit in the Table of Fits. The graphical fit results indicate that:

- The fits and residuals for the polynomial equations are all similar, making it difficult to choose the best one.

- The fit and residuals for the single-term exponential equation indicate it is a poor fit overall. Therefore, it is a poor choice and you can remove the exponential fit from the candidates for best fit.

2 Examine the behavior of the fits up to the year 2050. The goal of fitting the census data is to extrapolate the best fit to predict future population values.

   a Double-click the sixth-degree polynomial fit in the Table of Fits to view the plots for this fit.

   b Change the axes limits of the plots by selecting **Tools > Axes Limits**.

   c Alter the **X (cdate) Maximum** to 2050, and increase the **Main Y (pop) Maximum** to 400, and press **Enter**.

   d Examine the fit plot. The behavior of the sixth-degree polynomial fit beyond the data range makes it a poor choice for extrapolation and you can reject this fit.
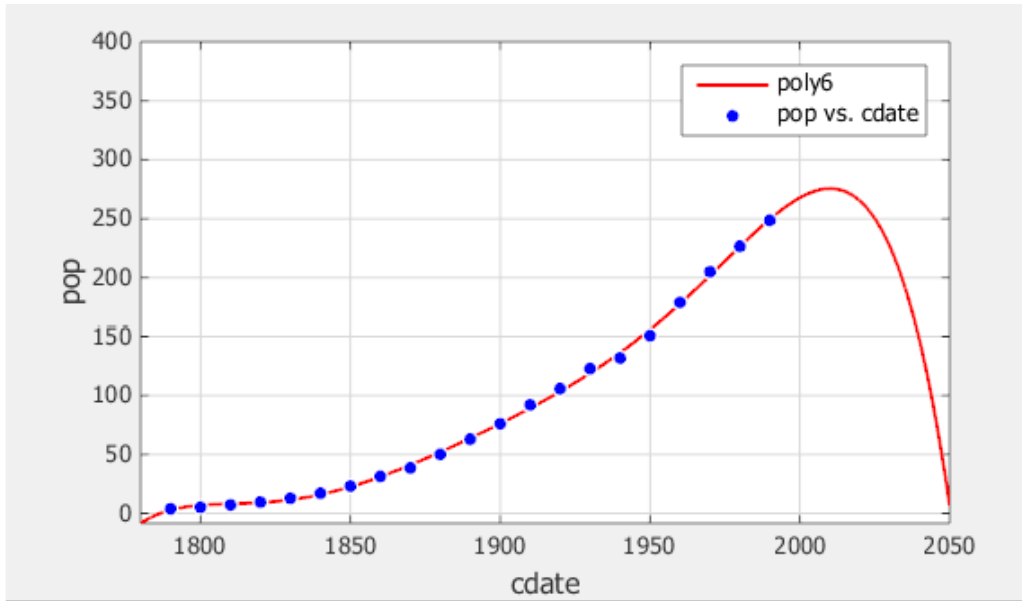
Evaluate the Numerical Fit Results

When you can no longer eliminate fits by examining them graphically, you should examine the numerical fit results. The Curve Fitting Tool displays two types of numerical fit results:

- Goodness-of-fit statistics

- Confidence bounds on the fitted coefficients

The goodness-of-fit statistics help you determine how well the curve fits the data. The confidence bounds on the coefficients determine their accuracy.

Examine the numerical fit results:

1 For each fit, view the goodness-of-fit statistics in the **Results** pane.

Results

Linear model Poly3:
    f(x) = p1*x^3 + p2*x^2 + p3*x + p4
        where x is normalized by mean 1890 and std 62.05
Coefficients (with 95% confidence bounds):
    p1 =   0.921 (-0.9743, 2.816)
    p2 =   25.18 (23.57, 26.79)
    p3 =   73.86 (70.33, 77.39)
    p4 =   61.74 (59.69, 63.8)

Goodness of fit:
  SSE: 149.8
  R-square: 0.9988
  Adjusted R-square: 0.9986
  RMSE: 2.968

2 Compare all fits simultaneously in the **Table of Fits**. Click the column headings to sort by statistics results.

Table of Fits

Fit name	Data	Fit type	SSE ▲	R-square	DFE	Adj R-sq	RMSE	# Coeff
poly6	pop vs. cd...	poly6	106.9276	0.9991	14	0.9988	2.7636	7
poly5	pop vs. cd...	poly5	144.1661	0.9988	15	0.9984	3.1002	6
poly4	pop vs. cd...	poly4	145.9689	0.9988	16	0.9985	3.0204	5
poly3	pop vs. cd...	poly3	149.7687	0.9988	17	0.9986	2.9682	4
poly2	pop vs. cd...	poly2	159.0293	0.9987	18	0.9986	2.9724	3
exp	pop vs. cd...	exp1	1.8942e+03	0.9847	19	0.9839	9.9847	2

3 Examine the sum of squares due to error (SSE) and the adjusted R-square statistics to help determine the best fit. The SSE statistic is the least-squares error of the fit, with a value closer to zero indicating a better fit. The adjusted R-square statistic is generally the best indicator of the fit quality when you add additional coefficients to your model.

The largest SSE for `exp1` indicates it is a poor fit, which you already determined by examining the fit and residuals. The lowest SSE value is associated with `poly6`. However, the behavior of this fit beyond the data

range makes it a poor choice for extrapolation, so you already rejected this fit by examining the plots with new axis limits.

The next best SSE value is associated with the fifth-degree polynomial fit, `poly5`, suggesting it might be the best fit. However, the SSE and adjusted *R*-square values for the remaining polynomial fits are all very close to each other. Which one should you choose?

4 Resolve the best fit issue by examining the confidence bounds for the remaining fits in the Results pane. Double-click a fit in the **Table of Fits** to open (or focus if already open) the fit figure and view the Results pane. A *fit figure* displays the fit settings, results pane and plots for a single fit.

Display the fifth-degree polynomial and the `poly2` fit figures side by side. Examining results side by side can help you assess fits.

a To show two fit figures simultaneously, use the layout controls at the top right of the Curve Fitting Tool or select **Window > Left/Right Tile** or **Top/Bottom Tile**.

b To change the displayed fits, click to select a fit figure and then double-click the fit to display in the **Table of Fits**.

c Compare the coefficients and bounds (`p1`, `p2`, and so on) in the Results pane for both fits, `poly5` and `poly2`. The toolbox calculates 95% confidence bounds on coefficients. The confidence bounds on the coefficients determine their accuracy. Check the equations in the Results pane (`f(x)=p1*x+p2*x...`) to see the model terms for each coefficient. Note that `p2` refers to the `p2*x` term in `Poly2` and the `p2*x^4` term in `Poly5`. Do not compare normalized coefficients directly with non-normalized coefficients.

> **Tip** Use the **View** menu to hide the **Fit Settings** or **Table of Fits** if you want more space to view and compare plots and results, as shown next. You can also hide the **Results** pane to show only plots.
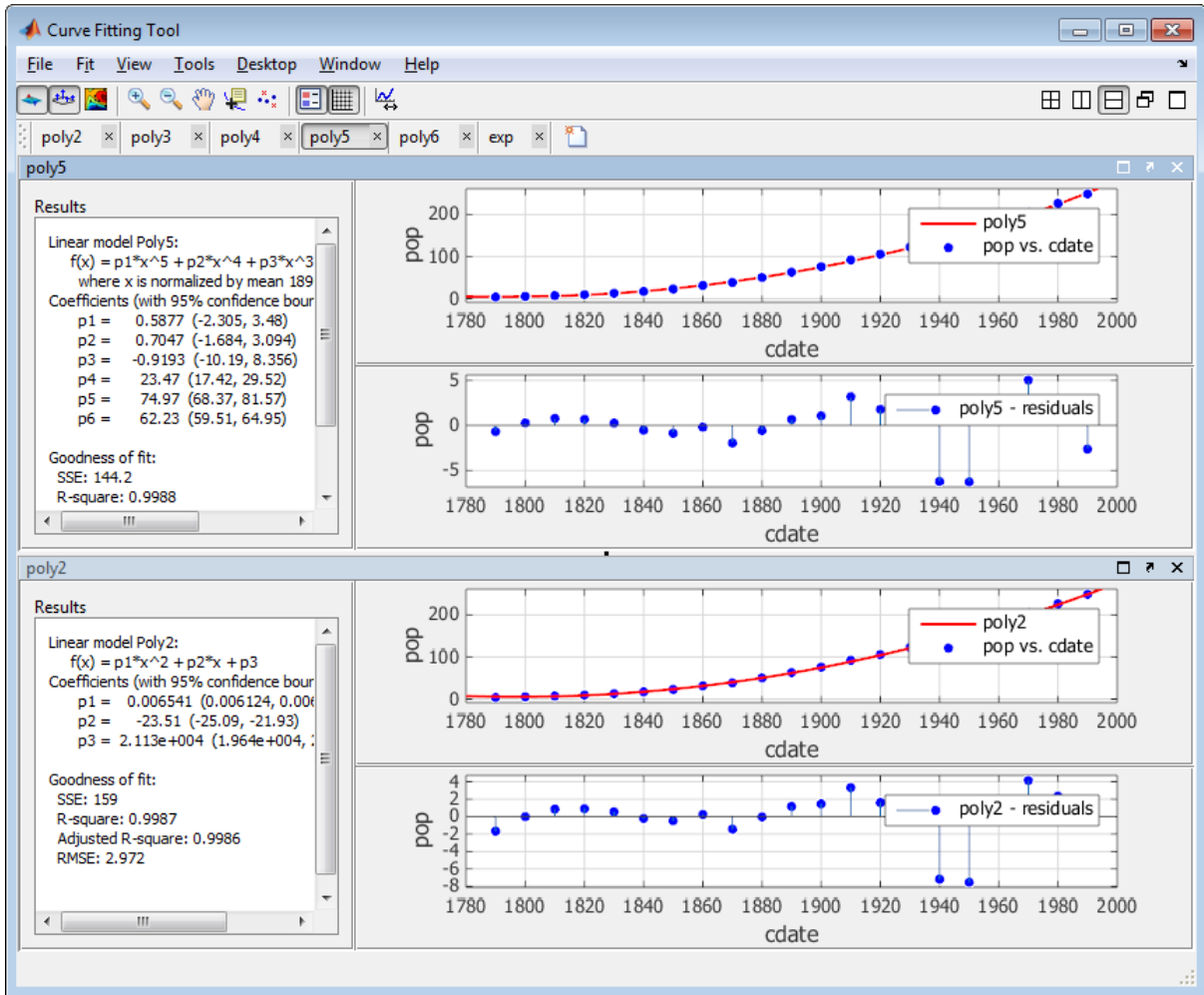
The bounds cross zero on the p1, p2, and p3 coefficients for the fifth-degree polynomial. This means you cannot be sure that these coefficients differ from zero. If the higher order model terms may have coefficients of zero, they are not helping with the fit, which suggests that this model overfits the census data.

```
Results

Linear model Poly5:
    f(x) = p1*x^5 + p2*x^4 + p3*x^3 + p4*x^2 + p5*x + p6
    where x is normalized by mean 1890 and std 62.05
Coefficients (with 95% confidence bounds):
    p1 =    0.5877  (-2.305, 3.48)
    p2 =    0.7047  (-1.684, 3.094)
    p3 =   -0.9193  (-10.19, 8.356)
    p4 =    23.47  (17.42, 29.52)
    p5 =    74.97  (68.37, 81.57)
    p6 =    62.23  (59.51, 64.95)

Goodness of fit:
  SSE: 144.2
  R-square: 0.9988
  Adjusted R-square: 0.9984
  RMSE: 3.1
```

However, the small confidence bounds do not cross zero on `p1`, `p2`, and `p3` for the quadratic fit, `poly2` indicate that the fitted coefficients are known fairly accurately.

```
Results

Linear model Poly2:
    f(x) = p1*x^2 + p2*x + p3
Coefficients (with 95% confidence bounds):
    p1 =  0.006541  (0.006124, 0.006958)
    p2 =    -23.51  (-25.09, -21.93)
    p3 = 2.113e+004  (1.964e+004, 2.262e+004)

Goodness of fit:
  SSE: 159
  R-square: 0.9987
  Adjusted R-square: 0.9986
  RMSE: 2.972
```

Therefore, after examining both the graphical and numerical fit results, you should select `poly2` as the best fit to extrapolate the census data.

Note The fitted coefficients associated with the constant, linear, and quadratic terms are nearly identical for each normalized polynomial equation. However, as the polynomial degree increases, the coefficient bounds associated with the higher degree terms cross zero, which suggests overfitting.

Analyzing Your Best Fit in the Workspace

You can use **Save to Workspace** to export the selected fit and the associated fit results to the MATLAB workspace. The fit is saved as a MATLAB object and the associated fit results are saved as structures.

1 Right-click the poly2 fit in the **Table of Fits** and select **Save "poly2" to Workspace** (or use the **Fit** menu).
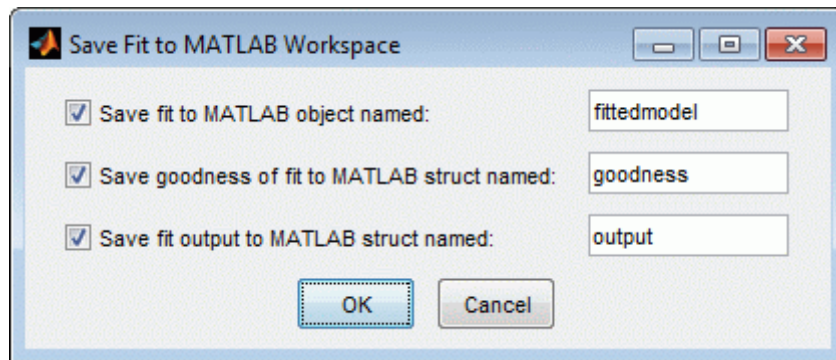
2 Click **OK** to save with the default names.

The fittedmodel is saved as a Curve Fitting Toolbox cfit object.

```
>> whos fittedmodel
  Name              Size            Bytes  Class

  fittedmodel       1x1               822  cfit
```

Examine the fittedmodel cfit object to display the model, the fitted coefficients, and the confidence bounds for the fitted coefficients:

```
fittedmodel
```

```
fittedmodel =
     Linear model Poly2:
       fittedmodel(x) = p1*x^2 + p2*x + p3
     Coefficients (with 95% confidence bounds):
       p1 =    0.006541  (0.006124, 0.006958)
       p2 =       -23.51  (-25.09, -21.93)
       p3 =  2.113e+004  (1.964e+004, 2.262e+004)
```

Examine the `goodness` structure to display goodness-of-fit results:

```
goodness
```

```
goodness =
           sse: 159.0293
       rsquare: 0.9987
           dfe: 18
    adjrsquare: 0.9986
          rmse: 2.9724
```

Examine the `output` structure to display additional information associated with the fit, such as the residuals:

```
output
```

```
output =
       numobs: 21
     numparam: 3
    residuals: [21x1 double]
     Jacobian: [21x3 double]
     exitflag: 1
    algorithm: 'QR factorization and solve'
   iterations: 1
```

You can evaluate (interpolate or extrapolate), differentiate, or integrate a fit over a specified data range with various postprocessing functions.

For example, to evaluate the `fittedmodel` at a vector of values to extrapolate to the year 2050, enter:

```
y = fittedmodel(2000:10:2050)
```

```
y =

   274.6221
   301.8240
   330.3341
   360.1524
   391.2790
   423.7137
```

Plot the fit to the census data and the extrapolated fit values:

```
plot(fittedmodel, cdate, pop)
hold on
plot(fittedmodel, 2000:10:2050, y)
hold off
```

For more examples and instructions for interactive and command-line fit analysis, and a list of all postprocessing functions, see "Fit Postprocessing".

For an example reproducing this interactive census data analysis using the command line, see "Custom Nonlinear Census Analysis" on page 5-6.

Saving Your Work

The toolbox provides several options for saving your work. You can save one or more fits and the associated fit results as variables to the MATLAB workspace. You can then use this saved information for documentation purposes, or to extend your data exploration and analysis. In addition to saving your work to MATLAB workspace variables, you can:

- Save the current curve fitting session by selecting **File > Save Session**. The session file contains all the fits and variables in your session and remembers your layout. See "Sessions" on page 2-20.

- Generate MATLAB code to recreate all fits and plots in your session by selecting **File > Generate Code**. The Curve Fitting Tool generates code from your session and displays the file in the MATLAB Editor.

  You can recreate your fits and plots by calling the file at the command line with your original data as input arguments. You can also call the file with new data, and automate the process of fitting multiple data sets.

For more information, see "Generating Code from the Curve Fitting Tool"
on page 7-12.

Surface Fitting to Franke Data

The Curve Fitting Tool provides some example data generated from Franke's bivariate test function. This data is suitable for trying various fit settings in Curve Fitting Tool.

To load the example data and create, compare, and export surface fits, follow these steps:

1 To load example data to use in the Curve Fitting Tool, enter `load franke` at the MATLAB command line. The variables x, y, and z appear in your workspace.

The example data is generated from Franke's bivariate test function, with added noise and scaling, to create suitable data for trying various fit settings in Curve Fitting Tool. For details on the Franke function, see the following paper:

Franke, R., Scattered Data Interpolation: Tests of Some Methods, Mathematics of Computation 38 (1982), pp. 181–200.

2 To divide the data into fitting and validation data, enter the following syntax:

```
xv = x(200:293);
yv = y(200:293);
zv = z(200:293);
x = x(1:199);
y = y(1:199);
z = z(1:199);
```

3 To fit a surface using this example data:

a Open Curve Fitting Tool. Enter `cftool`, or select **Curve Fitting** on the **Apps** tab.

b Select the variables x, y, and z interactively in the Curve Fitting Tool.
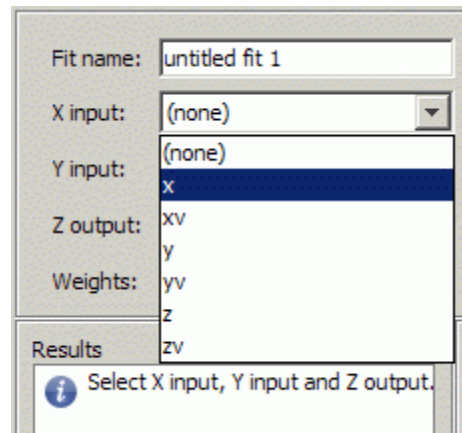
Alternatively, you can specify the variables when you enter `cftool(x,y,z)` to open Curve Fitting Tool (if necessary) and create a default fit.

The Curve Fitting Tool plots the data points as you select variables. When you select x, y, and z, the tool automatically creates a default surface fit. The default fit is an interpolating surface that passes through the data points.
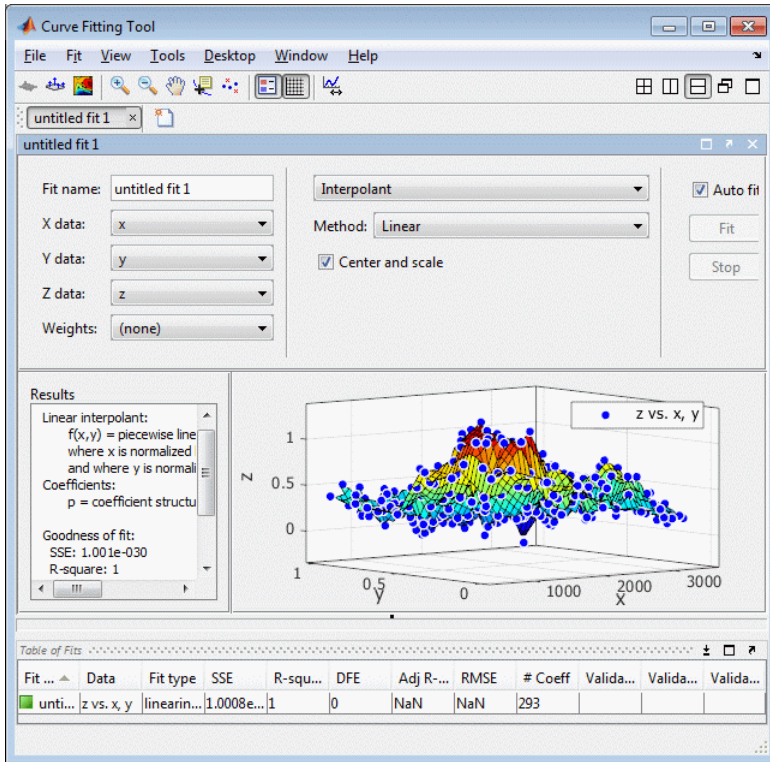
4 Try a Lowess fit type. Select the `Lowess` fit type from the drop-down list in the Curve Fitting Tool.
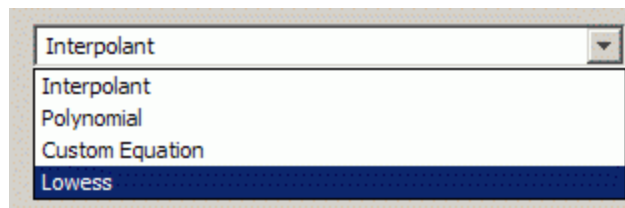
The Curve Fitting Tool creates a local smoothing regression fit.

5 Try altering the fit settings. Enter 10 in the **Span** edit box.

By reducing the span from the default to 10% of the total number of data points you produce a surface that follows the data more closely. The span defines the neighboring data points the toolbox uses to determine each smoothed value.
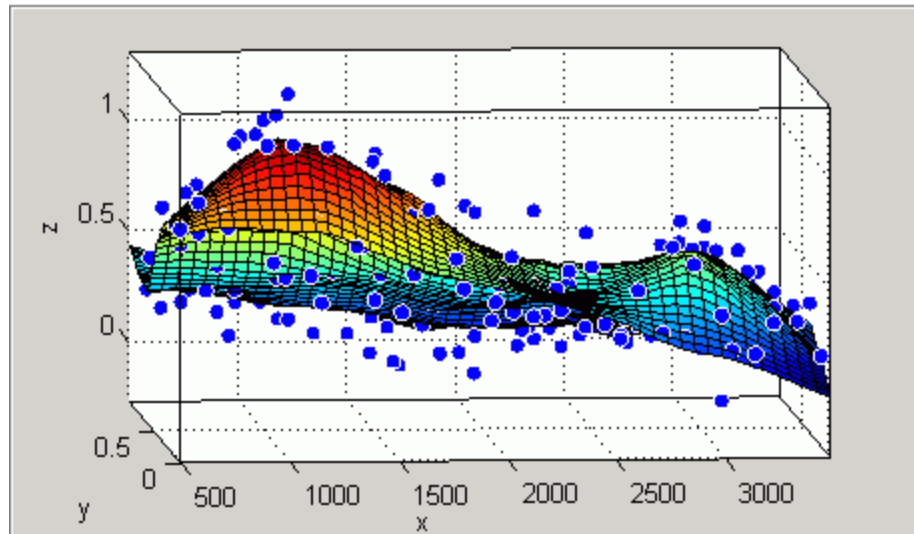
6 Edit the **Fit name** to Smoothing regression.

7 If you divided your data into fitting and validation data in step 2, select this validation data. Use the validation data to help you check that your surface is a good model, by comparing it against some other data not used for fitting.

    a Select **Fit > Specify Validation Data**. The Specify Validation Data dialog box opens.

    b Select the validation variables in the drop-down lists for **X input**, **Y input**, and **Z output**: xv, yv, and zv.

Review your selected validation data in the plots and the validation statistics (SSE and RMSE) in the **Results** pane and the **Table of Fits**.
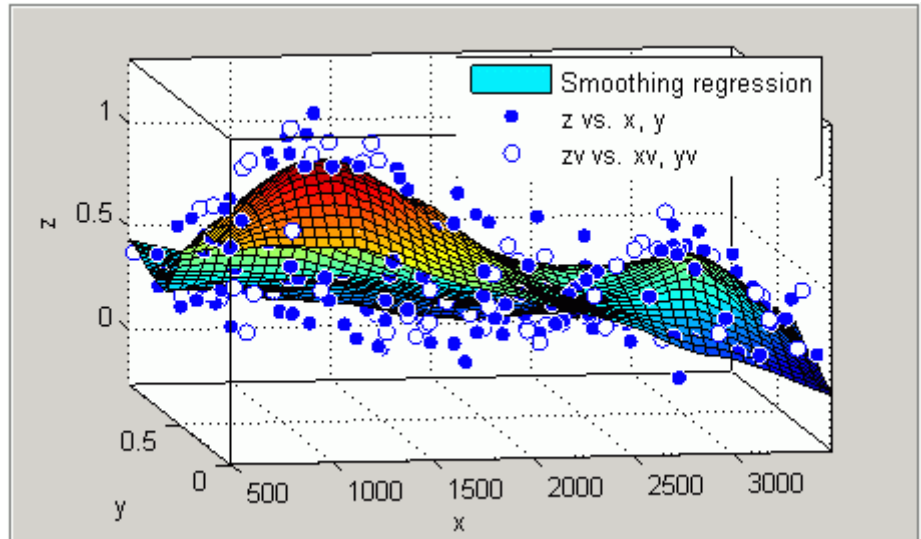
8 Create another fit to compare by making a copy of the current surface fit. Either select **Fit > Duplicate "Smoothing regression"**, or right-click the fit in the **Table of Fits**, and select **Duplicate**

The tool creates a new fit figure with the same fit settings, data, and validation data. It also adds a new row to the table of fits at the bottom.

9 Change the fit type to Polynomial and edit the fit name to Polynomial.
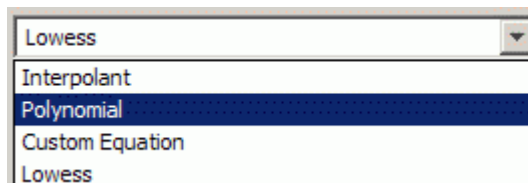
10 Change the **Degrees** of **x** and **y** to 3, to fit a cubic polynomial in both dimensions.

11 Look at the scales on the x and y axes, and read the warning message in the **Results** pane:

```
Equation is badly conditioned. Remove repeated data points
```

```
                        or try centering and scaling.
```

Select the **Center and scale** check box to normalize and correct for the large difference in scales in x and y.
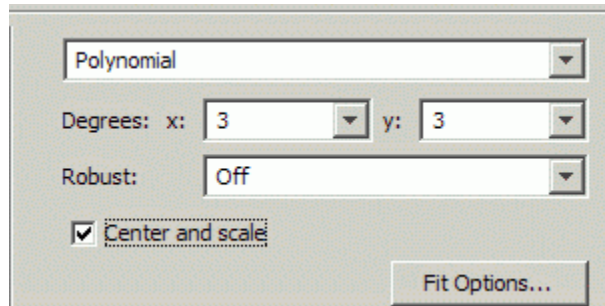
Normalizing the surface fit removes the warning message from the **Results** pane.

12 Look at the **Results** pane. You can view (and copy if desired):

- The model equation

- The values of the estimated coefficients

- The goodness-of-fit statistics

- The goodness of validation statistics

```
Linear model Poly33:
     f(x,y) = p00 + p10*x + p01*y + p20*x^2 + p11*x*y...
                    + p02*y^2 + p30*x^3 + p21*x^2*y
                    + p12*x*y^2 + p03*y^3
       where x is normalized by mean 1977 and std 866.5
       and where y is normalized by mean 0.4932 and std 0.29
Coefficients (with 95% confidence bounds):
       p00 =       0.4359  (0.3974, 0.4743)
       p10 =      -0.1375  (-0.194, -0.08104)
       p01 =      -0.4274  (-0.4843, -0.3706)
       p20 =       0.0161  (-0.007035, 0.03923)
       p11 =      0.07158  (0.05091, 0.09225)
       p02 =     -0.03668  (-0.06005, -0.01332)
       p30 =      0.02081  (-0.005475, 0.04709)
```

```
            p21 =       0.02432  (0.0012, 0.04745)
            p12 =      -0.03949  (-0.06287, -0.01611)
            p03 =        0.1185  (0.09164, 0.1453)

   Goodness of fit:
     SSE: 4.125
     R-square: 0.776
     Adjusted R-square: 0.7653
     RMSE: 0.1477

   Goodness of validation:
     SSE : 2.26745
     RMSE : 0.155312
```

13 To export this fit information to the workspace, select **Fit > Save to Workspace**. Executing this command also exports other information such as the numbers of observations and parameters, residuals, and the fitted model.

You can treat the fitted model as a function to make predictions or evaluate the surface at values of X and Y. For details see "Exporting a Fit to the Workspace" on page 7-13.

14 Display the residuals plot to check the distribution of points relative to the surface. Click the toolbar button [icon] or select **View > Residuals Plot**.
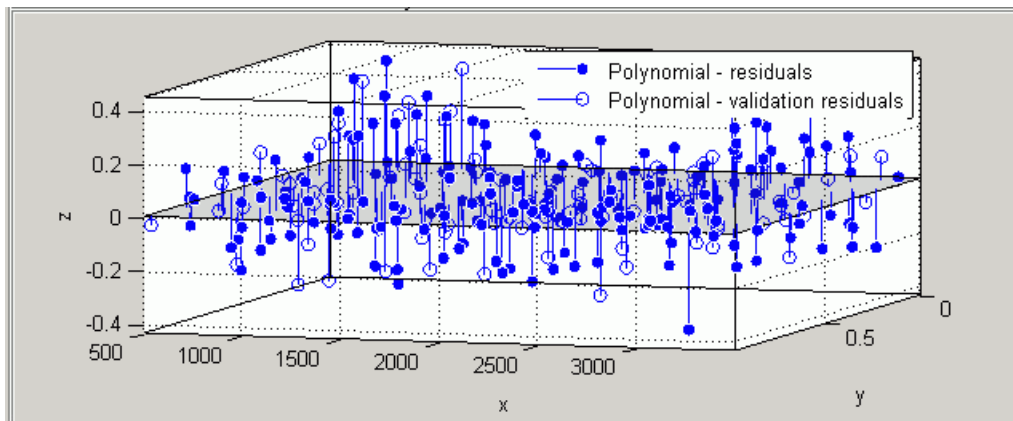
15 Right-click the residuals plot to select the **Go to X-Z view**. The X-Z view is not required, but the view makes it easier to see to remove outliers.

16 To remove outliers, click the toolbar button or select **Tools > Exclude Outliers**.

When you move the mouse cursor to the plot, it changes to a cross-hair to show you are in outlier selection mode.

a Click a point that you want to exclude in the surface plot or residuals plot. Alternatively, click and drag to define a rectangle and remove all enclosed points.

A removed plot point displays as a red star in the plots.
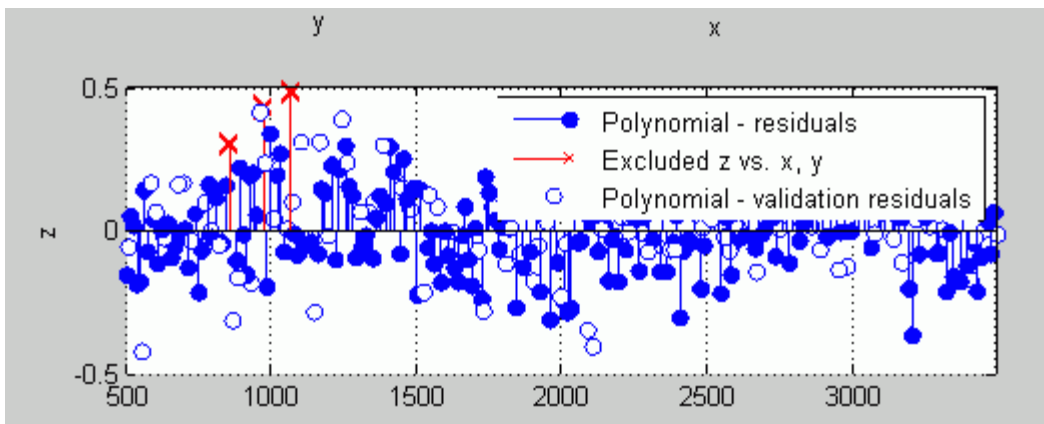
b If you have **Auto-fit** selected, the Curve Fitting Tool refits the surface without the point. Otherwise, you can click **Fit** to refit the surface.

c To return to rotation mode, click the toolbar button again to switch off **Exclude Outliers** mode.

17 To compare your fits side-by-side, use the tile tools. Select **Window > Left/Right Tile**, or use the toolbar buttons.
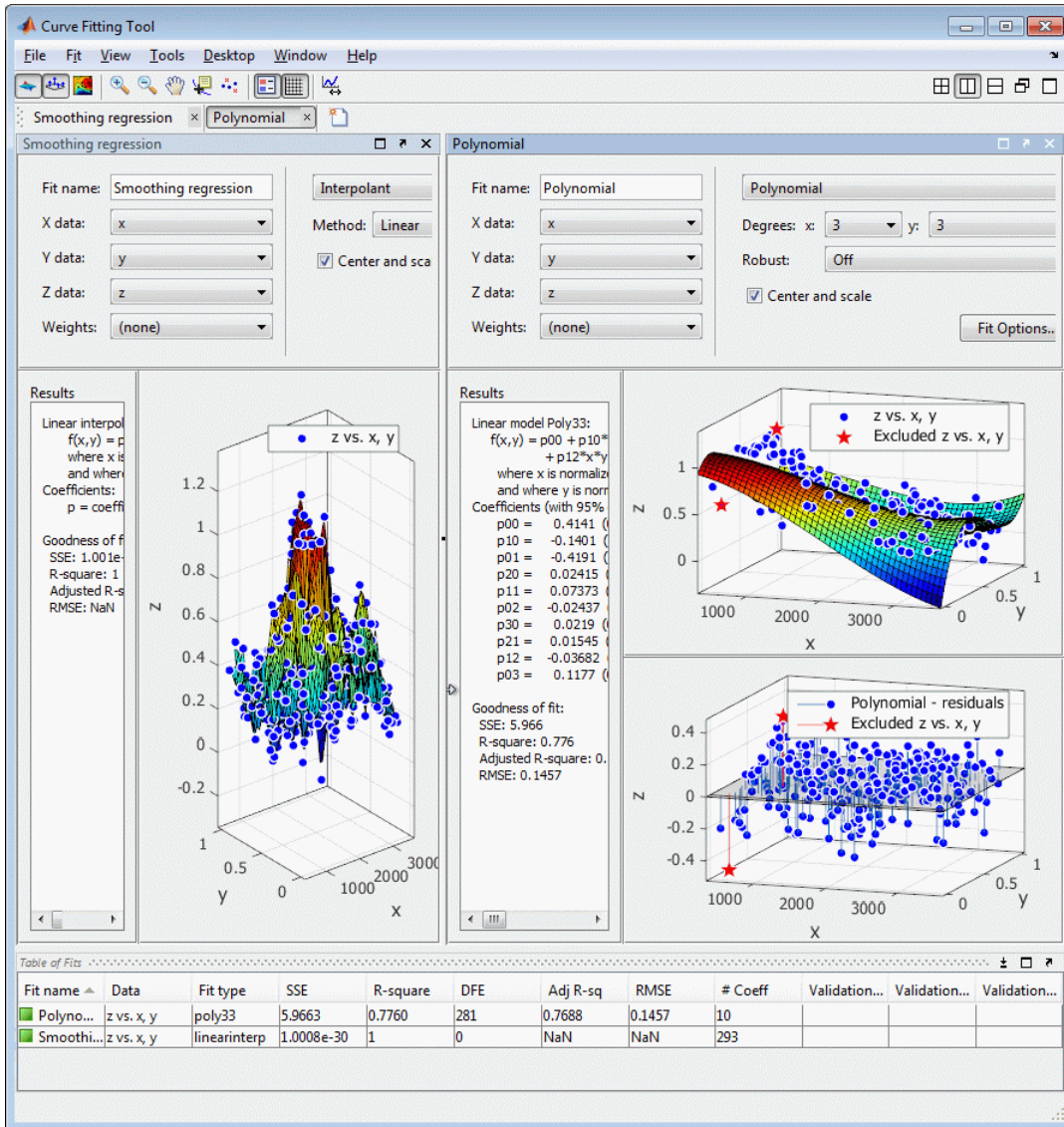
18 Review the information in the **Table of Fits**. Compare goodness-of-fit statistics for all fits in your session to determine which is best.

19 To save your interactive surface fitting session, select **File > Save Session**. You can save and reload sessions to access multiple fits. The session file contains all the fits and variables in your session and remembers your layout.

20 After interactively creating and comparing fits, you can generate code for all fits and plots in your Curve Fitting Tool session. Select **File > Generate Code**.

The Curve Fitting Tool generates code from your session and displays the file in the MATLAB Editor. The file includes all fits and plots in your current session.

21 Save the file with the default name, `createFits.m`.

22 You can recreate your fits and plots by calling the file from the command line (with your original data or new data as input arguments). In this case, your original variables still appear in the workspace.

- Highlight and evaluate the first line of the file (excluding the word `function`). Either right-click and select **Evaluate**, press **F9**, or copy and paste the following to the command line:

  ```
  [fitresult, gof] = createFits(x, y, z, xv, yv, zv)
  ```

- The function creates a figure window for each fit you had in your session. Observe that the polynomial fit figure shows both the surface and residuals plots that you created interactively in the Curve Fitting Tool.

- If you want you can use the generated code as a starting point to change the surface fits and plots to fit your needs. For a list of methods you can use, see `sfit`.

For more information on all fit settings and tools for comparing fits, see:

- "Interactive Fit Comparison" on page 2-13
- "Linear and Nonlinear Regression"
- "Interpolation"
- "Smoothing"
- "Fit Postprocessing"

Programmatic Curve and Surface Fitting

Curve and Surface Fitting

In this section...
"Fitting a Curve" on page 3-2
"Fitting a Surface" on page 3-2
"Model Types and Fit Analysis" on page 3-3
"Workflow for Command Line Fitting" on page 3-3

Fitting a Curve

To programmatically fit a curve, follow the steps in this simple example:

1 Load some data.

```
load hahn1
```

Create a fit using the `fit` function, specifying the variables and a model type (in this case `rat23` is the model type).

```
f = fit( temp, thermex, 'rat23' )
```

Plot your fit and the data.

```
plot( f, temp, thermex )
f( 600 )
```

Fitting a Surface

To programmatically fit a surface, follow the steps in this simple example:

1 Load some data.

```
load franke
```

2 Create a fit using the `fit` function, specifying the variables and a model type (in this case `poly23` is the model type).

```
 f = fit( [x, y], z, 'poly23' )
```

3 Plot your fit and the data.

```
plot(f, [x,y], z)
```

Model Types and Fit Analysis

For details and examples of specific model types and fit analysis, see the following sections:

1 "Linear and Nonlinear Regression"

2 "Interpolation"

3 "Smoothing"

4 "Fit Postprocessing"

Workflow for Command Line Fitting

Curve Fitting Toolbox software provides a variety of methods for data analysis and modeling. These methods are applied in a systematic manner, which can be represented in a workflow diagram such as the one below.
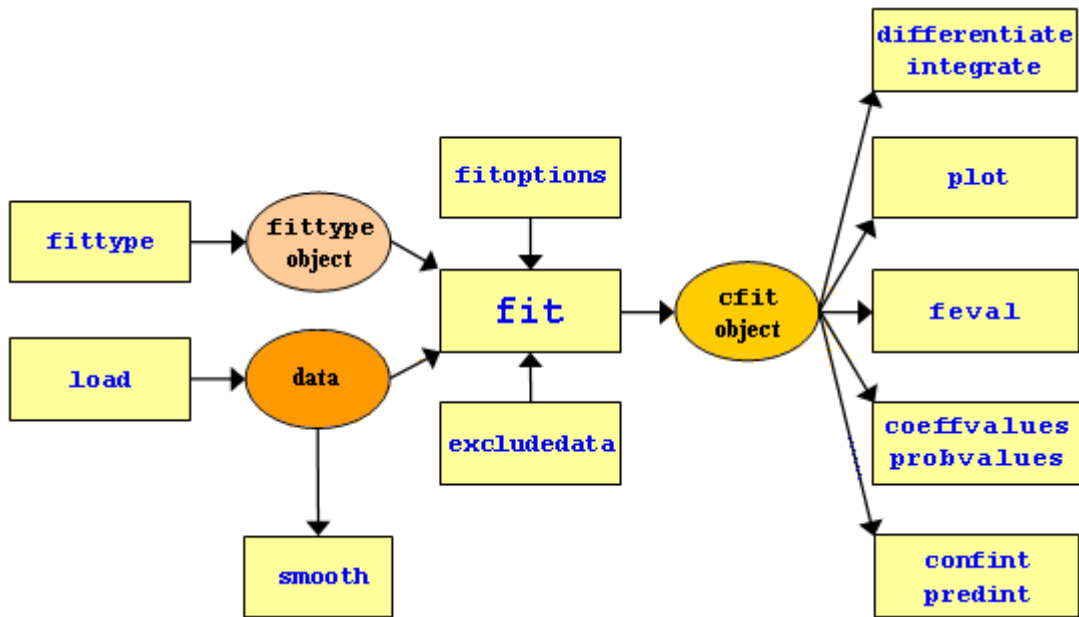
A typical analysis using curve fitting methods proceeds as follows:

1 Import your data into the MATLAB workspace using the `load` command
   (if your data has previously been stored in MATLAB variables) or any of
   the more specialized MATLAB functions for reading data from particular
   file types.

2 (Optional) If your data is noisy, you might want to smooth it using the
   `smooth` function. Smoothing is used to identify major trends in the data
   that can assist you in choosing an appropriate family of parametric models.
   If a parametric model is not evident or appropriate, smoothing can be an
   end in itself, providing a nonparametric fit of the data.

> **Note** Smoothing estimates the center of the distribution of the response at each predictor. It invalidates the assumption that errors in the data are independent, and so also invalidates the methods used to compute confidence and prediction intervals. Accordingly, once a parametric model is identified through smoothing, the *original* data should be passed to the `fit` function.

3 Specify a parametric model for the data—either a Curve Fitting Toolbox library model or a custom model that you define. You can specify the model by passing a string or expression to the `fit` function or (optional) with a `fittype` object you create with the `fittype` function.

To view available library models, see "List of Library Models for Curve and Surface Fitting" on page 4-13.

4 (Optional) You can create a fit options structure for the fit using the `fitoptions` function. Fit options specify things like weights for the data, fitting methods, and low-level options for the fitting algorithm.

5 (Optional) You can create an exclusion rule for the fit using the `excludedata` function. Exclusion rules indicate which data values will be treated as outliers and excluded from the fit.

6 You pass the data, a model (string, expression or `fittype` object), and (optionally) a fit options structure and an exclusion rule, to the `fit` function to perform the fit.

The `fit` function returns a `cfit` (for curves) or `sfit` (for surfaces) object that encapsulates the computed coefficients and the fit statistics.

7 You can then postprocess the `cfit` and `sfit` objects returned by the `fit` function, by passing them to a variety of functions, such as `feval`, `differentiate`, `integrate`, `plot`, `coeffvalues`, `probvalues`, `confint`, and `predint`.

For more details on functions, see "Curve and Surface Fitting Objects and Methods" on page 3-6.

Curve and Surface Fitting Objects and Methods

In this section...

This section describes how to use Curve Fitting Toolbox functions from the command-line or to write programs for curve and surface fitting applications.

The Curve Fitting Tool is a graphical user interface that allows convenient, interactive use of Curve Fitting Toolbox functions, without programming. You can, however, access Curve Fitting Toolbox functions directly, and write programs that combine curve fitting functions with MATLAB functions and functions from other toolboxes. This allows you to create a curve fitting environment that is precisely suited to your needs.

Models and fits in Curve Fitting Tool are managed internally as curve fitting *objects*. Objects are manipulated through a variety of functions called *methods*. You can create curve fitting objects, and apply curve fitting methods, outside of Curve Fitting Tool.

Curve Fitting Objects

In MATLAB programming, all workspace variables are *objects* of a particular *class*. Familiar examples of MATLAB classes are `double`, `char`, and `function_handle`. You can also create custom MATLAB classes, using object-oriented programming.

Methods are functions that operate exclusively on objects of a particular class. *Data types* package together objects and methods so that the methods operate exclusively on objects of their own type, and not on objects of other types. A clearly defined encapsulation of objects and methods is the goal of object-oriented programming.

Curve Fitting Toolbox software provides you with new MATLAB data types for performing curve fitting:

- `fittype` — Objects allow you to encapsulate information describing a parametric model for your data. Methods allow you to access and modify that information.

- `cfit` and `sfit` — Two subtypes of `fittype`, for curves and surfaces. Objects capture information from a particular fit by assigning values to coefficients, confidence intervals, fit statistics, etc. Methods allow you to post-process the fit through plotting, extrapolation, integration, etc.
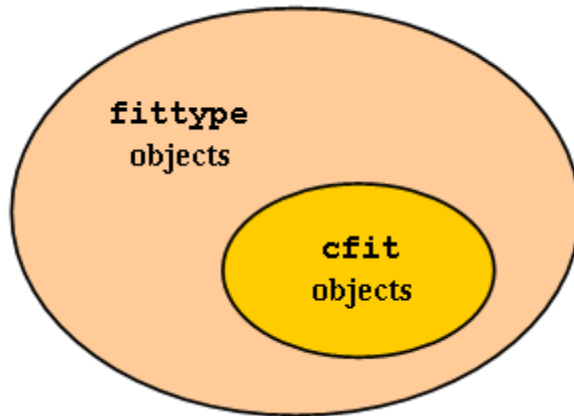
Because `cfit` is a subtype of `fittype`, `cfit` inherits all `fittype` methods. In other words, you can apply `fittype` methods to both `fittype` and `cfit` objects, but `cfit` methods are used exclusively with `cfit` objects. Similarly for `sfit` objects.

As an example, the `fittype` method `islinear`, which determines if a model is linear or nonlinear, would apply equally well before or after a fit; that is, to both `fittype` and `cfit` objects. On the other hand, the `cfit` methods `coeffvalues` and `confint`, which, respectively, return fit coefficients and their confidence intervals, would make no sense if applied to a general `fittype` object which describes a parametric model with undetermined coefficients.

Curve fitting objects have properties that depend on their type, and also on the particulars of the model or the fit that they encapsulate. For example, the following code uses the constructor methods for the two curve fitting types to create a `fittype` object `f` and a `cfit` object `c`:

```
f = fittype('a*x^2+b*exp(n*x)')
f =
     General model:
       f(a,b,n,x) = a*x^2+b*exp(n*x)
c = cfit(f,1,10.3,-1e2)
c =
     General model:
       c(x) = a*x^2+b*exp(n*x)
     Coefficients:
       a =             1
       b =          10.3
       n =          -100
```

Note that the display method for `fittype` objects returns only basic information, piecing together outputs from `formula` and `indepnames`.

`cfit` and `fittype` objects are evaluated at predictor values `x` using `feval`. You can call `feval` indirectly using the following functional syntax:

```
y = cfun(x) % cfit objects;
y = ffun(coef1,coef2,...,x) % fittype objects;
```

Curve Fitting Methods

Curve fitting methods allow you to create, access, and modify curve fitting objects. They also allow you, through methods like `plot` and `integrate`, to perform operations that uniformly process the entirety of information encapsulated in a curve fitting object.

The methods listed in the following table are available for all `fittype` objects, including `cfit` objects.

Fit Type Method	Description
argnames	Get input argument names
category	Get fit category
coeffnames	Get coefficient names
dependnames	Get dependent variable name
feval	Evaluate model at specified predictors

Fit Type Method	Description
fittype	Construct fittype object
formula	Get formula string
indepnames	Get independent variable name
islinear	Determine if model is linear
numargs	Get number of input arguments
numcoeffs	Get number of coefficients
probnames	Get problem-dependent parameter names
setoptions	Set model fit options
type	Get name of model

The methods listed in the following table are available exclusively for cfit objects.

Curve Fit Method	Description
cfit	Construct cfit object
coeffvalues	Get coefficient values
confint	Get confidence intervals for fit coefficients
differentiate	Differentiate fit
integrate	Integrate fit
plot	Plot fit
predint	Get prediction intervals
probvalues	Get problem-dependent parameter values

A complete list of methods for a curve fitting object can be obtained with the MATLAB methods command. For example,

```
f = fittype('a*x^2+b*exp(n*x)');
methods(f)

Methods for class fittype:
```

```
argnames     dependnames  fittype     islinear    probnames
category     feval        formula     numargs     setoptions
coeffnames   fitoptions   indepnames  numcoeffs   type
```

Note that some of the methods listed by `methods` do not appear in the tables above, and do not have reference pages in the Curve Fitting Toolbox documentation. These additional methods are generally low-level operations used by Curve Fitting Tool, and not of general interest when writing curve fitting applications.

There are no global accessor methods, comparable to `getfield` and `setfield`, available for `fittype` objects. Access is limited to the methods listed above. This is because many of the properties of `fittype` objects are derived from other properties, for which you do have access. For example,

```
f = fittype('a*cos( b*x-c )')
f =
     General model:
       f(a,b,c,x) = a*cos( b*x-c )

formula(f)
ans =
a*cos( b*x-c )

argnames(f)
ans =
    'a'
    'b'
    'c'
    'x'
```

You construct the `fittype` object `f` by giving the formula, so you do have write access to that basic property of the object. You have read access to that property through the `formula` method. You also have read access to the argument names of the object, through the `argnames` method. You don't, however, have direct write access to the argument names, which are derived from the formula. If you want to set the argument names, set the formula.

Surface Fitting Objects and Methods

Surface Fitting Objects and Methods

The surface fit object (`sfit`) stores the results from a surface fitting operation, making it easy to plot and analyze fits at the command line.

Like `cfit` objects, `sfit` objects are a subclass of `fittype` objects, so they inherit all the same methods of `fittype` listed in "Curve Fitting Methods" on page 3-8.

`sfit` objects also provide methods exclusively for `sfit` objects. See `sfit`.

One way to quickly assemble code for surface fits and plots into useful programs is to generate a file from a session in Curve Fitting Tool. In this way, you can transform your interactive analysis of a single data set into a reusable function for command-line analysis or for batch processing of multiple data sets. You can use the generated file without modification, or edit and customize the code as needed. See "Generating Code and Exporting Fits to the Workspace" on page 7-12.

4

Linear and Nonlinear Regression

Parametric Fitting

Parametric Fitting with Library Models

Parametric fitting involves finding coefficients (parameters) for one or more models that you fit to data. The data is assumed to be statistical in nature and is divided into two components:

data = deterministic component + random component

The deterministic component is given by a parametric model and the random component is often described as error associated with the data:

data = parametric model + error

The model is a function of the independent (predictor) variable and one or more coefficients. The error represents random variations in the data that follow a specific probability distribution (usually Gaussian). The variations can come from many different sources, but are always present at some level when you are dealing with measured data. Systematic variations can also exist, but they can lead to a fitted model that does not represent the data well.

The model coefficients often have physical significance. For example, suppose you collected data that corresponds to a single decay mode of a radioactive nuclide, and you want to estimate the half-life ($T_{1/2}$) of the decay. The law of radioactive decay states that the activity of a radioactive substance decays exponentially in time. Therefore, the model to use in the fit is given by

$$y = y_0 e^{-\lambda t}$$

where y_0 is the number of nuclei at time $t = 0$, and λ is the decay constant. The data can be described by

$$\text{data} = y_0 e^{-\lambda t} + \text{error}$$
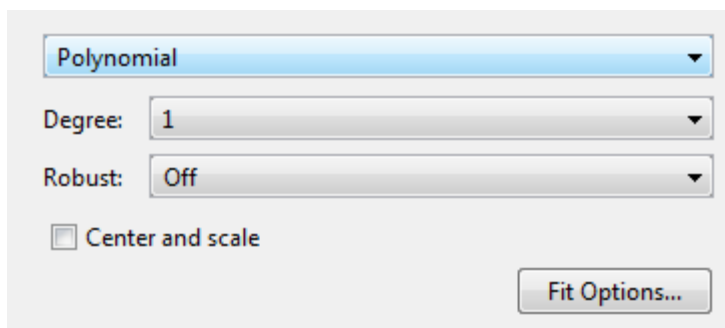
Both y_0 and λ are coefficients that are estimated by the fit. Because $T_{1/2}$ = $\ln(2)/\lambda$, the fitted value of the decay constant yields the fitted half-life. However, because the data contains some error, the deterministic component of the equation cannot be determined exactly from the data. Therefore, the coefficients and half-life calculation will have some uncertainty associated with them. If the uncertainty is acceptable, then you are done fitting the data. If the uncertainty is not acceptable, then you might have to take steps to reduce it either by collecting more data or by reducing measurement error and collecting new data and repeating the model fit.

With other problems where there is no theory to dictate a model, you might also modify the model by adding or removing terms, or substitute an entirely different model.

The Curve Fitting Toolbox parametric library models are described in the following sections.

Selecting a Model Type Interactively

Select a model type to fit from the drop-down list in the Curve Fitting Tool.

What fit types can you use for curves or surfaces? Based on your selected data, the fit category list shows either curve or surface categories. The following table describes the options for curves and surfaces.

Fit Category	Curves	Surfaces
Regression Models		
Polynomial	Yes (up to degree 9)	Yes (up to degree 5)
Exponential	Yes	
Fourier	Yes	
Gaussian	Yes	
Power	Yes	
Rational	Yes	
Sum of Sine	Yes	
Weibull	Yes	
Interpolation		
Interpolant	Yes Methods: Nearest neighbor Linear Cubic Shape-preserving (PCHIP)	Yes Methods: Nearest neighbor Linear Cubic Biharmonic
Smoothing		
Smoothing Spline	Yes	
Lowess		Yes
Custom		
Custom Equation	Yes	Yes

For all fit categories, look in the **Results** pane to see the model terms, the values of the coefficients, and the goodness-of-fit statistics.

> **Tip** If your fit has problems, messages in the **Results** pane help you identify better settings.

Selecting Fit Settings

The Curve Fitting Tool provides a selection of fit types and settings that you can change to try to improve your fit. Try the defaults first, then experiment with other settings.

For an overview of how to use the available fit options, see "Specifying Fit Options and Optimized Starting Points" on page 4-7.

You can try a variety of settings within a single fit figure, and you can also create multiple fits to compare. When you create multiple fits you can compare different fit types and settings side by side in the Curve Fitting Tool. See "Interactive Fit Comparison" on page 2-13.

Selecting Model Type Programmatically

You can specify a library model name as a string when you call the `fit` function. For example, to specify a quadratic `poly2`:

```
f = fit( x, y, 'poly2' )
```

See "List of Library Models for Curve and Surface Fitting" on page 4-13 to view all available library model names.

You can also use the `fittype` function to construct a `fittype` object for a library model, and use the `fittype` as an input to the `fit` function.

Use the `fitoptions` function to find out what parameters you can set, for example:

```
fitoptions(poly2)
```

For examples, see the sections for each model type, listed in the table in "Selecting a Model Type Interactively" on page 4-3. For details on all the functions for creating and analysing models, see "Curve and Surface Fitting" on page 3-2.

Using Normalize or Center and Scale

Most model types in the Curve Fitting Tool share the **Center and scale** option. When you select this option, the tool refits with the data centered and scaled, by applying the `Normalize` setting to the variables. At the command line, you can use `Normalize` as an input argument to the `fitoptions` function. See the `fitoptions` reference page.

Generally, it is a good idea to normalize inputs (also known as *predictor data*), which can alleviate numerical problems with variables of different scales. For example, suppose your surface fit inputs are engine speed with a range of 500–4500 r/min and engine load percentage with a range of 0–1. Then, **Center and scale** generally improves the fit because of the great difference in scale between the two inputs. However, if your inputs are in the same units or similar scale (e.g., eastings and northings for geographic data), then **Center and scale** is less useful. When you normalize inputs with this option, the values of the fitted coefficients change when compared to the original data.

If you are fitting a curve or surface to estimate coefficients, or the coefficients have physical significance, clear the **Center and scale** check box. The Curve Fitting Tool plots use the original scale with or without the **Center and scale** option.

At the command line, to set the option to center and scale the data before fitting, create the default fit options structure, set `Normalize` to `on`, then fit with the options:

```
options = fitoptions;
options.Normal = 'on';
options
options =
    Normalize: 'on'
    Exclude: [1xO double]
    Weights: [1xO double]
    Method: 'None'

load census
f1 = fit(cdate,pop,'poly3',options)
```

Specifying Fit Options and Optimized Starting Points

About Fit Options

Interactive fit options are described in the following sections. To specify the same fit options programmatically, see "Specifying Fit Options at the Command Line" on page 4-11.

To specify fit options interactively in the Curve Fitting Tool, click the **Fit Options** button to open the Fit Options dialog box. All fit categories except interpolants and smoothing splines have configurable fit options.

The available options depend on whether you are fitting your data using a linear model, a nonlinear model, or a nonparametric fit type:

- All the options described next are available for nonlinear models.

- **Lower** and **Upper** coefficient constraints are the only fit options available in the dialog box for polynomial linear models. For polynomials you can set **Robust** in the Curve Fitting Tool, without opening the Fit Options dialog box.

- Nonparametric fit types have no additional fit options dialog box (interpolant, smoothing spline, and lowess).

The fit options for the single-term exponential are shown next. The coefficient starting values and constraints are for the census data.
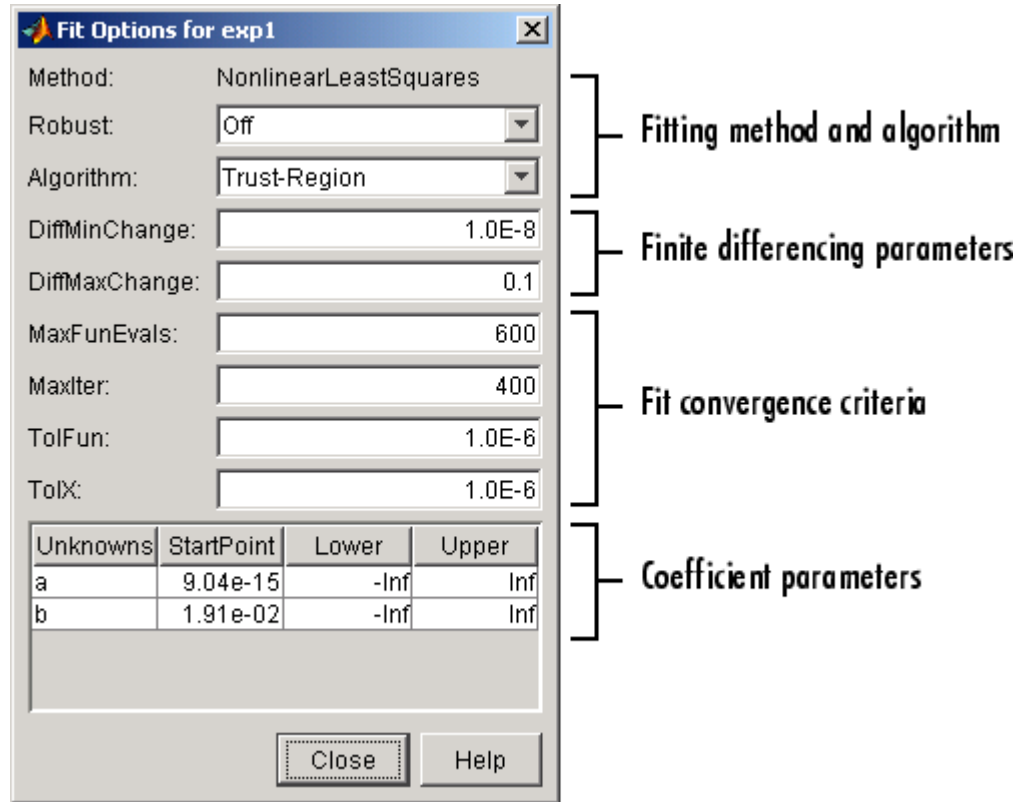
Fitting Method and Algorithm

- **Method** — The fitting method.

  The method is automatically selected based on the library or custom model you use. For linear models, the method is **LinearLeastSquares**. For nonlinear models, the method is **NonlinearLeastSquares**.

- **Robust** — Specify whether to use the robust least-squares fitting method.

  - **Off** — Do not use robust fitting (default).

  - **On** — Fit with the default robust method (bisquare weights).

  - **LAR** — Fit by minimizing the least absolute residuals (LAR).

- **Bisquare** — Fit by minimizing the summed square of the residuals, and reduce the weight of outliers using bisquare weights. In most cases, this is the best choice for robust fitting.

- **Algorithm** — Algorithm used for the fitting procedure:

  - **Trust-Region** — This is the default algorithm and must be used if you specify **Lower** or **Upper** coefficient constraints.

  - **Levenberg-Marquardt** — If the trust-region algorithm does not produce a reasonable fit, and you do not have coefficient constraints, try the Levenberg-Marquardt algorithm.

Finite Differencing Parameters

- **DiffMinChange** — Minimum change in coefficients for finite difference Jacobians. The default value is 10^{-8}.

- **DiffMaxChange** — Maximum change in coefficients for finite difference Jacobians. The default value is 0.1.

Note that **DiffMinChange** and **DiffMaxChange** apply to:

- Any nonlinear custom equation, that is, a nonlinear equation that you write

- Some of the nonlinear equations provided with Curve Fitting Toolbox software

However, **DiffMinChange** and **DiffMaxChange** do not apply to any linear equations.

Fit Convergence Criteria

- **MaxFunEvals** — Maximum number of function (model) evaluations allowed. The default value is 600.

- **MaxIter** — Maximum number of fit iterations allowed. The default value is 400.

- **TolFun** — Termination tolerance used on stopping conditions involving the function (model) value. The default value is 10^{-6}.

- **TolX** — Termination tolerance used on stopping conditions involving the coefficients. The default value is 10^{-6}.

Coefficient Parameters

- **Unknowns** — Symbols for the unknown coefficients to be fitted.

- **StartPoint** — The coefficient starting values. The default values depend on the model. For rational, Weibull, and custom models, default values are randomly selected within the range [0,1]. For all other nonlinear library models, the starting values depend on the data set and are calculated heuristically. See optimized starting points below.

- **Lower** — Lower bounds on the fitted coefficients. The tool only uses the bounds with the trust region fitting algorithm. The default lower bounds for most library models are -Inf, which indicates that the coefficients are unconstrained. However, a few models have finite default lower bounds. For example, Gaussians have the width parameter constrained so that it cannot be less than 0. See default contraints below.

- **Upper** — Upper bounds on the fitted coefficients. The tool only uses the bounds with the trust region fitting algorithm. The default upper bounds for all library models are Inf, which indicates that the coefficients are unconstrained.

For more information about these fit options, see the lsqcurvefit function in the Optimization Toolbox documentation.

Optimized Starting Points and Default Constraints

The default coefficient starting points and constraints for library and custom models are shown in the next table. If the starting points are optimized, then they are calculated heuristically based on the current data set. Random starting points are defined on the interval [0,1] and linear models do not require starting points.

If a model does not have constraints, the coefficients have neither a lower bound nor an upper bound. You can override the default starting points and constraints by providing your own values using the Fit Options dialog box.

Default Starting Points and Constraints

Model	Starting Points	Constraints
Custom linear	N/A	None
Custom nonlinear	Random	None
Exponential	Optimized	None
Fourier	Optimized	None
Gaussian	Optimized	$c_i > 0$
Polynomial	N/A	None
Power	Optimized	None
Rational	Random	None
Sum of Sine	Optimized	$b_i > 0$
Weibull	Random	$a, b > 0$

Note that the sum of sines and Fourier series models are particularly sensitive to starting points, and the optimized values might be accurate for only a few terms in the associated equations.

Specifying Fit Options at the Command Line

Create the default fit options structure and set the option to center and scale the data before fitting:

```
options = fitoptions;
options.Normal = 'on';
options
options =
    Normalize: 'on'
    Exclude: [1xO double]
    Weights: [1xO double]
    Method: 'None'
```

Modifying the default fit options structure is useful when you want to set the Normalize, Exclude, or Weights fields, and then fit your data using the same options with different fitting methods. For example:

```
load census
f1 = fit(cdate,pop,'poly3',options);
f2 = fit(cdate,pop,'exp1',options);
f3 = fit(cdate,pop,'cubicsp',options);
```

Data-dependent fit options are returned in the third output argument of the fit function. For example, the smoothing parameter for smoothing spline is data-dependent:

```
[f,gof,out] = fit(cdate,pop,'smooth');
smoothparam = out.p
smoothparam =
    0.0089
```

Use fit options to modify the default smoothing parameter for a new fit:

```
options = fitoptions('Method','Smooth','SmoothingParam',0.0098);
[f,gof,out] = fit(cdate,pop,'smooth',options);
```

For more details on using fit options, see the fitoptions reference page.

List of Library Models for Curve and Surface Fitting

In this section...
"Use Library Models to Fit Data" on page 4-13
"Library Model Types" on page 4-13
"Model Names and Equations" on page 4-14

Use Library Models to Fit Data

You can use the Curve Fitting Toolbox library of models for data fitting with the `fit` function. You use library model names as input arguments in the `fit`, `fitoptions`, and `fittype` functions.

Library Model Types

The following tables describe the library model types for curves and surfaces.

- Use the links in the table for examples and detailed information on each library type.

- If you want a quick reference of model names for input arguments to the `fit` function, see "Model Names and Equations" on page 4-14.

Library Model Types for Curves	Description
`distribution`	Distribution models such as Weibull. See "Weibull Distributions" on page 4-44.
`exponential`	Exponential function and sum of two exponential functions. See "Exponential Models" on page 4-25.
`fourier`	Up to eight terms of Fourier series. See "Fourier Series" on page 4-29.
`gaussian`	Sum of up to eight Gaussian models. See "Gaussian Models" on page 4-31.
`interpolant`	Interpolating models, including linear, nearest neighbor, cubic spline, and shape-preserving cubic spline. See "Nonparametric Fitting" on page 6-2.

Library Model Types for Curves	Description
polynomial	Polynomial models, up to degree nine. See "Polynomial Models" on page 4-19.
power	Power function and sum of two power functions. See "Power Series" on page 4-33.
rational	Rational equation models, up to 5th degree/5th degree (i.e., up to degree 5 in both the numerator and the denominator). See "Rational Polynomials" on page 4-35.
sin	Sum of up to eight sin functions. See "Sum of Sines Models" on page 4-42.
spline	Cubic spline and smoothing spline models. See "Nonparametric Fitting" on page 6-2.

Library Model Types for Surfaces	Description
interpolant	Interpolating models, including linear, nearest neighbor, cubic spline, and biharmonic interpolation. See "Interpolants" on page 6-3.
lowess	Lowess smoothing models. See "Lowess Smoothing" on page 6-16.
polynomial	Polynomial models, up to degree five. See "Polynomial Models" on page 4-19.

Model Names and Equations

To specify the model you want to fit, consult the following tables for a model name to use as an input argument to the fit function. For example, to specify a quadratic curve with model name "poly2":

```
f = fit(x, y,  'poly2')
```

Polynomial Model Names and Equations

Examples of Polynomial Model Names for Curves	Equations
`poly1`	`Y = p1*x+p2`
`poly2`	`Y = p1*x^2+p2*x+p3`
`poly3`	`Y = p1*x^3+p2*x^2+...+p4`
...etc., up to `poly9`	`Y = p1*x^9+p2*x^8+...+p10`

For polynomial surfaces, model names are `'polyij'`, where *i* is the degree in *x* and *j* is the degree in *y*. The maximum for both *i* and *j* is five. The degree of the polynomial is the maximum of *i* and *j*. The degree of *x* in each term will be less than or equal to *i*, and the degree of *y* in each term will be less than or equal to *j*. See the following table for some example model names and equations, of many potential examples.

Examples of Polynomial Model Names for Surfaces	Equations
`poly21`	`Z = p00 + p10*x + p01*y + p20*x^2 + p11*x*y`
`poly13`	`Z = p00 + p10*x + p01*y + p11*x*y + p02*y^2 + p12*x*y^2 + p03*y^3`
`poly55`	`Z = p00 + p10*x + p01*y +...+ p14*x*y^4 + p05*y^5`

Distribution Model Name and Equation

Distribution Model Names	Equations
`weibull`	`Y = a*b*x^(b-1)*exp(-a*x^b)`

Exponential Model Names and Equations

Exponential Model Names	Equations
exp1	Y = a*exp(b*x)
exp2	Y = a*exp(b*x)+c*exp(d*x)

Fourier Series Model Names and Equations

Fourier Series Model Names	Equations
fourier1	Y = a0+a1*cos(x*p)+b1*sin(x*p)
fourier2	Y = a0+a1*cos(x*p)+b1*sin(x*p)... +a2*cos(2*x*p)+b2*sin(2*x*p)
fourier3	Y = a0+a1*cos(x*p)+b1*sin(x*p)... +a3*cos(3*x*p)+b3*sin(3*x*p)
...etc., up to fourier8	Y = a0+a1*cos(x*p)+b1*sin(x*p)... +a8*cos(8*x*p)+b8*sin(8*x*p)

Where p = 2*pi/(max(xdata)-min(xdata)).

Gaussian Model Names and Equations

Gaussian Model Names	Equations
gauss1	Y = a1*exp(-((x-b1)/c1)^2)
gauss2	Y = a1*exp(-((x-b1)/c1)^2)+a2*... exp(-((x-b2)/c2)^2)
gauss3	Y = a1*exp(-((x-b1)/c1)^2)+... +a3*exp(-((x-b3)/c3)^2)
...etc., up to gauss8	Y = a1*exp(-((x-b1)/c1)^2) +... +a8*exp(-((x-b8)/c8)^2)

Power Model Names and Equations

Power Model Names	Equations
power1	Y = a*x^b
power2	Y = a*x^b+c

Rational Model Names and Equations

Rational models are polynomials over polynomials with the leading coefficient of the denominator set to 1. Model names are ratij, where i is the degree of the numerator and j is the degree of the denominator. The degrees go up to five for both the numerator and the denominator.

Examples of Rational Model Names	Equations
rat02	Y = (p1)/(x^2+q1*x+q2)
rat21	Y = (p1*x^2+p2*x+p3)/(x+q1)
rat55	Y = (p1*x^5+...+p6)/(x^5+...+q5)

Sum of Sine Model Names and Equations

Sum of Sine Model Names	Equations
sin1	Y = a1*sin(b1*x+c1)
sin2	Y = a1*sin(b1*x+c1)+a2*sin...(b2*x+c2)
sin3	Y = a1*sin(b1*x+c1)+...+a3*sin(b3*x+c3)
...etc., up to sin8	Y = a1*sin(b1*x+c1)+...+a8*sin(b8*x+c8)

Spline Model Names

Spline models are supported for curve fitting, not for surface fitting.

Spline Model Names	Description
cubicspline	Cubic interpolating spline
smoothingspline	Smoothing spline

Interpolant Model Names

Type	Interpolant Model Names	Description
Curves and Surfaces	linearinterp	Linear interpolation
	nearestinterp	Nearest neighbor interpolation
	cubicinterp	Cubic spline interpolation
Curves only	pchipinterp	Shape-preserving (pchip) interpolation
Surfaces only	biharmonicinterp	Biharmonic (MATLAB 4 griddata) interpolation

Lowess Model Names

Lowess models are supported for surface fitting, not for curve fitting.

Lowess Model Names	Description
lowess	Local linear regression
loess	Local quadratic regression

Polynomial Models

In this section...
"About Polynomial Models" on page 4-19
"Selecting a Polynomial Fit Interactively" on page 4-20
"Selecting a Polynomial Fit at the Command Line" on page 4-22
"Defining Polynomial Terms for Polynomial Surface Fits" on page 4-23

About Polynomial Models

Polynomial models for curves are given by

$$y = \sum_{i=1}^{n+1} p_i x^{n+1-i}$$

where $n + 1$ is the *order* of the polynomial, n is the *degree* of the polynomial, and $1 \le n \le 9$. The order gives the number of coefficients to be fit, and the degree gives the highest power of the predictor variable.

In this guide, polynomials are described in terms of their degree. For example, a third-degree (cubic) polynomial is given by

$$y = p_1 x^3 + p_2 x^2 + p_3 x + p_4$$

Polynomials are often used when a simple empirical model is required. You can use the polynomial model for interpolation or extrapolation, or to characterize data using a global fit. For example, the temperature-to-voltage conversion for a Type J thermocouple in the 0 to 760° temperature range is described by a seventh-degree polynomial.

Note If you do not require a global parametric fit and want to maximize the flexibility of the fit, piecewise polynomials might provide the best approach. Refer to "Nonparametric Fitting" on page 6-2 for more information.

The main advantages of polynomial fits include reasonable flexibility for data that is not too complicated, and they are linear, which means the fitting process is simple. The main disadvantage is that high-degree fits can become unstable. Additionally, polynomials of any degree can provide a good fit within the data range, but can diverge wildly outside that range. Therefore, exercise caution when extrapolating with polynomials.
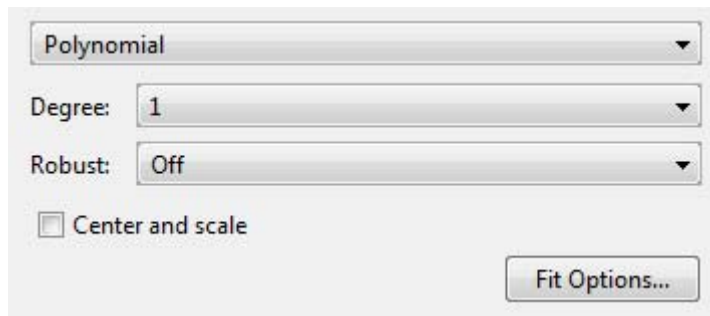
When you fit with high-degree polynomials, the fitting procedure uses the predictor values as the basis for a matrix with very large values, which can result in scaling problems. To handle this, you should normalize the data by centering it at zero mean and scaling it to unit standard deviation. Normalize data by selecting the **Center and scale** check box in the Curve Fitting Tool.

Selecting a Polynomial Fit Interactively

In the Curve Fitting Tool, select `Polynomial` from the model type list.
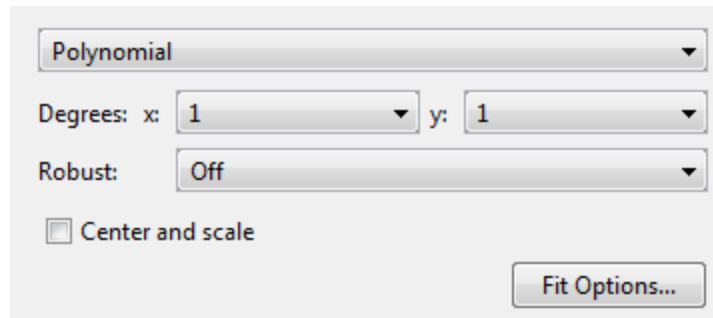
The Polynomial fit uses the Curve Fitting Toolbox `polynomial` library model.

For *curves*, the `Polynomial` model fits a polynomial in **x**.

For *surfaces*, the `Polynomial` model fits a polynomial in **x** and **y**.

You can specify the following options:

- The degree for the **x** and **y** inputs:

  - For curves, degree of **x** can be up to 9.

  - For surfaces, degree of **x** and **y** can be up to 5.

    The degree of the polynomial is the maximum of **x** and **y** degrees. See "Defining Polynomial Terms for Polynomial Surface Fits" on page 4-23.

- The robust linear least-squares fitting method to use (`Off`, `LAR`, or `Bisquare`). For details, see `Robust` on the `fitoptions` reference page.

- Set bounds or exclude terms by clicking **Fit Options**. You can exclude any term by setting its bounds to 0.

  Look in the **Results** pane to see the model terms, the values of the coefficients, and the goodness-of-fit statistics.

Tip If your input variables have very different scales, select and clear the **Center and scale** check box to see the difference in the fit. Messages in the **Results** pane prompt you when scaling might improve your fit.

For an example comparing various polynomial fits, see "Curve Fitting to Census Data" on page 2-22.

Selecting a Polynomial Fit at the Command Line

The `polynomial` library model is an input argument to the `fit` and `fittype` functions. Specify the model type `poly` followed by the degree in `x` (up to 9), or `x` and `y` (up to 5).

For example, to specify a quadratic curve `poly2`:

```
f = fit(x, y,  'poly2')
```

To specify a cubic surface `poly33`:

```
f = fit([x, y] z, 'poly33')
```

To view available polynomial models and equations, see "Polynomial Model Names and Equations" on page 4-15.

To find out what parameters you can set:

```
fitoptions poly2
```

For example, to load some data and fit a cubic polynomial with center and scale and robust fitting options:

```
load census;
f=fit(cdate, pop, 'poly3', 'Normalize', 'on', 'Robust', 'on')
plot(f,cdate,pop)
```

Robust `'on'` is equivalent to `'Bisquare'`, the default method.

All fitting methods have the default properties `Normalize`, `Exclude`, `Weights`, and `Method`. For an example, see "Specifying Fit Options at the Command Line" on page 4-11.

Polynomial models have the `Method` property value `LinearLeastSquares`, and the additional fit options properties shown in the next table. For details on all fit options, see the `fitoptions` reference page.

Property	Description
Robust	Specifies the robust linear least-squares fitting method to use. Values are `'on'`, `'off'`, `'LAR'`, or `'Bisquare'`. The default is `'off'`. `'LAR'` specifies the least absolute residual method and `'Bisquare'` specifies the bisquare weights method. `'on'` is equivalent to `'Bisquare'`, the default method.
Lower	A vector of lower bounds on the coefficients to be fitted. The default value is an empty vector, indicating that the fit is unconstrained by lower bounds. If bounds are specified, the vector length must equal the number of coefficients. Individual unconstrained lower bounds can be specified by `-Inf`.
Upper	A vector of upper bounds on the coefficients to be fitted. The default value is an empty vector, indicating that the fit is unconstrained by upper bounds. If bounds are specified, the vector length must equal the number of coefficients. Individual unconstrained upper bounds can be specified by `Inf`.

Defining Polynomial Terms for Polynomial Surface Fits

You can control the terms to include in the polynomial surface model by specifying the degrees for the x and y inputs. If i is the degree in x and j is the degree in y, the total degree of the polynomial is the maximum of i and j. The degree of x in each term is less than or equal to i, and the degree of y in each term is less than or equal to j. The maximum for both i and j is five.

For example:

```
poly21  Z = p00 + p10*x + p01*y + p20*x^2 + p11*x*y


poly13  Z = p00 + p10*x + p01*y + p11*x*y + p02*y^2
                     + p12*x*y^2 + p03*y^3
```

```
poly55  Z = p00 + p10*x + p01*y +...+ p14*x*y^4
                     + p05*y^5
```

For example, if you specify an x degree of 3 and a y degree of 2, the model name is `poly32`. The model terms follow the form in this table.

Degree of Term	0	1	2
0	1	y	y^2
1	x	xy	xy^2
2	x^2	x^2y	N/A
3	x^3	N/A	N/A

The total degree of the polynomial cannot exceed the maximum of i and j. In this example, terms such as x^3y and x^2y^2 are excluded because their degrees sum to more than 3. In both cases, the total degree is 4.

Exponential Models

About Exponential Models

The toolbox provides a one-term and a two-term exponential model as given by

$$y = ae^{bx}$$
$$y = ae^{bx} + ce^{dx}$$

Exponentials are often used when the rate of change of a quantity is proportional to the initial amount of the quantity. If the coefficient associated with b and/or d is negative, y represents exponential decay. If the coefficient is positive, y represents exponential growth.
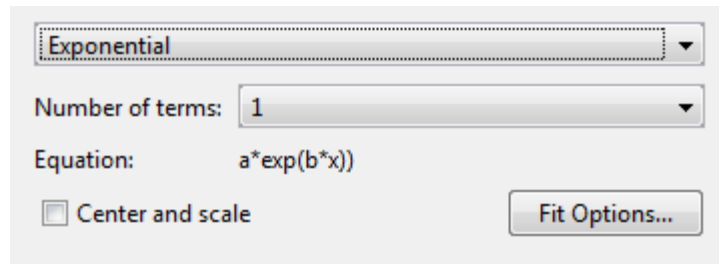
For example, a single radioactive decay mode of a nuclide is described by a one-term exponential. a is interpreted as the initial number of nuclei, b is the decay constant, x is time, and y is the number of remaining nuclei after a specific amount of time passes. If two decay modes exist, then you must use the two-term exponential model. For the second decay mode, you add another exponential term to the model.

Examples of exponential growth include contagious diseases for which a cure is unavailable, and biological populations whose growth is uninhibited by predation, environmental factors, and so on.

Selecting an Exponential Fit Interactively

In the Curve Fitting Tool, select `Exponential` from the model type list.

The exponential fit uses the `exp1` and `exp2` library models. These library models are input arguments to the `fit` and `fittype` functions.

You can specify the following options:

- Choose one or two terms to fit `exp1` or `exp2`.

  Look in the **Results** pane to see the model terms, the values of the coefficients, and the goodness-of-fit statistics.

- (Optional) Click **Fit Options** to specify coefficient starting values and constraint bounds appropriate for your data, or change algorithm settings.

  The toolbox calculates optimized start points for exponential fits, based on the current data set. You can override the start points and specify your own values in the Fit Options dialog box.

  The fit options for the single-term exponential are shown next. The coefficient starting values and constraints are for the census data.
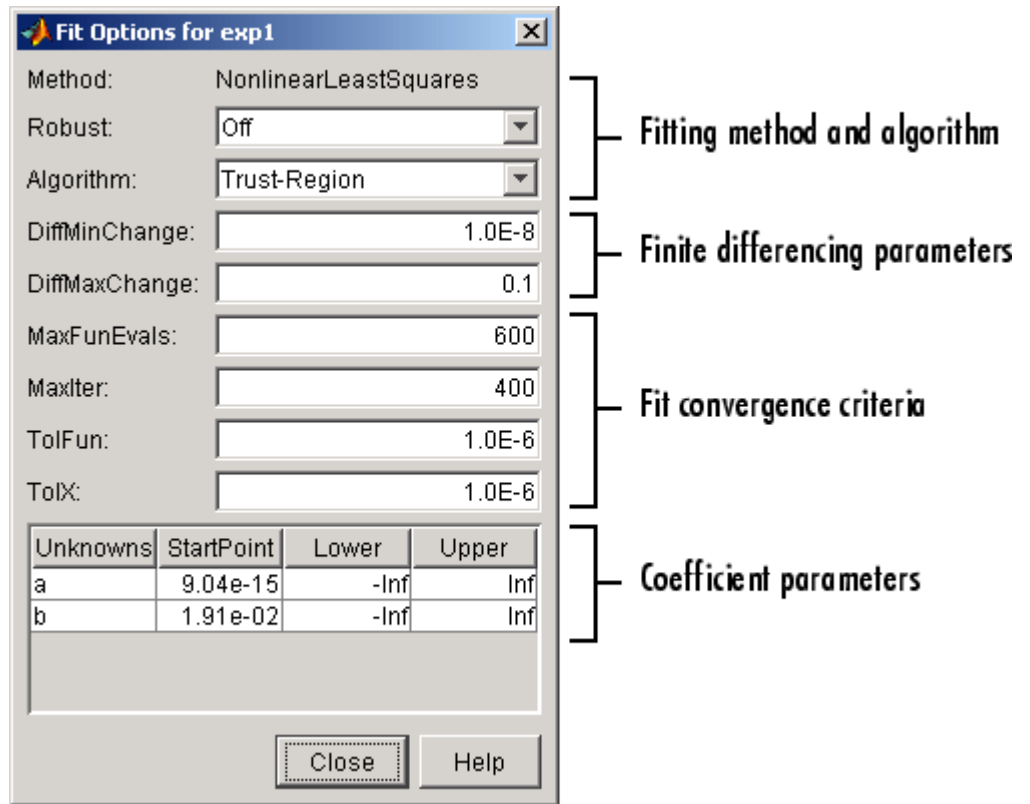
For an example specifying starting values appropriate to the data, see "Gaussian Fitting with an Exponential Background" on page 5-20.

For more information on the settings, see "Specifying Fit Options and Optimized Starting Points" on page 4-7.

Selecting an Exponential Fit at the Command Line

Specify the model type 'exp1' or 'exp2'.

For example, to generate data with an exponential trend and then fit the data using a single-term exponential:

```
x = (0:0.2:5)';
y = 2*exp(-0.2*x) + 0.5*randn(size(x));
```

```
f = fit(x,y,'exp1')
plot(f,x,y)
```

If you want to modify fit options such as coefficient starting values and constraint bounds appropriate for your data, or change algorithm settings, see the table of additional properties with `NonlinearLeastSquares` on the `fitoptions` reference page.

Fourier Series

In this section...
"About Fourier Series Models" on page 4-29
"Selecting a Fourier Fit Interactively" on page 4-29
"Selecting a Fourier Fit at the Command Line" on page 4-30

About Fourier Series Models

The Fourier series is a sum of sine and cosine functions that describes a periodic signal. It is represented in either the trigonometric form or the exponential form. The toolbox provides this trigonometric Fourier series form

$$y = a_0 + \sum_{i=1}^{n} a_i \cos(nwx) + b_i \sin(nwx)$$

where a_0 models a constant (intercept) term in the data and is associated with the $i = 0$ cosine term, w is the fundamental frequency of the signal, n is the number of terms (harmonics) in the series, and $1 \le n \le 8$.

For more information about the Fourier series, refer to "Fourier Analysis and Filtering".

Selecting a Fourier Fit Interactively

In the Curve Fitting Tool, select Fourier from the model type list.

You can specify the following options:

- Choose the number of terms: 1 to 8.

  Look in the **Results** pane to see the model terms, the values of the coefficients, and the goodness-of-fit statistics.

- (Optional) Click **Fit Options** to specify coefficient starting values and constraint bounds, or change algorithm settings.

  The toolbox calculates optimized start points for Fourier series models, based on the current data set. You can override the start points and specify your own values in the Fit Options dialog box.

  For more information on the settings, see "Specifying Fit Options and Optimized Starting Points" on page 4-7.

For an example comparing the library Fourier fit with custom equations, see "Fourier Analysis of ENSO Data" on page 5-9.

Selecting a Fourier Fit at the Command Line

Specify the model type `fourier` followed by the number of terms, e.g., `'fourier1'` to `'fourier8'`.

For example, to load some data and fit an eight-term Fourier model:

```
load enso;
f = fit(month,pressure,'fourier8')
plot(f,month,pressure)
```

If you want to modify fit options such as coefficient starting values and constraint bounds appropriate for your data, or change algorithm settings, see the table of additional properties with `NonlinearLeastSquares` on the `fitoptions` reference page.

Gaussian Models

About Gaussian Models

The Gaussian model fits peaks, and is given by

$$y = \sum_{i=1}^{n} a_i e^{\left[-\left(\frac{x - b_i}{c_i} \right)^2 \right]}$$

where a is the amplitude, b is the centroid (location), c is related to the peak width, n is the number of peaks to fit, and $1 \le n \le 8$.

Gaussian peaks are encountered in many areas of science and engineering. For example, Gaussian peaks can describe line emission spectra and chemical concentration assays.

Selecting a Gaussian Fit Interactively

In the Curve Fitting Tool, select Gaussian from the model type list.

You can specify the following options:

- Choose the number of terms: 1 to 8.

  Look in the **Results** pane to see the model terms, the values of the coefficients, and the goodness-of-fit statistics.

- (Optional) Click **Fit Options** to specify coefficient starting values and constraint bounds, or change algorithm settings.

  The toolbox calculates optimized start points for Gaussian models, based on the current data set. You can override the start points and specify your own values in the Fit Options dialog box.

  Gaussians have the width parameter c1 constrained with a lower bound of 0. The default lower bounds for most library models are -Inf, which indicates that the coefficients are unconstrained.

  For more information on the settings, see "Specifying Fit Options and Optimized Starting Points" on page 4-7.

Selecting a Gaussian Fit at the Command Line

Specify the model type gauss followed by the number of terms, e.g., 'gauss1' to 'gauss8'.

For example, to load some data and fit an eight-term Gaussian model:

```
load enso;
f = fit(month,pressure,'gauss8')
plot(f,month,pressure)
```

If you want to modify fit options such as coefficient starting values and constraint bounds appropriate for your data, or change algorithm settings, see the table of additional properties with NonlinearLeastSquares on the fitoptions reference page.

Power Series

About Power Series Models

The toolbox provides a one-term and a two-term power series model as given by

$$y = ax^b$$

$$y = a + bx^c$$

Power series models describe a variety of data. For example, the rate at which reactants are consumed in a chemical reaction is generally proportional to the concentration of the reactant raised to some power.

Selecting a Power Fit Interactively

In the Curve Fitting Tool, select Power from the model type list.
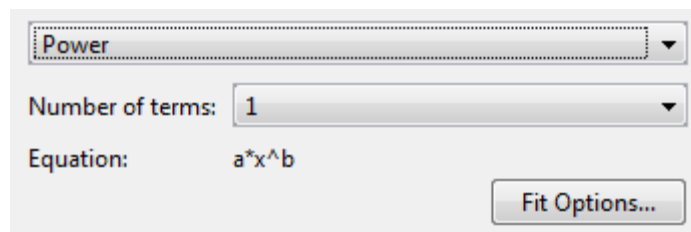
You can specify the following options:

- Choose the number of terms: 1 to 2.

  Look in the **Results** pane to see the model terms, the values of the coefficients, and the goodness-of-fit statistics.

4-33

- (Optional) Click **Fit Options** to specify coefficient starting values and constraint bounds, or change algorithm settings.

  The toolbox calculates optimized start points for power series models, based on the current data set. You can override the start points and specify your own values in the Fit Options dialog box.

  For more information on the settings, see "Specifying Fit Options and Optimized Starting Points" on page 4-7.

Selecting a Power Fit at the Command Line

Specify the model type power followed by the number of terms, e.g., `'power1'` or `'power2'`.

For example, to load some data and fit a two-term power series model:

```
load hahn1;
f = fit(temp,thermex,'power2')
plot(f,temp,thermex)
```

If you want to modify fit options such as coefficient starting values and constraint bounds appropriate for your data, or change algorithm settings, see the table of additional properties with `NonlinearLeastSquares` on the `fitoptions` reference page.

Rational Polynomials

About Rational Models

Rational models are defined as ratios of polynomials and are given by

$$y = \frac{\sum\limits_{i=1}^{n+1} p_i x^{n+1-i}}{x^m + \sum\limits_{i=1}^{m} q_i x^{m-1}}$$

where n is the degree of the numerator polynomial and $0 \leq n \leq 5$, while m is the degree of the denominator polynomial and $1 \leq m \leq 5$. Note that the coefficient associated with x^m is always 1. This makes the numerator and denominator unique when the polynomial degrees are the same.

In this guide, rationals are described in terms of the degree of the numerator/the degree of the denominator. For example, a quadratic/cubic rational equation is given by

$$y = \frac{p_1 x^2 + p_2 x + p_3}{x^3 + q_1 x^2 + q_2 x + q_3}$$

Like polynomials, rationals are often used when a simple empirical model is required. The main advantage of rationals is their flexibility with data that has a complicated structure. The main disadvantage is that they become unstable when the denominator is around 0. For an example that uses rational polynomials of various degrees, see "Example: Rational Fit" on page 4-37.

Selecting a Rational Fit Interactively

In the Curve Fitting Tool, select Rational from the model type list.

You can specify the following options:

• Choose the degree of the numerator and denominator polynomials. The numerator can have degree 0 to 5, and the denominator from 1 to 5.

  Look in the **Results** pane to see the model terms, the values of the coefficients, and the goodness-of-fit statistics.

• (Optional) Click **Fit Options** to specify coefficient starting values and constraint bounds, or change algorithm settings.

  The toolbox calculates random start points for rational models, defined on the interval [0,1]. You can override the start points and specify your own values in the Fit Options dialog box.

  For more information on the settings, see "Specifying Fit Options and Optimized Starting Points" on page 4-7.

Selecting a Rational Fit at the Command Line

Specify the model type ratij, where *i* is the degree of the numerator polynomial and *j* is the degree of the denominator polynomial. For example, 'rat02', 'rat21' or 'rat55'.

For example, to load some data and fit a rational model:

```
load hahn1;
f = fit( temp, thermex, 'rat32')
plot(f,temp,thermex)
```

See "Example: Rational Fit" on page 4-37 to fit this example interactively with various rational models.

If you want to modify fit options such as coefficient starting values and constraint bounds appropriate for your data, or change algorithm settings, see the table of additional properties with `NonlinearLeastSquares` on the `fitoptions` reference page.

Example: Rational Fit

This example fits measured data using a rational model. The data describes the coefficient of thermal expansion for copper as a function of temperature in degrees kelvin.

For this data set, you will find the rational equation that produces the best fit. Rational models are defined as a ratio of polynomials as given by:

$$y = \frac{p_1 x^n + p_2 x^{n-1} + \ldots + p_{n+1}}{x^m + q_1 x^{m-1} + \ldots + q_m}$$

where n is the degree of the numerator polynomial and m is the degree of the denominator polynomial. Note that the rational equations are not associated with physical parameters of the data. Instead, they provide a simple and flexible empirical model that you can use for interpolation and extrapolation.

1 Load the thermal expansion data from the file `hahn1.mat`, which is provided with the toolbox.

```
load hahn1
```

The workspace contains two new variables:

- `temp` is a vector of temperatures in degrees kelvin.

- `thermex` is a vector of thermal expansion coefficients for copper.
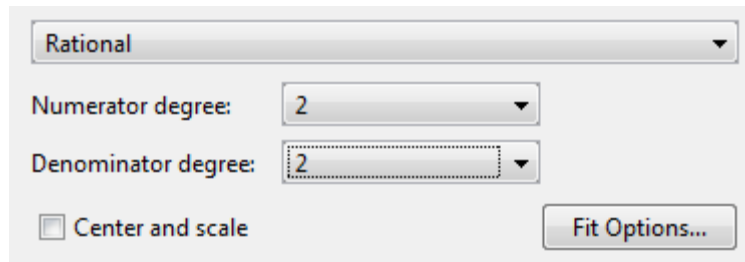
2 Open the Curve Fitting Tool by entering:

```
cftool
```

3 Select `temp` and `thermex` from the **X data** and **Y data** lists.

The Curve Fitting Tool fits and plots the data.

4 Select `Rational` in the fit category list.

5 Try an initial choice for the rational model of quadratic/quadratic. Select 2 for both **Numerator degree** and **Denominator degree**.

The Curve Fitting Tool fits a quadratic/quadratic rational.

6 Examine the residuals. Select **View > Residuals Plot** or click the toolbar button.

Examine the data, fit, and residuals. Observe that the fit misses the data for the smallest and largest predictor values. Additionally, the residuals show a strong pattern throughout the entire data set, indicating that a better fit is possible.
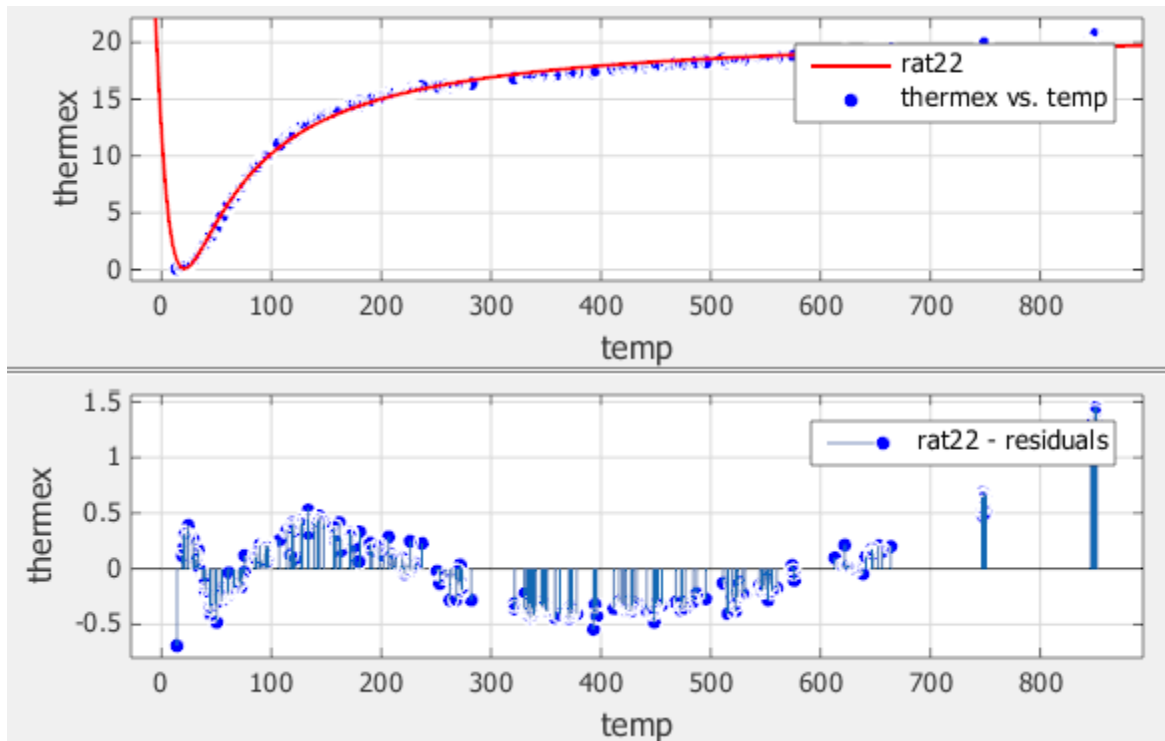
7 For the next fit, try a cubic/cubic equation. Select **3** for both **Numerator degree** and **Denominator degree**.

Examine the data, fit, and residuals. The fit exhibits several discontinuities around the zeros of the denominator.

Note Your results depend on random start points and may vary from those shown.
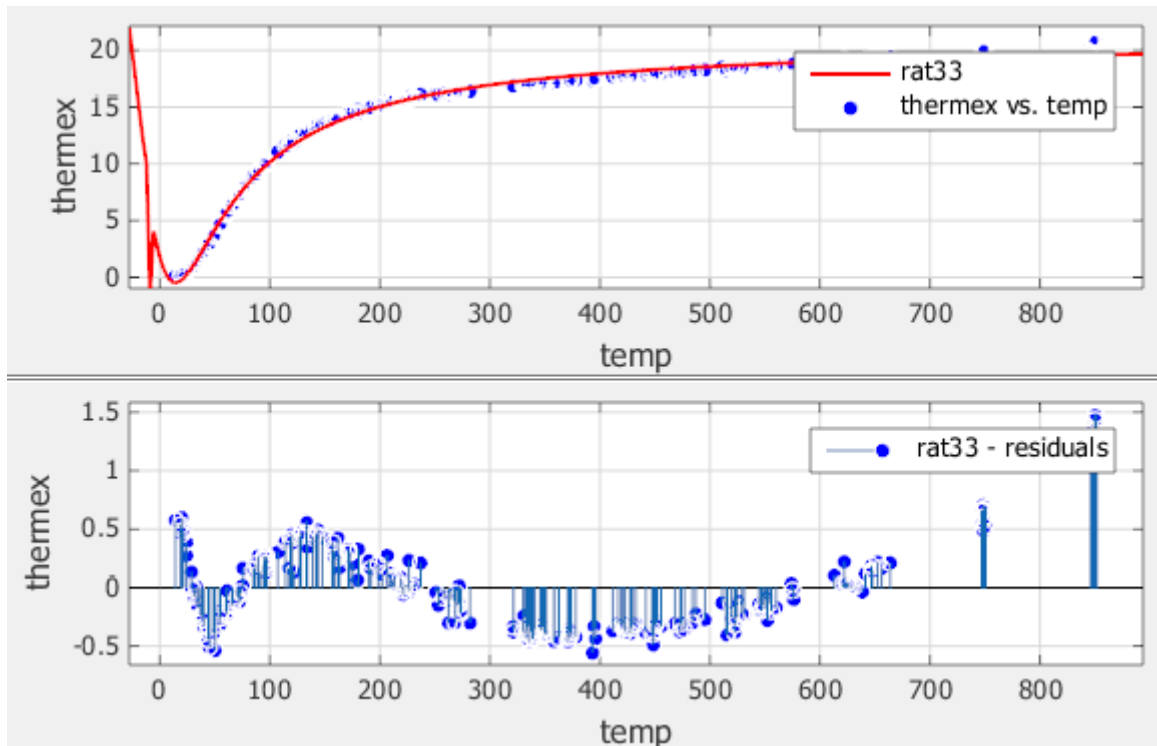
8 Look in the **Results** pane. The message and numerical results indicate that the fit did not converge.

```
Fit computation did not converge:
Fitting stopped because the number of iterations
 or function evaluations exceeded the specified maximum.
```

Although the message in the **Results** pane indicates that you might improve the fit if you increase the maximum number of iterations, a better choice at this stage of the fitting process is to use a different rational equation because the current fit contains several discontinuities. These discontinuities are due to the function blowing up at predictor values that correspond to the zeros of the denominator.

9 Try fitting the data using a cubic/quadratic equation. Select 2 for the **Denominator degree** and leave the **Numerator degree** set to 3.

10 The input variables have very different scales, so select the **Center and scale** option.

The data, fit, and residuals are shown below.
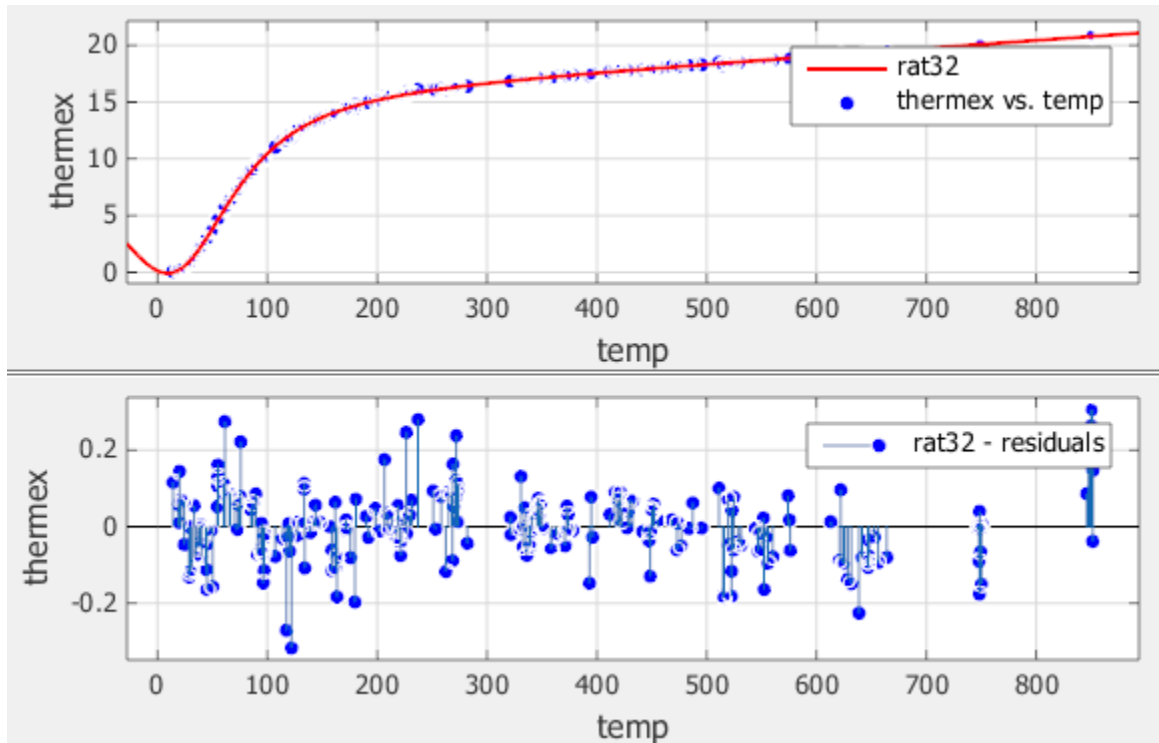
The fit is well behaved over the entire data range, and the residuals are randomly scattered about zero. Therefore, you can confidently use this fit for further analysis.

Sum of Sines Models

In this section...
"About Sum of Sines Models" on page 4-42
"Selecting a Sum of Sine Fit Interactively" on page 4-42
"Selecting a Sum of Sine Fit at the Command Line" on page 4-43

About Sum of Sines Models

The sum of sines model fits periodic functions, and is given by

$$y = \sum_{i=1}^{n} a_i \sin(b_i x + c_i)$$

where a is the amplitude, b is the frequency, and c is the phase constant for each sine wave term. n is the number of terms in the series and $1 \le n \le 8$. This equation is closely related to the Fourier series described in "Fourier Series" on page 4-29. The main difference is that the sum of sines equation includes the phase constant, and does not include a constant (intercept) term.

Selecting a Sum of Sine Fit Interactively

In the Curve Fitting Tool, select Sum of Sine from the model type list.
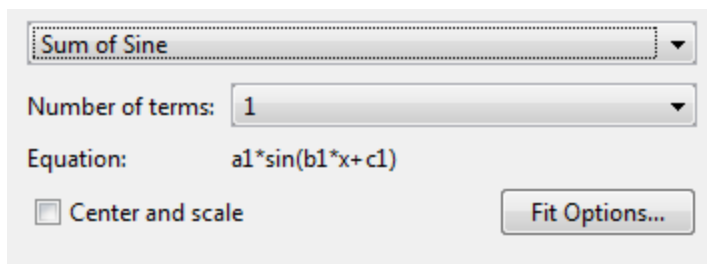
You can specify the following options:

- Choose the number of terms: 1 to 8.

Look in the **Results** pane to see the model terms, the values of the coefficients, and the goodness-of-fit statistics.

- (Optional) Click **Fit Options** to specify coefficient starting values and constraint bounds, or change algorithm settings.

  The toolbox calculates optimized start points for sum of sine models, based on the current data set. You can override the start points and specify your own values in the Fit Options dialog box.

  The sum of sine model has a lower bound constraint on c_i of `0`. The default lower bounds for most library models are `-Inf`.

  For more information on the settings, see "Specifying Fit Options and Optimized Starting Points" on page 4-7.

Selecting a Sum of Sine Fit at the Command Line

Specify the model type `sin` followed by the number of terms, e.g., `'sin1'` to `'sin8'`.

For example, to load some periodic data and fit a six-term sum of sine model:

```
load enso;
f = fit( month, pressure, 'sin6')
plot(f,month,pressure)
```

If you want to modify fit options such as coefficient starting values and constraint bounds appropriate for your data, or change algorithm settings, see the table of additional properties with `NonlinearLeastSquares` on the `fitoptions` reference page.

Weibull Distributions

About Weibull Distribution Models

The Weibull distribution is widely used in reliability and life (failure rate) data analysis. The toolbox provides the two-parameter Weibull distribution

$$y = abx^{b-1}e^{-ax^b}$$

where a is the scale parameter and b is the shape parameter.

Note that there are other Weibull distributions but you must create a custom equation to use these distributions:

- A three-parameter Weibull distribution with x replaced by $x - c$ where c is the location parameter
- A one-parameter Weibull distribution where the shape parameter is fixed and only the scale parameter is fitted.

Curve Fitting Toolbox does not fit Weibull probability distributions to a sample of data. Instead, it fits curves to response and predictor data such that the curve has the same shape as a Weibull distribution.

Selecting a Weibull Fit Interactively

In the Curve Fitting Tool, select Weibull from the model type list.
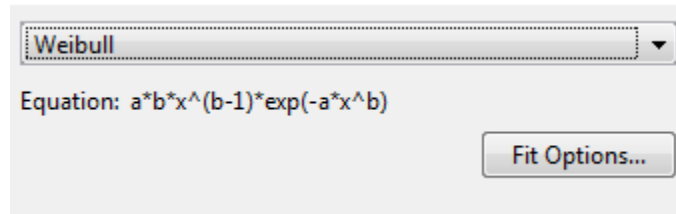
There are no fit settings to configure.

(Optional) Click **Fit Options** to specify coefficient starting values and constraint bounds, or change algorithm settings.

The toolbox calculates random start points for Weibull models, defined on the interval [0,1]. You can override the start points and specify your own values in the Fit Options dialog box.

For more information on the settings, see "Specifying Fit Options and Optimized Starting Points" on page 4-7.

Selecting a Weibull Fit at the Command Line

Specify the model type `weibull`.

For example, to load some example data measuring blood concentration of a compound against time, and fit and plot a Weibull model specifying a start point:

```
time = [ 0.1;   0.1;   0.3;   0.3;   1.3;   1.7;   2.1;...
   2.6;   3.9;   3.9; ...
         5.1;   5.6;   6.2;   6.4;   7.7;   8.1;   8.2;...
   8.9;   9.0;   9.5; ...
         9.6;  10.2;  10.3;  10.8;  11.2;  11.2;  11.2;...
  11.7;  12.1;  12.3; ...
        12.3;  13.1;  13.2;  13.4;  13.7;  14.0;  14.3;...
  15.4;  16.1;  16.1; ...
        16.4;  16.4;  16.7;  16.7;  17.5;  17.6;  18.1;...
  18.5;  19.3;  19.7;];
conc = [0.01;  0.08;  0.13;  0.16;  0.55;  0.90;  1.11;...
  1.62;  1.79;  1.59; ...
```

```
        1.83;  1.68;  2.09;  2.17;  2.66;  2.08;  2.26;...
   1.65;  1.70;  2.39; ...
        2.08;  2.02;  1.65;  1.96;  1.91;  1.30;  1.62;...
   1.57;  1.32;  1.56; ...
        1.36;  1.05;  1.29;  1.32;  1.20;  1.10;  0.88;...
   0.63;  0.69;  0.69; ...
        0.49;  0.53;  0.42;  0.48;  0.41;  0.27;  0.36;...
   0.33;  0.17;  0.20;];

f=fit(time, conc/25, 'Weibull', ...
'StartPoint', [0.01, 2] )
plot(f,time,conc/25, 'o');
```

If you want to modify fit options such as coefficient starting values and constraint bounds appropriate for your data, or change algorithm settings, see the table of additional properties with `NonlinearLeastSquares` on the `fitoptions` reference page.

Appropriate start point values and scaling `conc/25` for the two-parameter Weibull model were calculated by fitting a 3 parameter Weibull model using this custom equation:

```
f=fit(time, conc, ' c*a*b*x^(b-1)*exp(-a*x^b)', 'StartPoint', [0.01, 2, 5]

f =
     General model:
     f(x) = c*a*b*x^(b-1)*exp(-a*x^b)
     Coefficients (with 95% confidence bounds):
       a =    0.009854  (0.007465, 0.01224)
       b =        2.003  (1.895, 2.11)
       c =        25.65  (24.42, 26.89)
```

This Weibull model is defined with three parameters: the first scales the curve along the horizontal axis, the second defines the shape of the curve, and the third scales the curve along the vertical axis. Notice that while this curve has almost the same form as the Weibull probability density function, it is not a density because it includes the parameter c, which is necessary to allow the curve's height to adjust to data. This Weibull model is based on a Statistics Toolbox example, Curve Fitting and Distribution Fitting.

Least-Squares Fitting

In this section...

Introduction

Curve Fitting Toolbox software uses the method of least squares when fitting data. Fitting requires a parametric model that relates the response data to the predictor data with one or more coefficients. The result of the fitting process is an estimate of the model coefficients.

To obtain the coefficient estimates, the least-squares method minimizes the summed square of residuals. The residual for the ith data point r_i is defined as the difference between the observed response value y_i and the fitted response value $\hat{y}_i$, and is identified as the error associated with the data.

$$r_i = y_i - \hat{y}_i$$
$$\text{residual} = \text{data} - \text{fit}$$

The summed square of residuals is given by

$$S = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

where n is the number of data points included in the fit and S is the sum of squares error estimate. The supported types of least-squares fitting include:

- Linear least squares

- Weighted linear least squares

- Robust least squares

- Nonlinear least squares

Error Distributions

When fitting data that contains random variations, there are two important assumptions that are usually made about the error:

- The error exists only in the response data, and not in the predictor data.

- The errors are random and follow a normal (Gaussian) distribution with zero mean and constant variance, σ^2.

The second assumption is often expressed as

$$error \sim N\left(0, \sigma^2\right)$$

The errors are assumed to be normally distributed because the normal distribution often provides an adequate approximation to the distribution of many measured quantities. Although the least-squares fitting method does not assume normally distributed errors when calculating parameter estimates, the method works best for data that does not contain a large number of random errors with extreme values. The normal distribution is one of the probability distributions in which extreme random errors are uncommon. However, statistical results such as confidence and prediction bounds do require normally distributed errors for their validity.

If the mean of the errors is zero, then the errors are purely random. If the mean is not zero, then it might be that the model is not the right choice for your data, or the errors are not purely random and contain systematic errors.

A constant variance in the data implies that the "spread" of errors is constant. Data that has the same variance is sometimes said to be of *equal quality*.

The assumption that the random errors have constant variance is not implicit to weighted least-squares regression. Instead, it is assumed that the weights provided in the fitting procedure correctly indicate the differing levels of

quality present in the data. The weights are then used to adjust the amount of influence each data point has on the estimates of the fitted coefficients to an appropriate level.

Linear Least Squares

Curve Fitting Toolbox software uses the linear least-squares method to fit a linear model to data. A *linear* model is defined as an equation that is linear in the coefficients. For example, polynomials are linear but Gaussians are not. To illustrate the linear least-squares fitting process, suppose you have n data points that can be modeled by a first-degree polynomial.

$$y = p_1 x + p_2$$

To solve this equation for the unknown coefficients p_1 *and* p_2, you write S as a system of n simultaneous linear equations in two unknowns. If n is greater than the number of unknowns, then the system of equations is *overdetermined*.

$$S = \sum_{i=1}^{n} \left(y_i - (p_1 x_i + p_2) \right)^2$$

Because the least-squares fitting process minimizes the summed square of the residuals, the coefficients are determined by differentiating S with respect to each parameter, and setting the result equal to zero.

$$\frac{\partial S}{\partial p_1} = -2 \sum_{i=1}^{n} x_i \left(y_i - (p_1 x_i + p_2) \right) = 0$$

$$\frac{\partial S}{\partial p_2} = -2 \sum_{i=1}^{n} \left(y_i - (p_1 x_i + p_2) \right) = 0$$

The estimates of the true parameters are usually represented by b. Substituting b_1 and b_2 for p_1 and p_2, the previous equations become

$$\sum x_i \left(y_i - (b_1 x_i + b_2) \right) = 0$$
$$\sum \left(y_i - (b_1 x_i + b_2) \right) = 0$$

where the summations run from $i = 1$ to n. The *normal equations* are defined as

$$b_1 \sum x_i^2 + b_2 \sum x_i = \sum x_i y_i$$
$$b_1 \sum x_i + n b_2 = \sum y_i$$

Solving for b_1

$$b_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \left(\sum x_i \right)^2}$$

Solving for b_2 using the b_1 value

$$b_2 = \frac{1}{n} \left(\sum y_i - b_1 \sum x_i \right)$$

As you can see, estimating the coefficients p_1 and p_2 requires only a few simple calculations. Extending this example to a higher degree polynomial is straightforward although a bit tedious. All that is required is an additional normal equation for each linear term added to the model.

In matrix form, linear models are given by the formula

$$y = X\beta + \varepsilon$$

where

- y is an n-by-1 vector of responses.
- β is a m-by-1 vector of coefficients.
- X is the n-by-m design matrix for the model.
- ε is an n-by-1 vector of errors.

For the first-degree polynomial, the n equations in two unknowns are expressed in terms of y, X, and β as

$$
\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 1 \\ x_2 1 \\ x_3 1 \\ . \\ . \\ . \\ x_n 1 \end{bmatrix} \times \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}
$$

The least-squares solution to the problem is a vector b, which estimates the unknown vector of coefficients β. The normal equations are given by

$$(X^T X)b = X^T y$$

where X^T is the transpose of the design matrix X. Solving for b,

$$b = (X^T X)^{-1} X^T y$$

Use the MATLAB backslash operator (`mldivide`) to solve a system of simultaneous linear equations for unknown coefficients. Because inverting $X^T X$ can lead to unacceptable rounding errors, the backslash operator uses QR decomposition with pivoting, which is a very stable algorithm numerically. Refer to Arithmetic Operators in the MATLAB documentation for more information about the backslash operator and QR decomposition.

You can plug b back into the model formula to get the predicted response values, $\hat{y}$.

$$\hat{y} = Xb = Hy$$

$$H = X(X^T X)^{-1} X^T$$

A hat (circumflex) over a letter denotes an estimate of a parameter or a prediction from a model. The projection matrix H is called the hat matrix, because it puts the hat on y.

The residuals are given by

$$r = y - \hat{y} = (1 - H)y$$

Weighted Least Squares

It is usually assumed that the response data is of equal quality and, therefore, has constant variance. If this assumption is violated, your fit might be unduly influenced by data of poor quality. To improve the fit, you can use weighted least-squares regression where an additional scale factor (the weight) is included in the fitting process. Weighted least-squares regression minimizes the error estimate

$$s = \sum_{i=1}^{n} w_i \left(y_i - \hat{y}_i \right)^2$$

where w_i are the weights. The weights determine how much each response value influences the final parameter estimates. A high-quality data point influences the fit more than a low-quality data point. Weighting your data is recommended if the weights are known, or if there is justification that they follow a particular form.

The weights modify the expression for the parameter estimates b in the following way,

$$b = \hat{\beta} = (X^T W X)^{-1} X^T W y$$

where W is given by the diagonal elements of the weight matrix w.

You can often determine whether the variances are not constant by fitting the data and plotting the residuals. In the plot shown below, the data contains replicate data of various quality and the fit is assumed to be correct. The poor quality data is revealed in the plot of residuals, which has a "funnel" shape where small predictor values yield a bigger scatter in the response values than large predictor values.
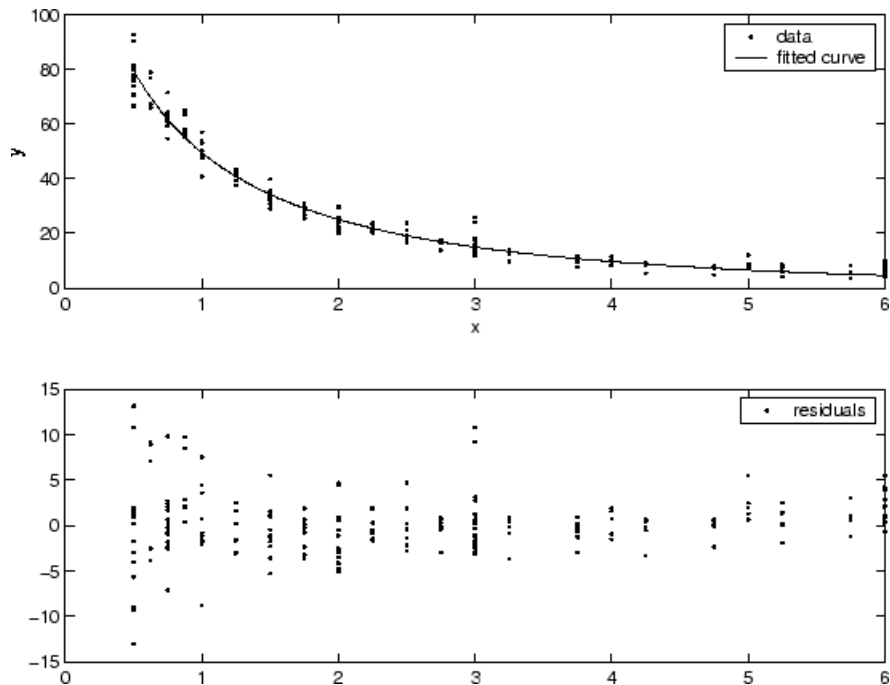
The weights you supply should transform the response variances to a constant value. If you know the variances of the measurement errors in your data, then the weights are given by

$$w_i = 1/\sigma_i^2$$

Or, if you only have estimates of the error variable for each data point, it usually suffices to use those estimates in place of the true variance. If you do not know the variances, it suffices to specify weights on a relative scale. Note that an overall variance term is estimated even when weights have been specified. In this instance, the weights define the relative weight to each point in the fit, but are not taken to specify the exact variance of each point.

For example, if each data point is the mean of several independent measurements, it might make sense to use those numbers of measurements as weights.

Robust Least Squares

It is usually assumed that the response errors follow a normal distribution, and that extreme values are rare. Still, extreme values called *outliers* do occur.

The main disadvantage of least-squares fitting is its sensitivity to outliers. Outliers have a large influence on the fit because squaring the residuals magnifies the effects of these extreme data points. To minimize the influence of outliers, you can fit your data using robust least-squares regression. The toolbox provides these two robust regression methods:

- Least absolute residuals (LAR) — The LAR method finds a curve that minimizes the absolute difference of the residuals, rather than the squared differences. Therefore, extreme values have a lesser influence on the fit.

- Bisquare weights — This method minimizes a weighted sum of squares, where the weight given to each data point depends on how far the point is from the fitted line. Points near the line get full weight. Points farther from the line get reduced weight. Points that are farther from the line than would be expected by random chance get zero weight.

  For most cases, the bisquare weight method is preferred over LAR because it simultaneously seeks to find a curve that fits the bulk of the data using the usual least-squares approach, and it minimizes the effect of outliers.

Robust fitting with bisquare weights uses an iteratively reweighted least-squares algorithm, and follows this procedure:

1 Fit the model by weighted least squares.

2 Compute the *adjusted residuals* and standardize them. The adjusted residuals are given by

$$r_{adj} = \frac{r_i}{\sqrt{1 - h_i}}$$

r_i are the usual least-squares residuals and h_i are *leverages* that adjust the residuals by reducing the weight of high-leverage data points, which have a large effect on the least-squares fit. The standardized adjusted residuals are given by

$$u = \frac{r_{adj}}{Ks}$$

K is a tuning constant equal to 4.685, and *s* is the robust variance given by *MAD*/0.6745 where *MAD* is the median absolute deviation of the residuals.

3 Compute the robust weights as a function of *u*. The bisquare weights are given by

$$w_i = \begin{cases} \left(1 - (u_i)^2\right)^2 & |u_i| < 1 \\ 0 & |u_i| \geq 1 \end{cases}$$

Note that if you supply your own regression weight vector, the final weight is the product of the robust weight and the regression weight.

4 If the fit converges, then you are done. Otherwise, perform the next iteration of the fitting procedure by returning to the first step.

The plot shown below compares a regular linear fit with a robust fit using bisquare weights. Notice that the robust fit follows the bulk of the data and is not strongly influenced by the outliers.
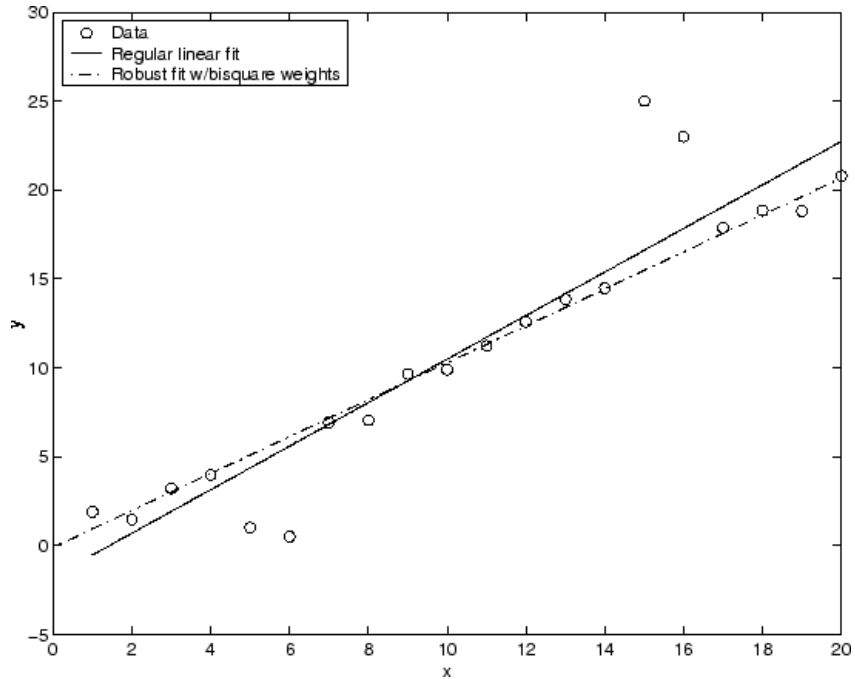
Instead of minimizing the effects of outliers by using robust regression, you can mark data points to be excluded from the fit. Refer to "Removing Outliers" on page 7-10 for more information.

Nonlinear Least Squares

Curve Fitting Toolbox software uses the nonlinear least-squares formulation to fit a nonlinear model to data. A nonlinear model is defined as an equation that is nonlinear in the coefficients, or a combination of linear and nonlinear in the coefficients. For example, Gaussians, ratios of polynomials, and power functions are all nonlinear.

In matrix form, nonlinear models are given by the formula

$$y = f(X, \beta) + \varepsilon$$

where

- *y* is an *n*-by-1 vector of responses.

- *f* is a function of ß and *X*.

- ß is a *m*-by-1 vector of coefficients.

- *X* is the *n*-by-*m* design matrix for the model.

- ε is an *n*-by-1 vector of errors.

Nonlinear models are more difficult to fit than linear models because the coefficients cannot be estimated using simple matrix techniques. Instead, an iterative approach is required that follows these steps:

1 Start with an initial estimate for each coefficient. For some nonlinear models, a heuristic approach is provided that produces reasonable starting values. For other models, random values on the interval [0,1] are provided.

2 Produce the fitted curve for the current set of coefficients. The fitted response value $\hat{y}$ is given by

$\hat{y} = f(X,b)$

and involves the calculation of the *Jacobian* of *f*(*X*,*b*), which is defined as a matrix of partial derivatives taken with respect to the coefficients.

3 Adjust the coefficients and determine whether the fit improves. The direction and magnitude of the adjustment depend on the fitting algorithm. The toolbox provides these algorithms:

- Trust-region — This is the default algorithm and must be used if you specify coefficient constraints. It can solve difficult nonlinear problems more efficiently than the other algorithms and it represents an improvement over the popular Levenberg-Marquardt algorithm.

- Levenberg-Marquardt — This algorithm has been used for many years and has proved to work most of the time for a wide range of nonlinear models and starting values. If the trust-region algorithm does not produce a reasonable fit, and you do not have coefficient constraints, you should try the Levenberg-Marquardt algorithm.

4 Iterate the process by returning to step 2 until the fit reaches the specified convergence criteria.

You can use weights and robust fitting for nonlinear models, and the fitting process is modified accordingly.

Because of the nature of the approximation process, no algorithm is foolproof for all nonlinear models, data sets, and starting points. Therefore, if you do not achieve a reasonable fit using the default starting points, algorithm, and convergence criteria, you should experiment with different options. Refer to "Specifying Fit Options and Optimized Starting Points" on page 4-7 for a description of how to modify the default options. Because nonlinear models can be particularly sensitive to the starting points, this should be the first fit option you modify.

Example: Robust Fitting

1 Create a baseline sinusoidal signal:

```
xdata = (0:0.1:2*pi)';
y0 = sin(xdata);
```

2 Add noise to the signal with non-constant variance:

```
% Response-dependent Gaussian noise
gnoise = y0.*randn(size(y0));

% Salt-and-pepper noise
spnoise = zeros(size(y0));
p = randperm(length(y0));
sppoints = p(1:round(length(p)/5));
spnoise(sppoints) = 5*sign(y0(sppoints));

ydata = y0 + gnoise + spnoise;
```

3 Fit the noisy data with a baseline sinusoidal model:

```
f = fittype('a*sin(b*x)');
fit1 = fit(xdata,ydata,f,'StartPoint',[1 1]);
```

4 Identify "outliers" as points at a distance greater than 1.5 standard deviations from the baseline model, and refit the data with the outliers excluded:

```
fdata = feval(fit1,xdata);
I = abs(fdata - ydata) > 1.5*std(ydata);
outliers = excludedata(xdata,ydata,'indices',I);

fit2 = fit(xdata,ydata,f,'StartPoint',[1 1],'Exclude',outliers);
```

5 Compare the effect of excluding the outliers with the effect of giving them lower bisquare weight in a robust fit:

```
fit3 = fit(xdata,ydata,f,'StartPoint',[1 1],'Robust','on');
```

6 Plot the data, the outliers, and the results of the fits:

```
plot(fit1,'r-',xdata,ydata,'k.',outliers,'m*')
hold on
plot(fit2,'c--')
plot(fit3,'b:')
xlim([O 2*pi])
```


7 Plot the residuals for the two fits considering outliers:

```
figure
plot(fit2,xdata,ydata,'co','residuals')
```

```
hold on
plot(fit3,xdata,ydata,'bx','residuals')
```

5
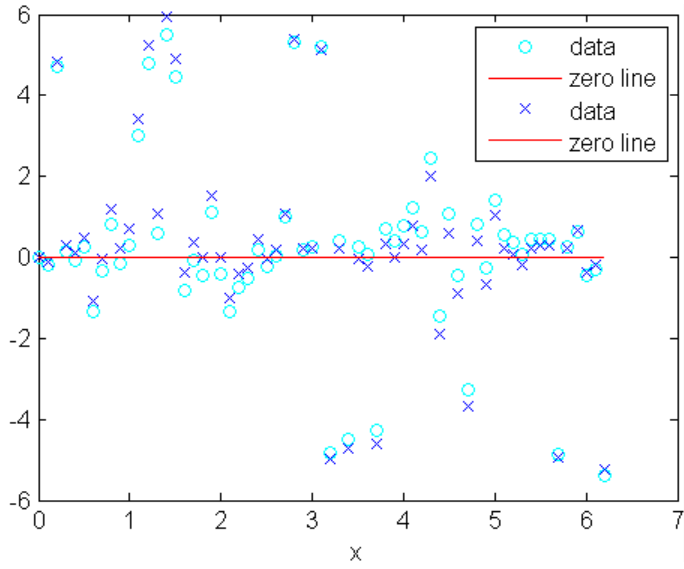
Custom Linear and Nonlinear Regression

Custom Models

Custom Models vs. Library Models

If the toolbox library does not contain a desired parametric equation, you can create your own custom equation. Library models, however, offer the best chance for rapid convergence. This is because:

- For most library models, the toolbox calculates optimal default coefficient starting points. For custom models, the toolbox chooses random default starting points on the interval [0,1].

- Library models use an analytic Jacobian. Custom models use finite differencing.

Selecting a Custom Equation Fit Interactively

In the Curve Fitting Tool, select `Custom Equation` from the model type list.

Use the custom equation fit to define your own equations. An example custom equation appears when you select **Custom Equation** from the list, as shown here for curve data.

If you have surface data, the example custom equation uses both **x** and **y**.

1 You can edit x, y, and z to any valid variable names.

2 In the lower box, edit the example to define your own custom equation. You can enter any valid MATLAB expression in terms of your variable names.

You can create custom general equations. General models are nonlinear combinations of (perhaps nonlinear) terms. They are defined by equations that might be nonlinear in the parameters. The custom equation fit uses the nonlinear least-squares fitting procedure.

You can define a custom linear equation, but the nonlinear fitting is less efficient and usually slower than linear least-squares fitting. If you need linear least-squares fitting for custom equations, you must use the legacy Curve Fitting Tool. See "Creating Custom Models Using the Legacy Curve Fitting Tool" on page 5-46.

You can save your custom equations as part of your saved Curve Fitting Tool sessions.

Your function can execute a number of times, both during fitting and during preprocessing before fitting. Be aware this may be time-consuming if you are using functions with side effects such as writing data to a file, or displaying diagnostic information to the Command Window.

For examples, see:

• "Fourier Analysis of ENSO Data" on page 5-9

Selecting a Custom Equation Fit at the Command Line

To fit custom models, either:

- Use an anonymous function with the `fit` function.

- Create a `fittype` object with the `fittype` function to use as an input argument for the `fit` function.

This example loads some data and uses a custom equation defining a Weibull model as an input to the `fit` function:

```
time = [ 0.1;   0.1;   0.3;   0.3;   1.3;   1.7;   2.1;   2.6;   3.9;   3.9
         5.1;   5.6;   6.2;   6.4;   7.7;   8.1;   8.2;   8.9;   9.0;   9.5
         9.6;  10.2;  10.3;  10.8;  11.2;  11.2;  11.2;  11.7;  12.1;  12.3
        12.3;  13.1;  13.2;  13.4;  13.7;  14.0;  14.3;  15.4;  16.1;  16.1
        16.4;  16.4;  16.7;  16.7;  17.5;  17.6;  18.1;  18.5;  19.3;  19.7
conc = [0.01;  0.08;  0.13;  0.16;  0.55;  0.90;  1.11;  1.62;  1.79;  1.59
        1.83;  1.68;  2.09;  2.17;  2.66;  2.08;  2.26;  1.65;  1.70;  2.39
        2.08;  2.02;  1.65;  1.96;  1.91;  1.30;  1.62;  1.57;  1.32;  1.56
        1.36;  1.05;  1.29;  1.32;  1.20;  1.10;  0.88;  0.63;  0.69;  0.69
        0.49;  0.53;  0.42;  0.48;  0.41;  0.27;  0.36;  0.33;  0.17;  0.20

f = fit( time, conc, 'c*a*b*x^(b-1)*exp(-a*x^b)', 'StartPoint', [0.01, 2, 5
plot( f, time, conc )
```

To define a custom model using `fittype`, use the form:

```
f = fittype(expr)
```

which constructs a custom model `fittype` object for the MATLAB expression contained in the string, cell array, or anonymous function *expr*.

For details on specifying dependent and independent variables, problem parameters, and coefficients using `fittype`, see "Using Custom Models" on page 13-75 on the `fittype` reference page.

If *expr* is a string or anonymous function, then the toolbox uses a nonlinear fitting algorithm. To use a linear fitting algorithm for your custom equation, use a cell array of terms. See "Using Linear Models" on page 13-76 on the `fittype` reference page.

For examples of linear and nonlinear custom models, see the `fittype` reference page. For a step-by-step example, see "Custom Nonlinear Census Analysis" on page 5-6.

Custom Nonlinear Census Analysis

The following code, using Curve Fitting Toolbox functions and methods, reproduces an interactive analysis of the census data in the Curve Fitting Tool in "Curve Fitting to Census Data" on page 2-22.

1 Load and plot the data in `census.mat`:

```
load census
plot(cdate,pop,'o')
hold on
```


2 Create a fit options structure and a `fittype` object for the custom nonlinear model $y = a(x-b)^n$, where a and b are coefficients and n is a problem-dependent parameter:

```
s = fitoptions('Method','NonlinearLeastSquares',...
               'Lower',[0,0],...
               'Upper',[Inf,max(cdate)],...
               'Startpoint',[1 1]);
f = fittype('a*(x-b)^n','problem','n','options',s);
```

3 Fit the data using the fit options and a value of $n = 2$:

```
[c2,gof2] = fit(cdate,pop,f,'problem',2)
c2 =
     General model:
       c2(x) = a*(x-b)^n
     Coefficients (with 95% confidence bounds):
       a =    0.006092  (0.005743, 0.006441)
       b =         1789  (1784, 1793)
     Problem parameters:
       n =            2
gof2 =
           sse: 246.1543
       rsquare: 0.9980
           dfe: 19
     adjrsquare: 0.9979
          rmse: 3.5994
```

4 Fit the data using the fit options and a value of $n = 3$:

```
[c3,gof3] = fit(cdate,pop,f,'problem',3)
c3 =
     General model:
       c3(x) = a*(x-b)^n
     Coefficients (with 95% confidence bounds):
       a =  1.359e-005  (1.245e-005, 1.474e-005)
       b =         1725  (1718, 1731)
     Problem parameters:
       n =            3
gof3 =
           sse: 232.0058
       rsquare: 0.9981
           dfe: 19
     adjrsquare: 0.9980
          rmse: 3.4944
```

5 Plot the fit results and the data:

```
plot(c2,'m')
plot(c3,'c')
```
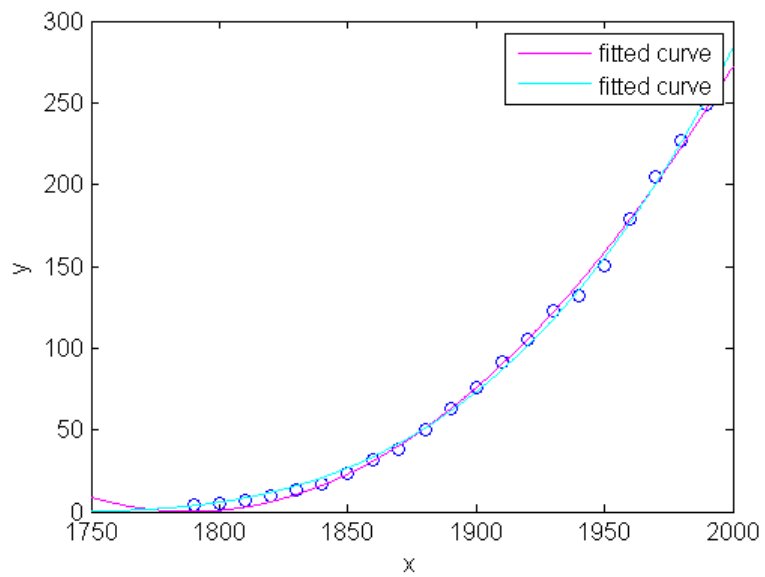
Fourier Analysis of ENSO Data

This example fits the ENSO data using several custom nonlinear equations. The ENSO data consists of monthly averaged atmospheric pressure differences between Easter Island and Darwin, Australia. This difference drives the trade winds in the southern hemisphere.

The ENSO data is clearly periodic, which suggests it can be described by a Fourier series:

$$y(x) = a_0 + \sum_{i=1}^{\infty} a_i \cos\left(2\pi \frac{x}{c_i}\right) + b_i \sin\left(2\pi \frac{x}{c_i}\right)$$

where a_i and b_i are the amplitudes, and c_i are the periods (cycles) of the data. The question to answer here is how many cycles exist?

As a first attempt, assume a single cycle and fit the data using one cosine term and one sine term.

$$y_1(x) = a_0 + a_1 \cos\left(2\pi \frac{x}{c_1}\right) + b_1 \sin\left(2\pi \frac{x}{c_1}\right)$$

If the fit does not describe the data well, add additional cosine and sine terms with unique period coefficients until a good fit is obtained.

The equation is nonlinear because an unknown coefficient c_1 is included as part of the trigonometric function arguments.

Load Data and Fit Library and Custom Fourier Models

1 Load the data and open the Curve Fitting Tool:

```
load enso
cftool
```

2 The toolbox includes the Fourier series as a nonlinear library equation. However, the library equation does not meet the needs of this example because its terms are defined as fixed multiples of the fundamental frequency *w*. Refer to "Fourier Series" on page 4-29 for more information. Create the built-in library Fourier fit to compare with your custom equations:

a Select month for **X data** and pressure for **Y data**.

b Select Fourier for the model type.

c Enter Fourier for the **Fit name**.

d Change the number of terms to 8.

Observe that this library model does not describe the data well enough.
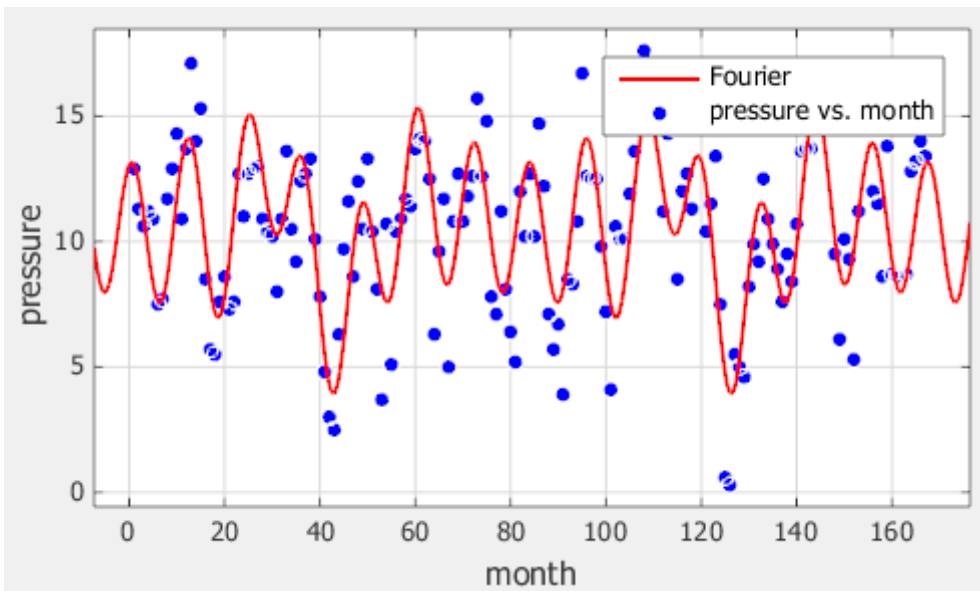
3 Duplicate your fit. Right-click your fit in the **Table of Fits** and select **Duplicate 'Fourier'**.

4 Name the new fit `Enso1Period`.

5 Change the fit type from `Fourier` to `Custom Equation`.

6 Replace the example text in the equation edit box with

```
a0+a1*cos(2*pi*x/c1)+b1*sin(2*pi*x/c1)
```


The toolbox applies the fit to the `enso` data.

The graphical and numerical results shown here indicate that the fit does not describe the data well. In particular, the fitted value for `c1` is unreasonably small. Your initial fit results might differ from these results because the starting points are randomly selected.
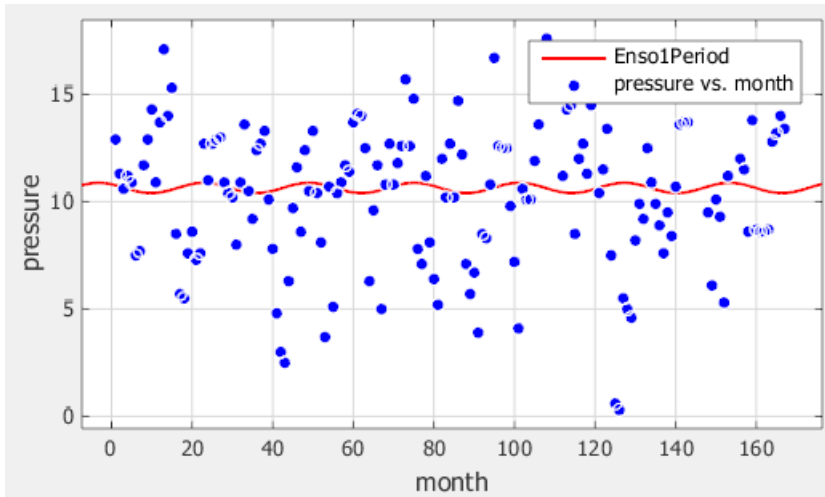
Results

```
General model:
    f(x) = a0+a1*cos(2*pi*x/c1)+b1*sin(2*pi*x/c1)
Coefficients (with 95% confidence bounds):
    a0 =      10.64  (10.11, 11.17)
    a1 =    -0.2215  (-1.111, 0.6678)
    b1 =     0.0915  (-1.302, 1.485)
    c1 =     0.6447  (0.6404, 0.6489)

Goodness of fit:
  SSE: 1958
  R-square: 0.002451
  Adjusted R-square: -0.0158
  RMSE: 3.455
```

By default, the coefficients are unbounded and have random starting values from 0 to 1. The data include a periodic component with a period of about 12 months. However, with `c1` unconstrained and with a random starting point, this fit failed to find that cycle.

Use Fit Options to Constrain a Coefficient

1 To assist the fitting procedure, constrain c1 to a value from 10 to 14. Click the **Fit Options** button to view and edit constraints for unknown coefficients.

2 In the Fit Options dialog box, observe that by default the coefficients are unbounded (bounds of -Inf and Inf).

3 Change the **Lower** and **Upper** bounds for c1 to constrain the cycle from 10 to 14 months, as shown next.
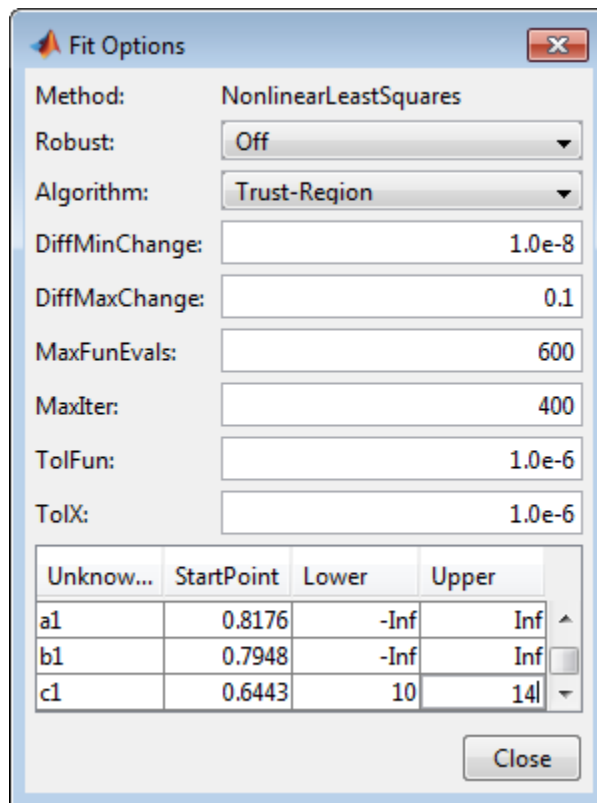
4 Click **Close**. The Curve Fitting Tool refits.

5 Observe the new fit and the residuals plot. If necessary, select **View > Residuals Plot** or use the toolbar button.
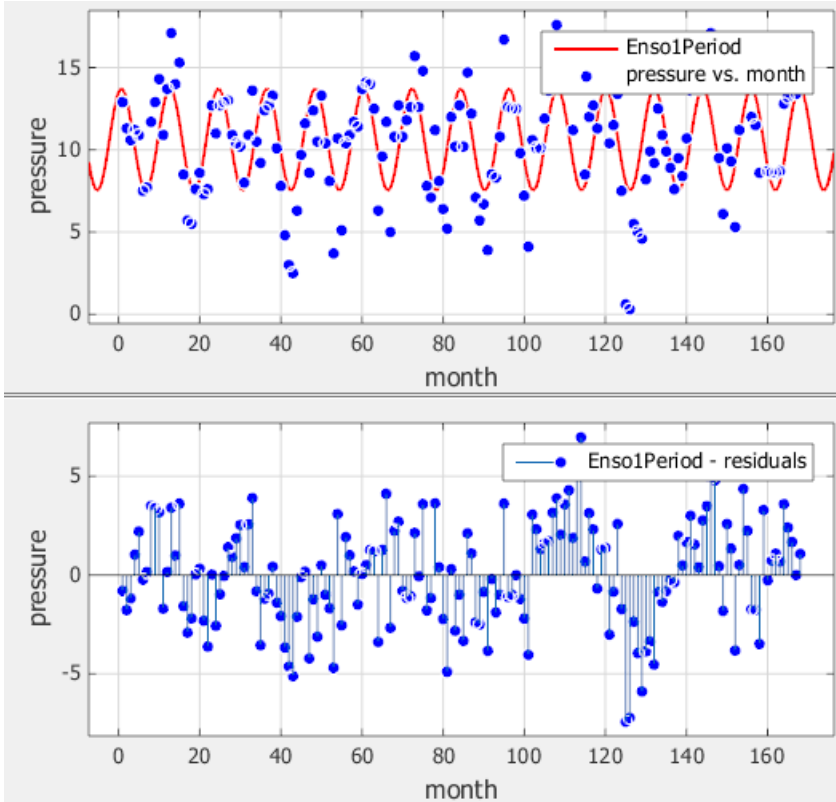
The fit appears to be reasonable for some data points but clearly does not describe the entire data set very well. As predicted, the numerical results in the **Results** pane (c1=11.94) indicate a cycle of approximately 12 months. However, the residuals show a systematic periodic distribution, indicating that at least one more cycle exists. There are additional cycles that you should include in the fit equation.

Create Second Custom Fit with Additional Terms and Constraints

To refine your fit, you need to add an additional sine and cosine term to $y_1(x)$ as follows:

$$y_2(x) = y_1(x) + a_2 \cos\left(2\pi\, \frac{x}{c_2}\right) + b_2 \sin\left(2\pi\, \frac{x}{c_2}\right)$$

and constrain the upper and lower bounds of c_2 to be roughly twice the bounds used for c_1.

1 Duplicate your fit by right-clicking it in the **Table of Fits** and selecting **Duplicate 'Enso1Period'**.

2 Name the new fit `Enso2Period`.

3 Add these terms to the end of the previous equation:

```
+a2*cos(2*pi*x/c2)+b2*sin(2*pi*x/c2)
```

4 Click **Fit Options**. When you edit the custom equation, the tool refits with default fit options. So:

   a Change the **Lower** and **Upper** for `c1` to constrain the cycle from 10 to 14 months.

   b Change the **Lower** and **Upper** for `c2` to be roughly twice the bounds used for `c1` (20<c2<30).

   c Change the **StartPoint** for `a0` to `5`.

When you close the dialog box, the Curve Fitting Tool refits. The fit and residuals are shown next.
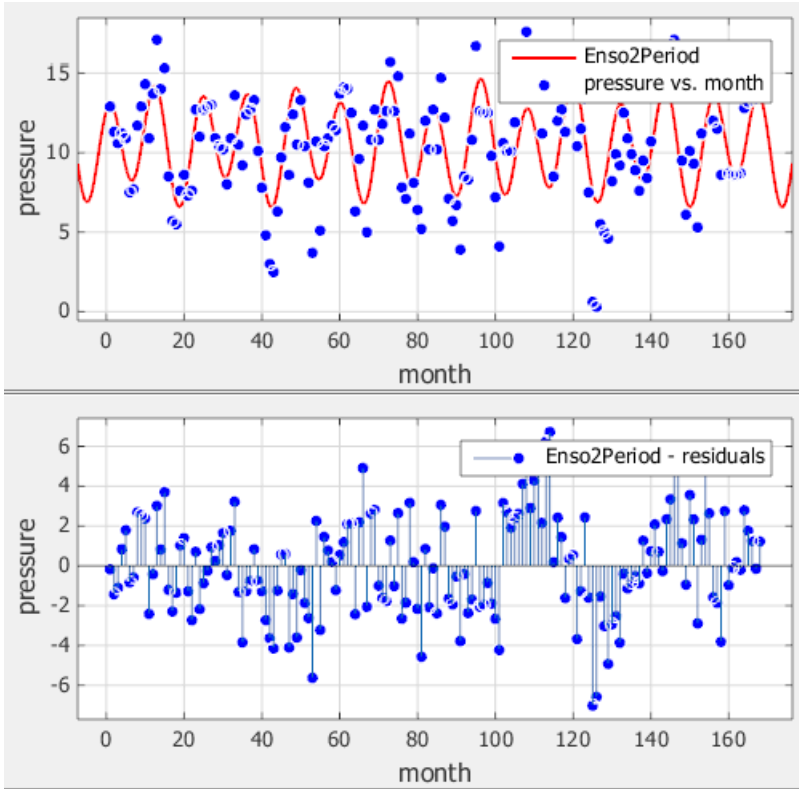
The fit appears reasonable for most data points. However, the residuals indicate that you should include another cycle to the fit equation.

Create a Third Custom Fit with Additional Terms and Constraints

As a third attempt, add an additional sine and cosine term to $y_2(x)$

$$y_3(x) = y_2(x) + a_3 \cos\left(2\pi \frac{x}{c_3}\right) + b_3 \sin\left(2\pi \frac{x}{c_3}\right)$$

and constrain the lower bound of c_3 to be roughly triple the value of c_1.

1 Duplicate your fit by right-clicking it in the **Table of Fits** and selecting **Duplicate 'Enso1Period'**.

2 Name the new fit `Enso3Period`.

3 Add these terms to the end of the previous equation:

```
+a3*cos(2*pi*x/c3)+b3*sin(2*pi*x/c3)
```

4 Click **Fit Options**.

a Change the **Lower** and **Upper** for `c1` to constrain the cycle from 10 to 14 months.

b Change the **Lower** and **Upper** for `c2` to be roughly twice the bounds used for `c1` ($20 < c2 < 30$).

c Change the **Lower** bound for `c3` to be `36`, which is roughly triple the value of `c1`.

d Change the **StartPoint** for `a0` to be `5`.
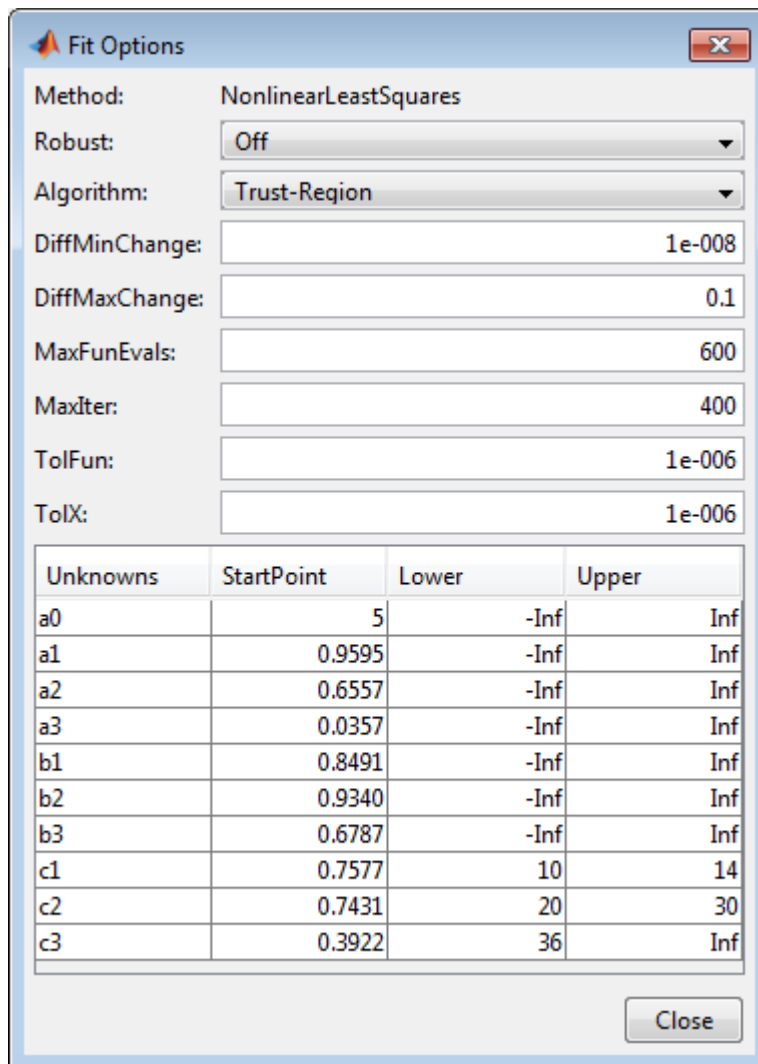
e Close the dialog box. The Curve Fitting Tool refits. The fit and residuals appear next.
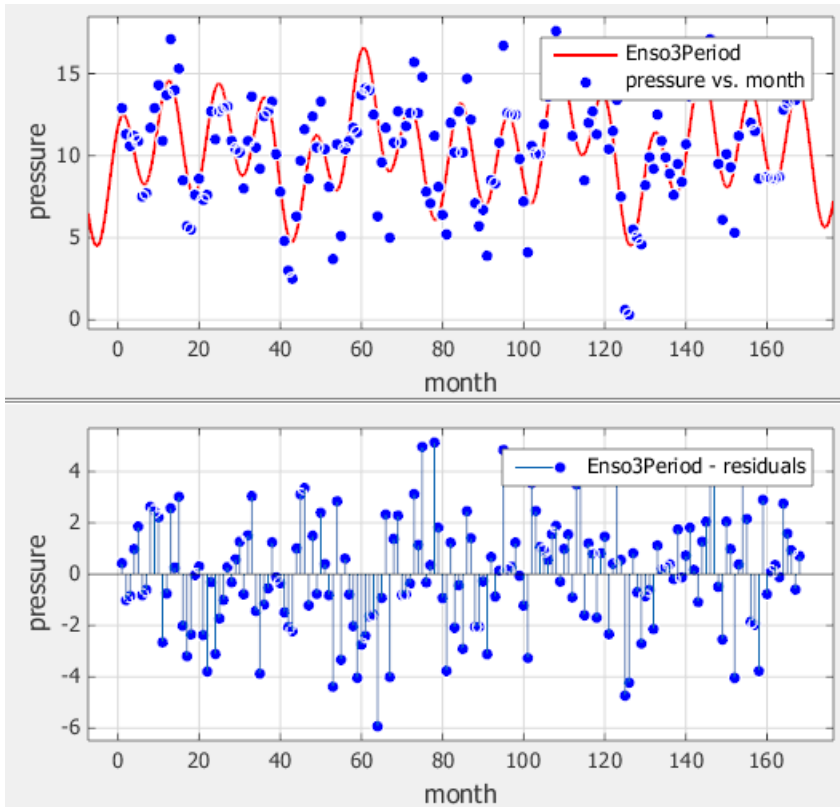
The fit is an improvement over the previous two fits, and appears to account for most of the cycles in the ENSO data set. The residuals appear random for most of the data, although a pattern is still visible indicating that additional cycles might be present, or you can improve the fitted amplitudes.

In conclusion, Fourier analysis of the data reveals three significant cycles. The annual cycle is the strongest, but cycles with periods of approximately 44 and 22 months are also present. These cycles correspond to El Nino and the Southern Oscillation (ENSO).

Gaussian Fitting with an Exponential Background

This example fits two poorly resolved Gaussian peaks on a decaying exponential background using a general (nonlinear) custom model.

Fit the data using this equation

$$y(x) = ae^{-bx} + a_1 e^{-\left(\frac{x-b_1}{c_1}\right)^2} + a_2 e^{-\left(\frac{x-b_2}{c_2}\right)^2}$$

where a_i are the peak amplitudes, b_i are the peak centroids, and c_i are related to the peak widths. Because unknown coefficients are part of the exponential function arguments, the equation is nonlinear.

1 Load the data and open the Curve Fitting Tool:

```
load gauss3
cftool
```

The workspace contains two new variables:

- xpeak is a vector of predictor values.
- ypeak is a vector of response values.

2 In the Curve Fitting Tool, select xpeak for **X data** and ypeak for **Y data**.

3 Enter Gauss2exp1 for the **Fit name**.

4 Select Custom Equation for the model type.

5 Replace the example text in the equation edit box with these terms:

```
a*exp(-b*x)+a1*exp(-((x-b1)/c1)^2)+a2*exp(-((x-b2)/c2)^2)
```

The fit is poor (or incomplete) at this point because the starting points are randomly selected and no coefficients have bounds.

6 Specify reasonable coefficient starting points and constraints. Deducing the starting points is particularly easy for the current model because the Gaussian coefficients have a straightforward interpretation and

the exponential background is well defined. Additionally, as the peak amplitudes and widths cannot be negative, constrain a_1, a_2, c_1, and c_2 to be greater then 0.

a Click **Fit Options**.

b Change the **Lower** bound for a_1, a_2, c_1, and c_2 to **0**, as the peak amplitudes and widths cannot be negative.

c Enter start points as shown for the unknown coefficients.

Unknowns	Start Point
a	100
a1	100
a2	80
b	0.1
b1	110
b2	140
c1	20
c2	20
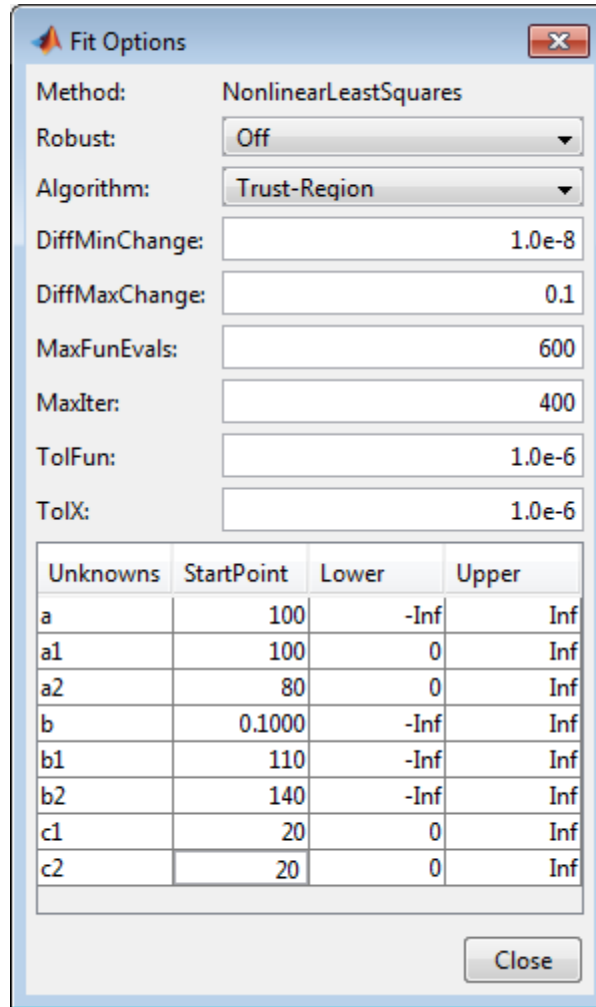
When you close the dialog box, the Curve Fitting Tool refits. Following are the fit and residuals.
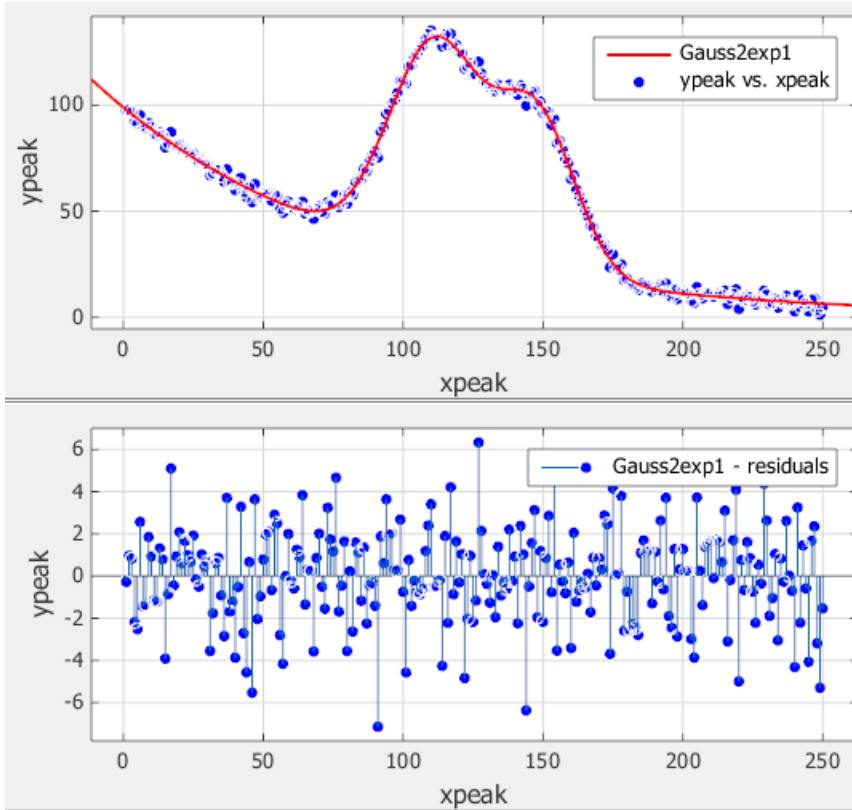
Biopharmaceutical Drug Interaction Surface Fitting

Surface Fitting to Biopharmaceutical Data

Curve Fitting Toolbox software provides some example data for an anesthesia drug interaction study. You can use Curve Fitting Tool to fit response surfaces to this data to analyze drug interaction effects. Response surface models provide a good method for understanding the pharmacodynamic interaction behavior of drug combinations.

This data is based on the results in this paper:

- Kern SE, Xie G, White JL, Egan TD. Opioid-hypnotic synergy: A response surface analysis of propofol-remifentanil pharmacodynamic interaction in volunteers. Anesthesiology 2004; 100: 1373–81.

Anesthesia is typically at least a two-drug process, consisting of an opioid and a sedative hypnotic. This example uses Propofol and Reminfentanil as drug class prototypes. Their interaction is measured by four different measures of the analgesic and sedative response to the drug combination. Algometry, Tetany, Sedation, and Laryingoscopy comprise the four measures of surrogate drug effects at various concentration combinations of Propofol and Reminfentanil.

To interactively create response surfaces for this drug combination:

1 Use the Current Folder browser to locate and view the folder `matlab\toolbox\curvefit\curvefit`.

2 Right-click the file `OpioidHypnoticSynergy.txt`, and select **Import Data**. The Import Wizard appears.

   a Leave the default **Column delimiters** set to **Tab** and **Column vectors** in the Imported Data section.

Review the six variables selected for import: Propofol, Reminfentanil, Algometry, Tetany, Sedation, and Laryingoscopy.

b On the Import tab, in the Import section, click **Import Selection** to import the dose-response data into the MATLAB workspace.

Alternatively, you can import the data programmatically. Enter the following code to read the dose-response data from the file into the MATLAB workspace.

```
data = importdata( 'OpioidHypnoticSynergy.txt' );
Propofol      = data.data(:,1);
Remifentanil  = data.data(:,2);
Algometry     = data.data(:,3);
Tetany        = data.data(:,4);
Sedation      = data.data(:,5);
Laryingoscopy = data.data(:,6);
```

3 To create response surfaces you must select the two drugs for the X and Y inputs, and one of the four effects for the Z output. After you load the variables into your workspace, you can either open the tool and select variables interactively, or specify the initial fit variables with the cftool command.

Enter the following to open Curve Fitting Tool (if necessary) and create a new response surface for Algometry:

```
cftool(Propofol, Remifentanil, Algometry)
```

Review the Curve Fitting Tool X, Y, and Z input and output controls. The tool displays the selected variables Propofol, Remifentanil and Algometry, with a surface fit. The default fit is an interpolating surface that passes through the data points.
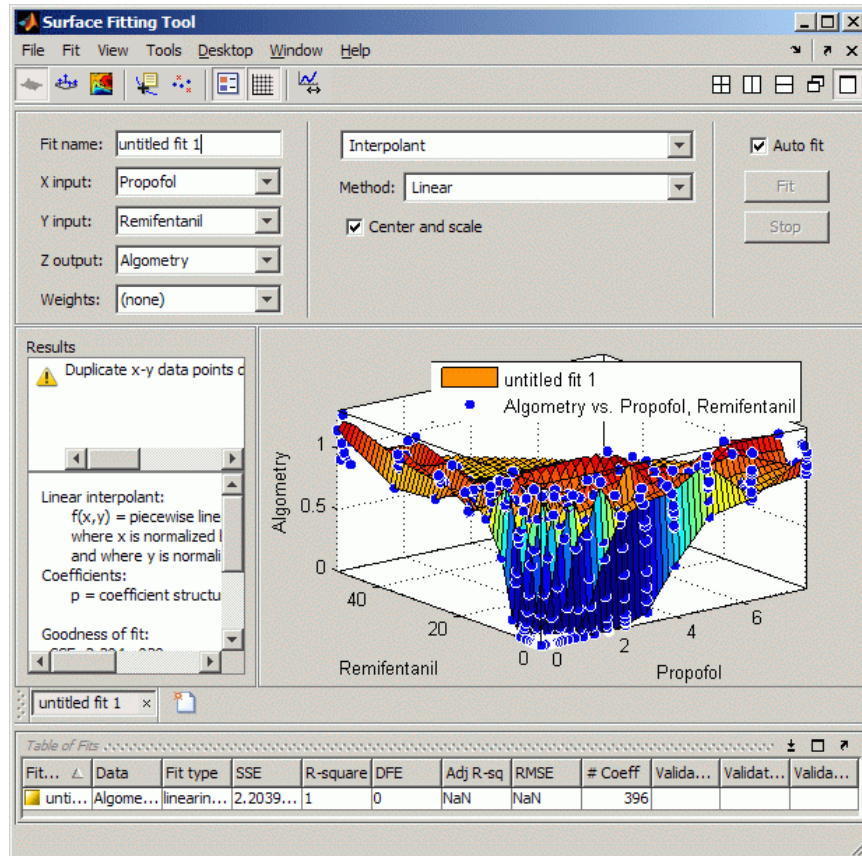
4 Create a copy of the current surface fit by either:

  a Selecting **Fit > Duplicate "*Current Fit Name*"**.

  b Right-clicking a fit in the **Table of Fits**, and selecting **Duplicate**.

5 Select the **Custom Equation** fit type from the drop-down list to define your own equation to fit the data.

6 Select and delete the example custom equation text in the edit box.

  You can use the custom equation edit box to enter MATLAB code to define your model. The equation that defines the model must depend on the input

variables x and y and a list of fixed parameters, estimable parameters, or both.

The model from the paper is:

$$E = \frac{E_{\max} \cdot \left(\dfrac{C_A}{IC50_A} + \dfrac{C_B}{IC50_B} + \alpha \cdot \dfrac{C_A}{IC50_A} \cdot \dfrac{C_B}{IC50_B} \right)^n}{1 + \left(\dfrac{C_A}{IC50_A} + \dfrac{C_B}{IC50_B} + \alpha \cdot \dfrac{C_A}{IC50_A} \cdot \dfrac{C_B}{IC50_B} \right)^n}$$

where C_A and C_B are the drug concentrations, and IC50A, IC50B, alpha, and n are the coefficients to be estimated.

You can define this in MATLAB code as

```
Effect = Emax*( CA/IC50A + CB/IC50B + alpha*( CA/IC50A )...
        .* ( CB/IC50B ) ).^n ./(( CA/IC50A + CB/IC50B + ...
        alpha*( CA/IC50A ) .* ( CB/IC50B ) ).^n  + 1);
```

Telling the tool which variables to fit and which parameters to estimate, requires rewriting the variable names CA and CB to x, and y. You must include x and y when you enter a custom equation in the edit box. Assume Emax = 1 because the effect output is normalized.

7 Enter the following text in the custom equation edit box.

```
( x/IC50A + y/IC50B + alpha*( x/IC50A ) .* ( y/IC50B ) ).^n
  ./(( x/IC50A + y/IC50B + alpha*( x/IC50A ) .*
  ( y/IC50B ) ).^n  + 1);
```

Curve Fitting Tool fits a surface to the data using the custom equation model.
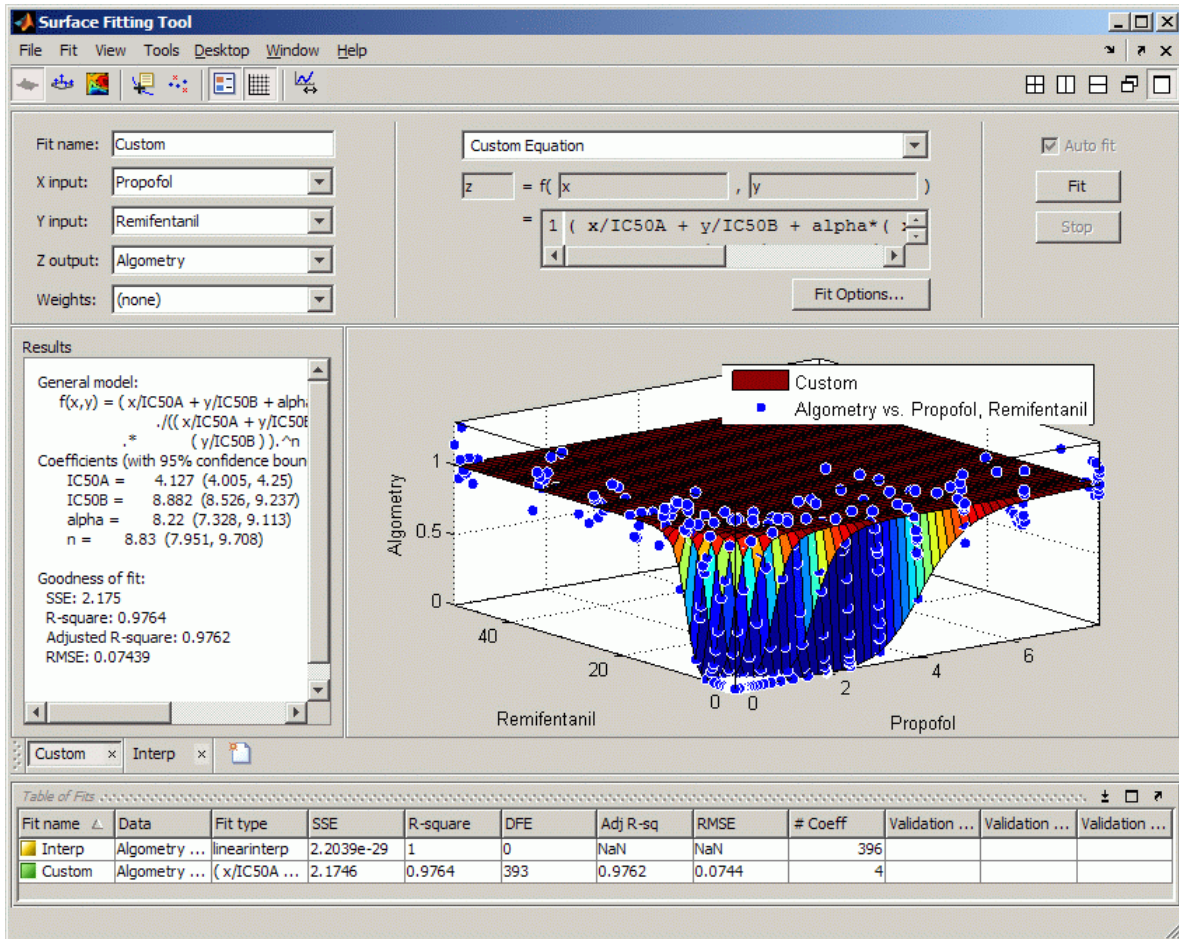
8 Set some of the fit options by clicking **Fit Options** under your custom equation.

In the Fit Options dialog box:

a Set **Robust** to Lar

b Set the **alpha StartPoint** to 1 and lower bound to –5.
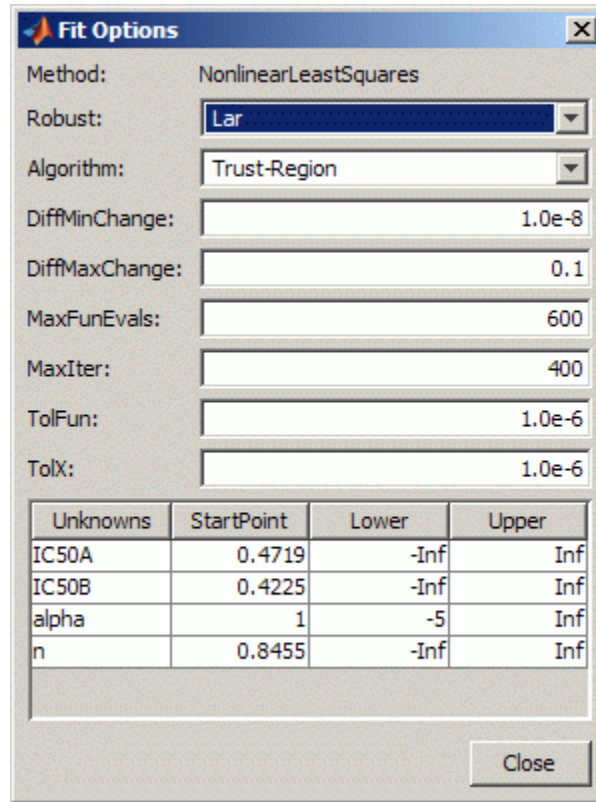
c Leave the other defaults, and click **Close**.

The tool refits with your new options.

9 Review the **Results** pane. View (and, optionally, copy) any of these results:

- The model equation
- The values of the estimated coefficients
- The goodness-of-fit statistics

10 Display the residuals plot to check the distribution of points relative to the surface by clicking the toolbar button ![icon] or selecting **View > Residuals Plot**.
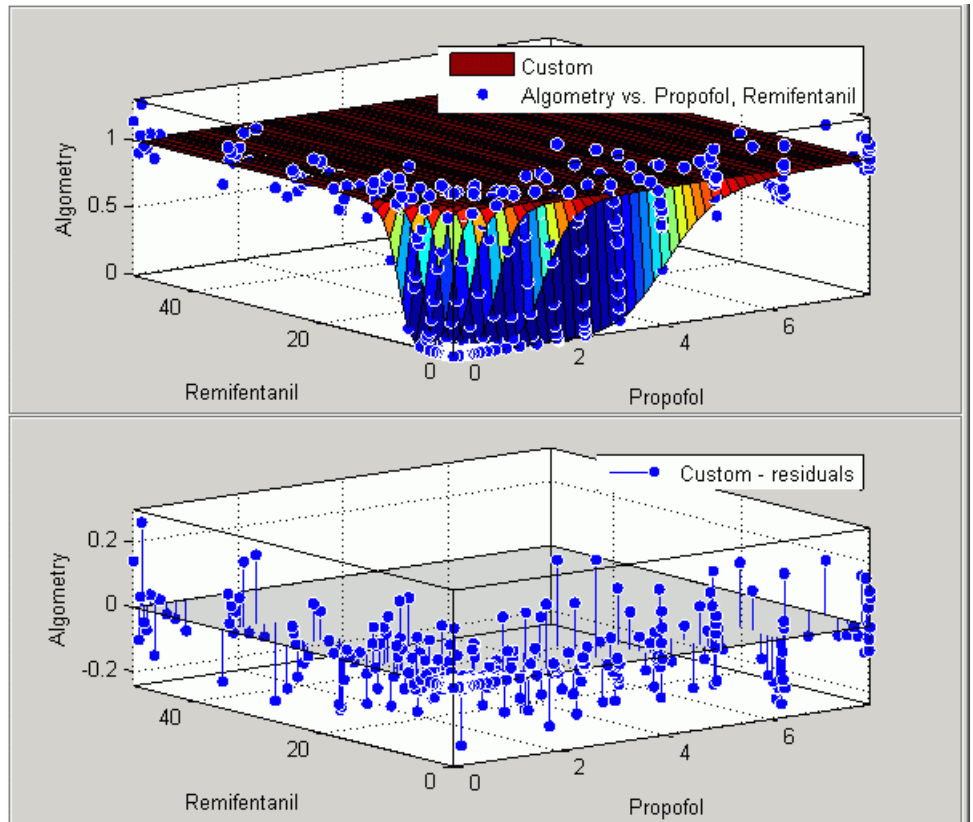
11 To generate code for all fits and plots in your Curve Fitting Tool session, select **File > Generate Code**.

The Curve Fitting Tool generates code from your session and displays the file in the MATLAB Editor. The file includes all fits and plots in your current session.

12 Save the file with the default name, `createFits.m`.

13 You can recreate your fits and plots by calling the file from the command line (with your original data or new data as input arguments). In this case, your original data still appears in the workspace.

Highlight the first line of the file (excluding the word function), and evaluate it by either right-clicking and selecting **Evaluate**, pressing **F9**, or copying and pasting the following to the command line:

```
[fitresult, gof] = createFits(Propofol,...
 Remifentanil, Algometry)
```

The function creates a figure window for each fit you had in your session. The custom fit figure shows both the surface and residuals plots that you created interactively in the Curve Fitting Tool.

14 Create a new fit to the Tetany response instead of Algometry by entering:

```
[fitresult, gof] = createFits(Propofol,...
 Remifentanil, Tetany)
```

You need to edit the file if you want the new response label on the plots. You can use the generated code as a starting point to change the surface fits and plots to fit your needs. For a list of methods you can use, see sfit.

To see how to programmatically fit surfaces to the same example problem, see "Custom Equation Fitting to Biopharmaceutical Data" on page 5-31.

Custom Equation Fitting to Biopharmaceutical Data

- "Load Data" on page 5-32
- "Create the Model Fit Type" on page 5-32
- "Fit a Surface to Algometry" on page 5-33
- "Fit a Surface to Tetany" on page 5-34
- "Fit a Surface to Sedation" on page 5-34
- "Fit a Surface to Laryingoscopy" on page 5-35

The following code, using Curve Fitting Toolbox methods, reproduces the interactive surface building with the Curve Fitting Tool described in "Surface Fitting to Biopharmaceutical Data" on page 5-24.

The following code models the response surfaces of four different measures of the analgesic and sedative response to the drug combination of Propofol and Reminfentanil.

Load Data

Load the data from file as follows:

```
data = importdata( 'OpioidHypnoticSynergy.txt' );
Propofol     = data.data(:,1);
Remifentanil = data.data(:,2);
Algometry    = data.data(:,3);
Tetany       = data.data(:,4);
Sedation     = data.data(:,5);
Laryingoscopy = data.data(:,6);
```

Create the Model Fit Type

Create the model fit type as follows:

```
ft = fittype( 'Emax*( CA/IC50A + CB/IC50B + alpha*( CA/IC50A
) * ( CB/IC50B ) )^n /(( CA/IC50A + CB/IC50B + alpha*(
CA/IC50A ) * ( CB/IC50B ) )^n  + 1 )', ...
    'indep', {'CA', 'CB'}, 'depend', 'z', 'problem', 'Emax' )
```

Output:

```
ft =

     General model:
ft(IC50A,IC50B,alpha,n,Emax,CA,CB) = Emax*...
( CA/IC50A + CB/IC50B + alpha*(CA/IC50A )...
 * ( CB/IC50B ) )^n /(( CA/IC50A + CB/IC50B...
+ alpha*( CA/IC50A ) * ( CB/IC50B ) )^n  + 1 )
```

Assume Emax = 1:

```
Emax = 1;
```

Fit a Surface to Algometry

Set fit options as follows:

```
opts = fitoptions( ft );
opts.Lower = [0 0 -5 -0];
opts.Robust = 'LAR';
opts.StartPoint = [0.00893838724332152 0.706165672266879...
 1 0.746030748284422];
```

Fit and plot a surface for Algometry:

```
[f, gof] = fit( [Propofol, Remifentanil], Algometry, ft,...
 opts, 'problem', Emax )
plot( f, [Propofol, Remifentanil], Algometry );
```
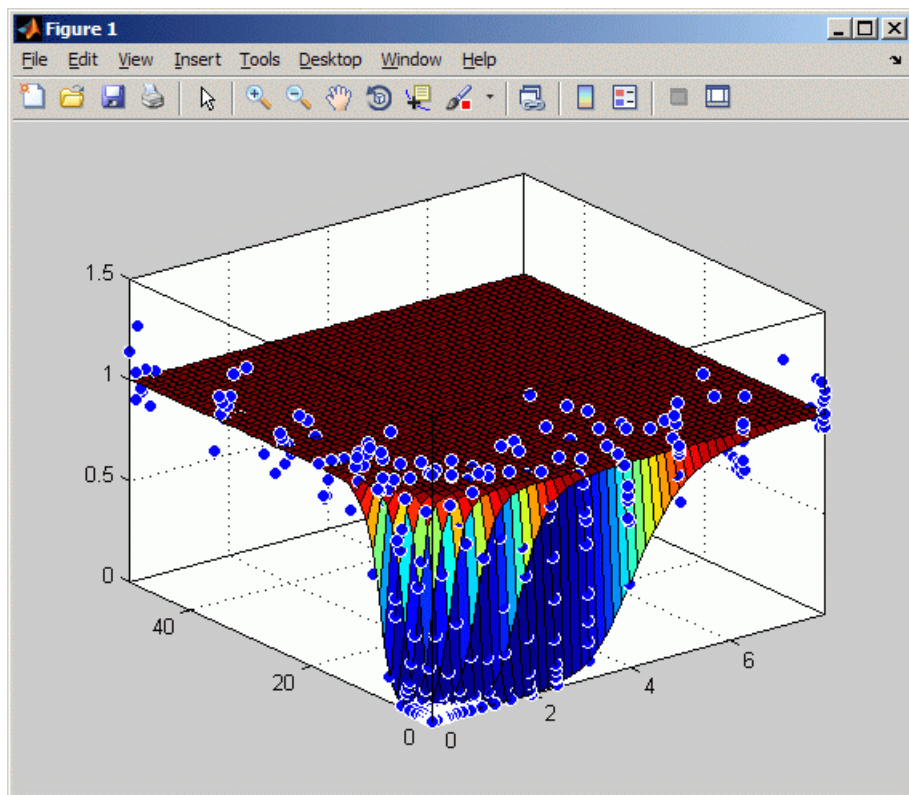
Fit a Surface to Tetany

Fit a surface to Tetany as follows:

```
[f, gof] = fit( [Propofol, Remifentanil], Tetany, ft,
opts, 'problem', Emax )
plot( f, [Propofol, Remifentanil], Tetany );
```
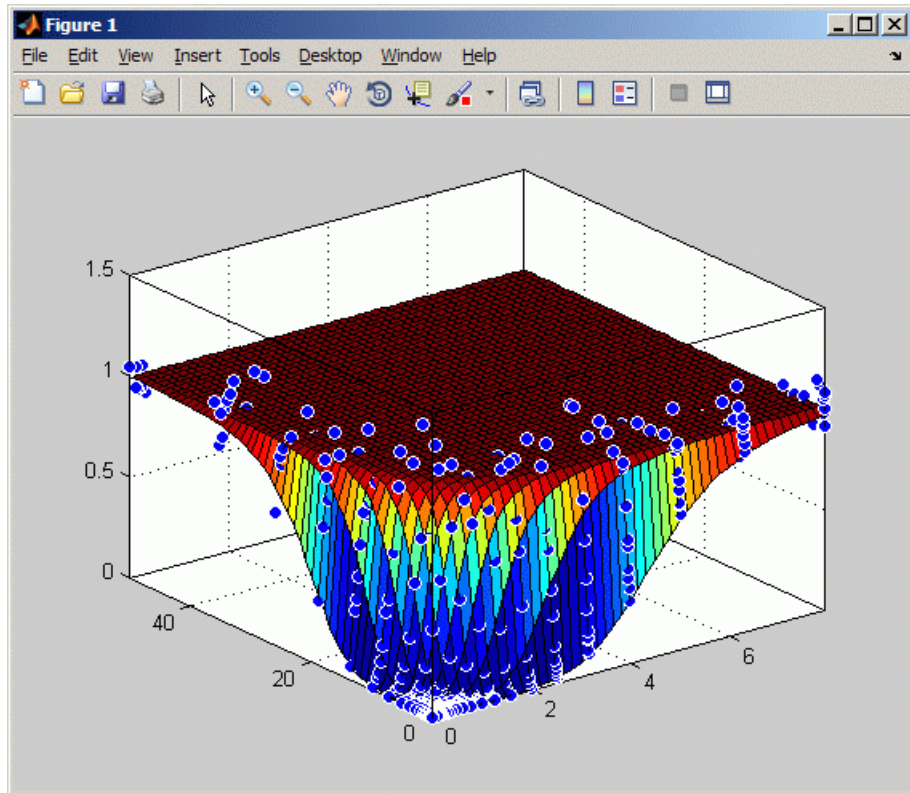

Fit a Surface to Sedation

Fit a surface to Sedation as follows:

```
[f, gof] = fit( [Propofol, Remifentanil], Sedation,
ft, opts, 'problem', Emax )
plot( f, [Propofol, Remifentanil], Sedation );
```
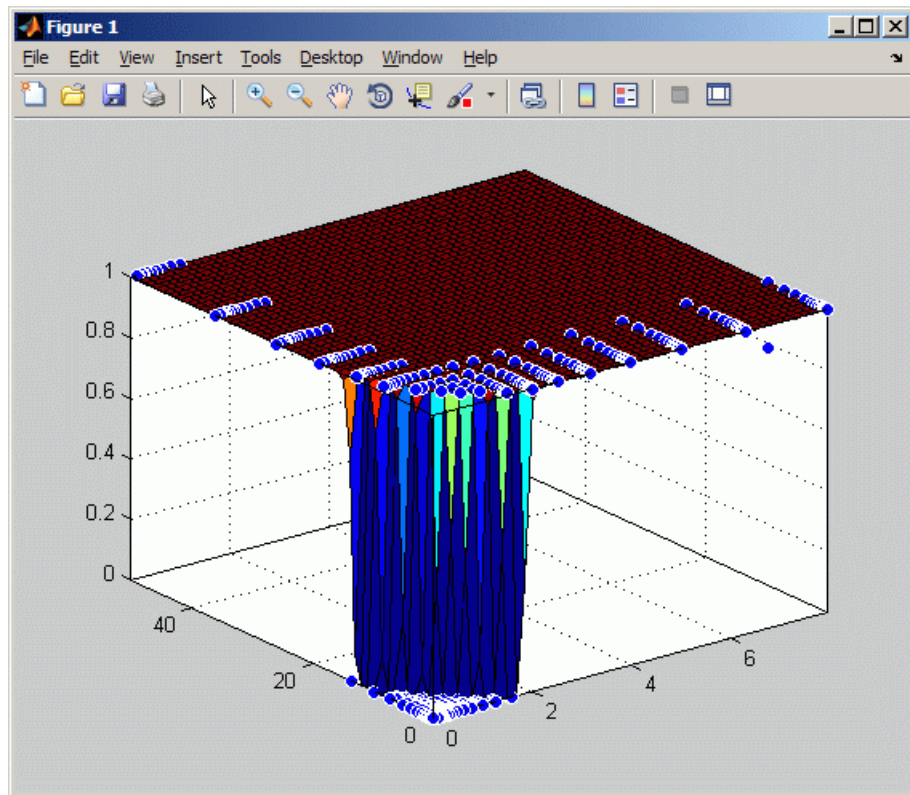
Fit a Surface to Laryingoscopy

Fit a surface to Laryingoscopy as follows:

```
[f, gof] = fit( [Propofol, Remifentanil], Laryingoscopy,
ft, opts, 'problem', Emax )
plot( f, [Propofol, Remifentanil], Laryingoscopy );
```
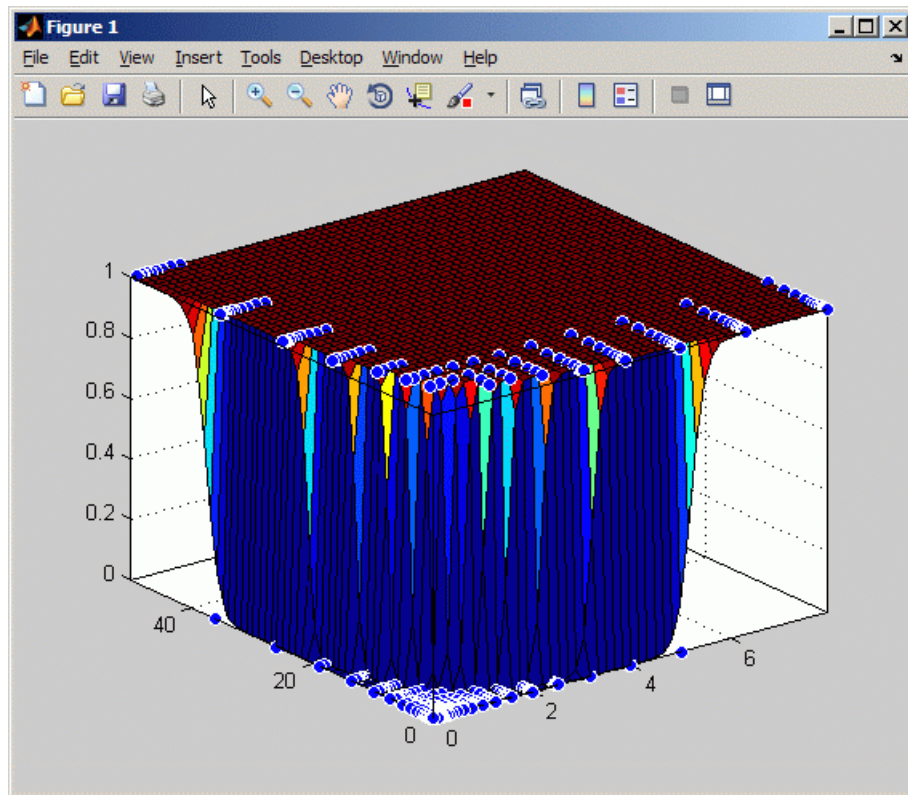
Custom Linear Models

In this section...
"About Custom Linear Models" on page 5-37
"Example: Legendre Polynomial" on page 5-37
"Creating Custom Models Using the Legacy Curve Fitting Tool" on page 5-46

About Custom Linear Models

You can use the Curve Fitting Tool custom equation fit to define your own linear or nonlinear equations. The custom equation fit uses the nonlinear least-squares fitting procedure. See "Selecting a Custom Equation Fit Interactively" on page 5-2.

You can define a custom linear equation, but the nonlinear fitting is less efficient and usually slower than linear least-squares fitting. If you need linear least-squares fitting for custom equations, you must use the legacy Curve Fitting Tool. For an example, see "Example: Legendre Polynomial" on page 5-37. For a procedure, see "Creating Custom Models Using the Legacy Curve Fitting Tool" on page 5-46.

Example: Legendre Polynomial

This example uses the legacy Curve Fitting Tool to define custom linear equations.

This example fits data using several custom linear equations. The data is generated, and is based on the nuclear reaction $^{12}C(e,e'\alpha)^{8}Be$. The equations use sums of Legendre polynomial terms.

Consider an experiment in which 124 MeV electrons are scattered from ^{12}C nuclei. In the subsequent reaction, alpha particles are emitted and produce the residual nuclei ^{8}Be. By analyzing the number of alpha particles emitted as a function of angle, you can deduce certain information regarding the nuclear dynamics of ^{12}C. The reaction kinematics are shown next.
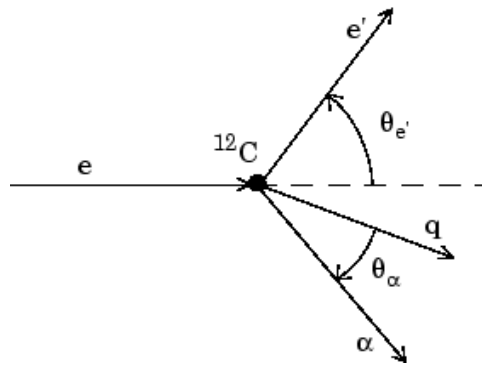
e is the incident electron.
^{12}C is the carbon target.
q is the momentum transferred to ^8Be.
e' is the scattered electron.
α is the emitted alpha particle.
$\Theta_{e'}$ is the electron scattering angle.
Θ_α is the alpha scattering angle.

The data is collected by placing solid state detectors at values of Θ_α ranging from $10°$ to $240°$ in $10°$ increments.

It is sometimes useful to describe a variable expressed as a function of angle in terms of Legendre polynomials

$$y(x) = \sum_{n=0}^{\infty} a_n P_n(x)$$

where $P_n(x)$ is a Legendre polynomial of degree n, x is $\cos(\Theta_\alpha)$, and a_n are the coefficients of the fit. Refer to the `legendre` function for information about generating Legendre polynomials.

For the alpha-emission data, you can directly associate the coefficients with the nuclear dynamics by invoking a theoretical model. Additionally, the theoretical model introduces constraints for the infinite sum shown above. In particular, by considering the angular momentum of the reaction, a fourth-degree Legendre polynomial using only even terms should describe the data effectively.

You can generate Legendre polynomials with Rodrigues' formula:

$$P_n(x) = \frac{1}{2^n n!} \left(\frac{d}{dx} \right)^n (x^2 - 1)^n$$

Legendre Polynomials up to Fourth Degree

n	$P_n(x)$
0	1
1	x
2	$(1/2)(3x^2 - 1)$
3	$(1/2)(5x^3 - 3x)$
4	$(1/8)(35x^4 - 30x^2 + 3)$

Fit the data using a fourth-degree Legendre polynomial with only even terms:

$$y_1(x) = a_0 + a_2 \left(\frac{1}{2} \right)(3x^2 - 1) + a_4 \left(\frac{1}{8} \right)(35x^4 - 30x^2 + 3)$$

1 Load the ^{12}C alpha-emission data from the file `carbon12alpha.mat`, which is provided with the toolbox:

```
load carbon12alpha
```

The workspace now contains two new variables:

- `angle` is a vector of angles (in radians) ranging from 10º to 240º in 10º increments.

- `counts` is a vector of raw alpha particle counts that correspond to the emission angles in `angle`.

2 Open the legacy Curve Fitting Tool by entering:

```
cftool -v1
```

3 Import these two variables into the Curve Fitting Tool.

In the Curve Fitting Tool, click **Data** to open the Data dialog box.

In the Data dialog box:

a Select `angle` and `counts` for X and Y.

b Name the data set `C12Alpha`

c Click **Create data set**.

d Click **Close**.

4 In the Curve Fitting Tool, click Fitting to open the Fitting dialog box.

5 In the Fitting dialog box:

a Click **New fit**.

b Specify a meaningful fit name.

c Select the `C12Alpha` data set.

d Select `Custom Equations` for the type of fit.

The Fit Editor for a custom equation fit type is shown following.
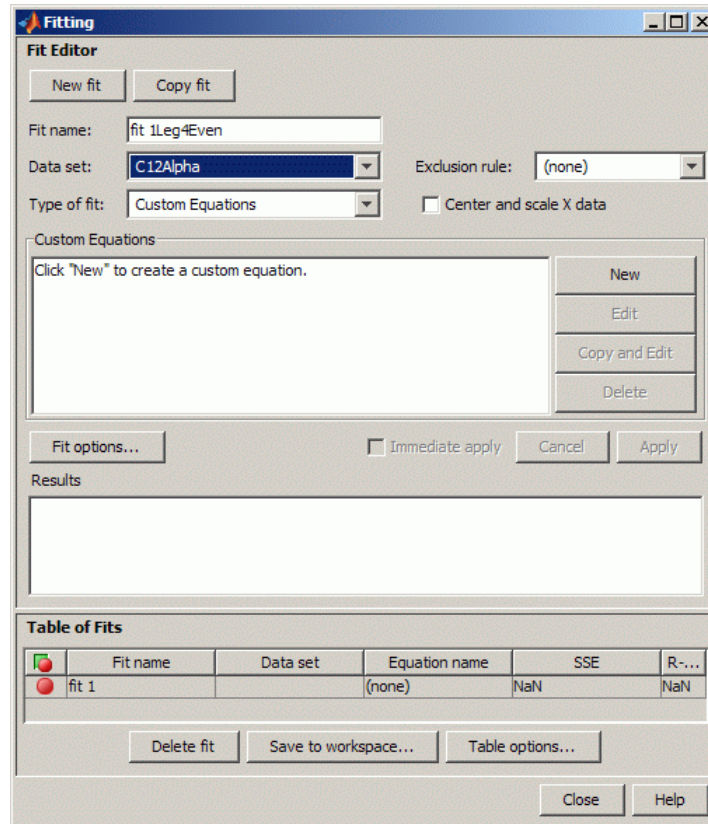
e Click **New** to open the New Custom Equation dialog box.

You use the **Linear Equations** tab in the New Custom Equation dialog box, because the Legendre polynomials depend only on the predictor variable and constants. This tab is shown for the model given by $y_1(x)$ (that is, the equation given at the beginning of this procedure). Because `angle` is given in radians, the argument of the Legendre terms is given by $\cos(\Theta_\alpha)$.
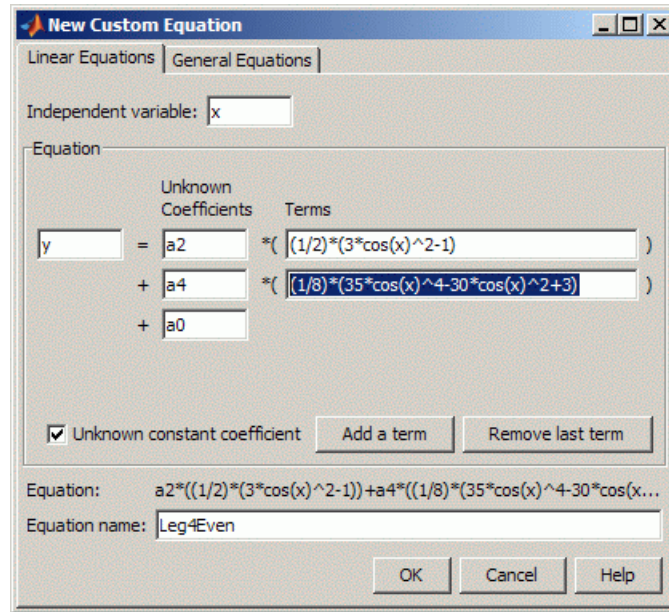
Edit these until your New Custom Equation dialog box resembles the
example shown.

i Change the coefficent names to a2, a4, and a0 (click **Add a term** to
show the third term).

ii Change the **Terms** for a2 to (1/2)*(3*cos(x)^2-1).

iii Change the **Terms** for a4 to (1/8)*(35*cos(x)^4-30*cos(x)^2+3).

iv Specify the **Equation name** Leg4Even.

v Click **OK**.

6 In the Fitting dialog box, click **Apply** to save the changes to the fit.

The fit and residuals appear in the Curve Fitting Tool. The fit appears to
follow the trend of the data well, while the residuals appear to be randomly
distributed and do not exhibit any systematic behavior.
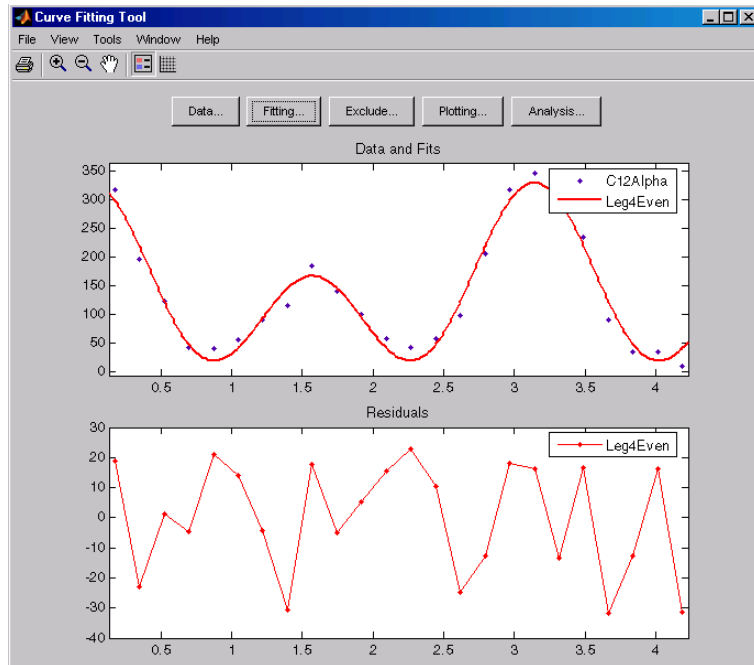
The numerical fit results in the Fitting dialog box appear next. The 95% confidence bounds indicate that the coefficients associated with $P_0(x)$ and $P_4(x)$ are known fairly accurately, but that the $P_2(x)$ coefficient has a relatively large uncertainty.

To confirm the theoretical argument that the alpha-emission data is best described by a fourth-degree Legendre polynomial with only even terms, fit the data using both even and odd terms:

$$y_2(x) = y_1(x) + a_1 x + a_3 \left(\frac{1}{2} \right) (5x^3 - 3x)$$

1 In the Fitting dialog box, click **Copy and Edit** to make a modified copy of your previous Legendre polynomial.

2 Use the **Linear Equations** tab of the New Custom Equation dialog box as follows to fit the model given by $y_2(x)$:

  a Click **Add a term** twice to add the odd Legendre terms.

  b Change the new coefficient names to `a1` and `a3`.

  c Change the **Terms** for `a1` to `cos(x)`.

  d Change the **Terms** for `a3` to `cos(x)`.

  e Specify the **Equation name** `Leg4EvenOdd`.

   The Copy and Edit Custom Equation dialog box should resemble this example.
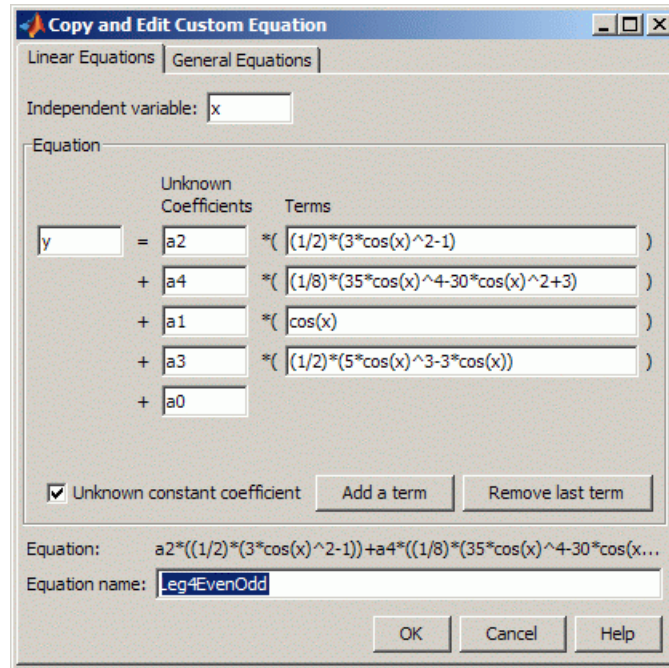
f Click **OK** to close the dialog box.

3 In the Fitting dialog box select your new equation Leg4EvenOdd, and click **Apply** to save the changes to the new fit. Observe the new fit plotted in Curve Fitting Tool.

The numerical results in the Fitting dialog box indicate that the odd Legendre terms do not contribute significantly to the fit, and the even Legendre terms are essentially unchanged from the previous fit. This confirms that the initial model choice is the best one.

```
Results

Linear model:
        f(x) = a2*(1/2)*(3*cos(x)^2-1) + a4*(1/8)*
Coefficients (with 95% confidence bounds):
        a2 =       24.19  (3.291, 45.1)
        a4 =       201.5  (177.6, 225.5)
        a1 =       1.837  (-12.89, 16.56)
        a3 =       -1.21  (-22.52, 20.1)
        a0 =       103.1  (92.71, 113.6)

Goodness of fit:
  SSE: 7940
  R-square: 0.9669
  Adjusted R-square: 0.96
  RMSE: 20.44
```

The odd Legendre coefficients are likely candidates for removal to simplify the fit because their values are small and their confidence bounds contain zero.

Creating Custom Models Using the Legacy Curve Fitting Tool

If you need linear least-squares fitting for custom equations, open the legacy Curve Fitting Tool by entering:

```
cftool -v1
```

Use the **Linear Equations** pane to define custom linear equations.

Create custom equations in the New Custom Equation dialog box. Open the dialog box in one of two ways:

- From the Curve Fitting Tool, select **Tools > Custom Equation**.

- From the Fitting dialog box, select **Custom Equations** from the **Type of fit** list, and then click the **New** button.

The dialog box contains two tabs: one for creating linear custom equations and one for creating general (nonlinear) custom equations.

Linear Equations

Linear models are linear combinations of (perhaps nonlinear) terms. They are defined by equations that are linear in the parameters. Use the **Linear Equations** pane on the New Custom Equation dialog box to create custom linear equations.

- **Independent variable** — Symbol representing the independent (predictor) variable. The default symbol is x.

- **Equation** — Symbol representing the dependent (response) variable, followed by the linear equation. The default symbol is y.

  - **Unknown Coefficients** — The unknown coefficients to be determined by the fit. The default symbols are a, b, c, and so on.

  - **Terms** — Functions of the independent variable. These can be nonlinear. Terms might not contain a coefficient to be fitted.

  - **Unknown constant coefficient** — If selected, a constant term (*y*-intercept) is included in the equation. Otherwise, a constant term is not included.

- ▪ **Add a term** — Add a term to the equation. An unknown coefficient is automatically added for each new term.

  - ▪ **Remove last term** — Remove the last term added to the equation.

- **Equation name** — The name of the equation. By default, the name is automatically updated to match the custom equation given by **Equation**. If you override the default, the name is no longer automatically updated.

General Equations

General models are, in general, nonlinear combinations of (perhaps nonlinear) terms. They are defined by equations that might be nonlinear in the parameters. Use the **General Equations** tab in the New Custom Equation dialog box to create custom general equations.

- **Independent variable** — Symbol representing the independent (predictor) variable. The default symbol is x.

- **Equation** — Symbol representing the dependent (response) variable, followed by the general equation. The default symbol is y. As you type in the terms of the equation, the unknown coefficients, associated starting values, and constraints automatically populate the table. By default, the starting values are randomly selected on the interval [0,1] and are unconstrained.

  You can immediately change the default starting values and constraints in this table, or you can change them later using the Fit Options dialog box.

- **Equation name** — The name of the equation. By default, the name is automatically updated to match the custom equation given by **Equation**. If you override the default, the name is no longer automatically updated.

Tip If you use the **General Equations** pane to define a linear equation, a nonlinear fitting procedure is used. While this is allowed, it is inefficient, and can result in less than optimal fitted coefficients. Use the **Linear Equations** tab to define custom linear equations.

Editing and Saving Custom Models

When you click **OK** in the New Custom Equation dialog box, the displayed **Equation name** is saved for the current session in the **Custom Equations** list in the Fitting dialog box. The list is highlighted as shown next.
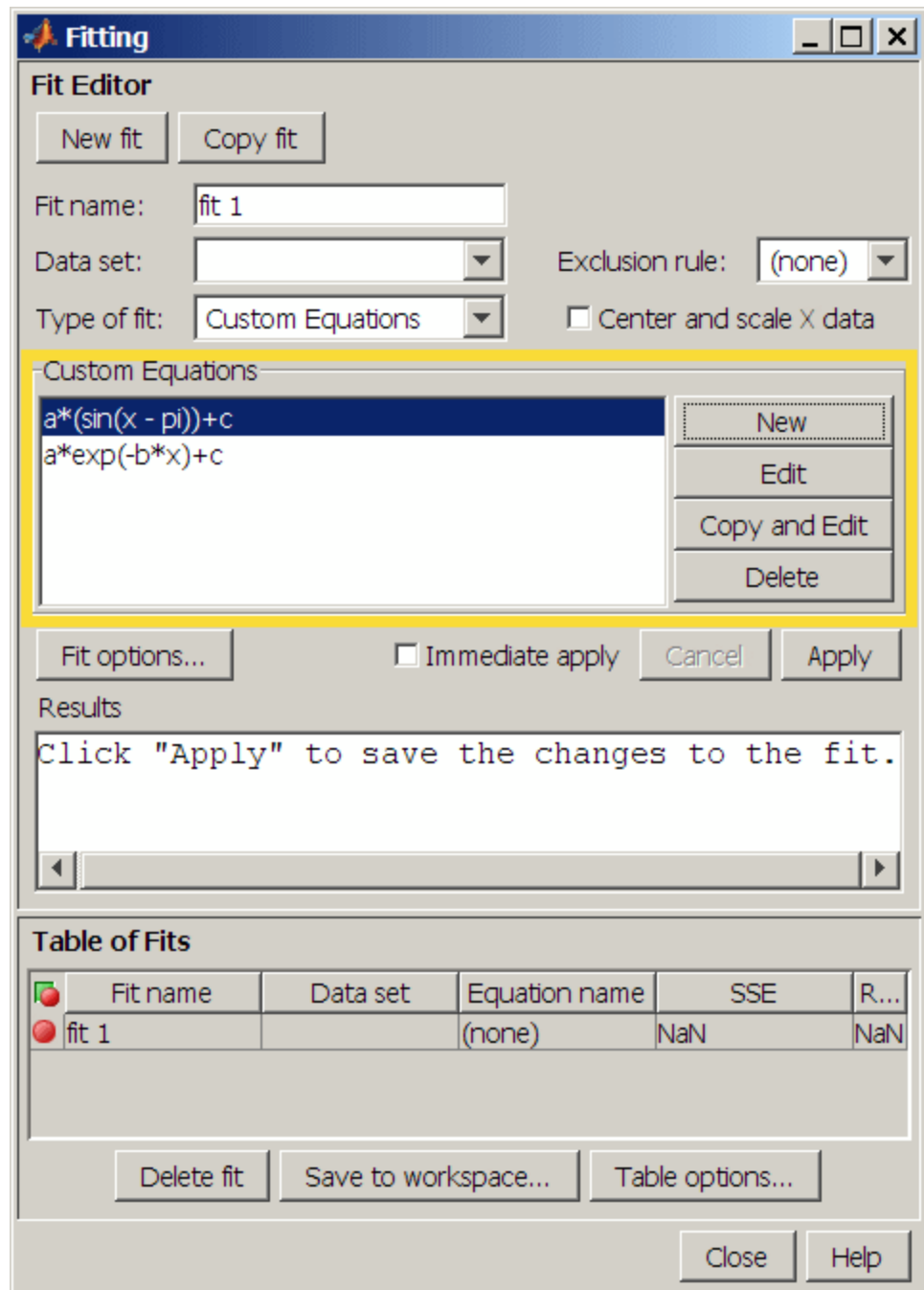
To edit a custom equation, select the equation in the **Custom Equations** list and click the **Edit** button. The Edit Custom Equation dialog box appears. It is identical to the New Custom Equation dialog box, but is prepopulated with the selected equation. After editing an equation in the Edit Custom Equation dialog box, click **OK** to save it back to the **Custom Equations** list for further use in the current session. A **Copy and Edit** button is also available, if you want to save both the original and edited equations for the current session.

To save custom equations for future sessions, select **File > Save Session** in the Curve Fitting Tool.

6

Interpolation and Smoothing

Nonparametric Fitting

In some cases, you are not concerned about extracting or interpreting fitted parameters. Instead, you might simply want to draw a smooth curve through your data. Fitting of this type is called *nonparametric fitting*. The Curve Fitting Toolbox software supports these nonparametric fitting methods:

- "Interpolants" on page 6-3 — Estimate values that lie between known data points.

- "Smoothing Splines" on page 6-8 — Create a smooth curve through the data. You adjust the level of smoothness by varying a parameter that changes the curve from a least-squares straight-line approximation to a cubic spline interpolant.

- "Lowess Smoothing" on page 6-16 — Create a smooth surface through the data using locally weighted linear regression to smooth data.

For details about interpolation, see "1-D Interpolation" and "Scattered Data Interpolation" in the MATLAB documentation.

You can also use smoothing techniques on response data. See "Filtering and Smoothing Data" on page 6-30.

To view all available model types, see "List of Library Models for Curve and Surface Fitting" on page 4-13.

Interpolants

Interpolation Methods

Interpolation is a process for estimating values that lie between known data points.

Interpolant Methods

Method	Description
Linear	Linear interpolation. This method fits a different linear polynomial between each pair of data points for curves, or between sets of three points for surfaces.
Nearest neighbor	Nearest neighbor interpolation. This method sets the value of an interpolated point to the value of the nearest data point. Therefore, this method does not generate any new data points.
Cubic spline	Cubic spline interpolation. This method fits a different cubic polynomial between each pair of data points for curves, or between sets of three points for surfaces.
Shape-preserving	Piecewise cubic Hermite interpolation (PCHIP). This method preserves monotonicity and the shape of the data. For curves only.
Biharmonic (v4)	MATLAB 4 `griddata` method. For surfaces only.

6-3

For surfaces, the Interpolant fit type uses the MATLAB `TriScatteredInterp` function for linear and nearest methods, and the MATLAB `griddata` function for cubic and biharmonic methods.

The type of interpolant to use depends on the characteristics of the data being fit, the required smoothness of the curve, speed considerations, post-fit analysis requirements, and so on. The linear and nearest neighbor methods are fast, but the resulting curves are not very smooth. The cubic spline and shape-preserving and v4 methods are slower, but the resulting curves are very smooth.

For example, the nuclear reaction data from the `carbon12alpha.mat` file is shown here with a nearest neighbor interpolant fit and a shape-preserving (PCHIP) interpolant fit. Clearly, the nearest neighbor interpolant does not follow the data as well as the shape-preserving interpolant. The difference between these two fits can be important if you are interpolating. However, if you want to integrate the data to get a sense of the total strength of the reaction, then both fits provide nearly identical answers for reasonable integration bin widths.
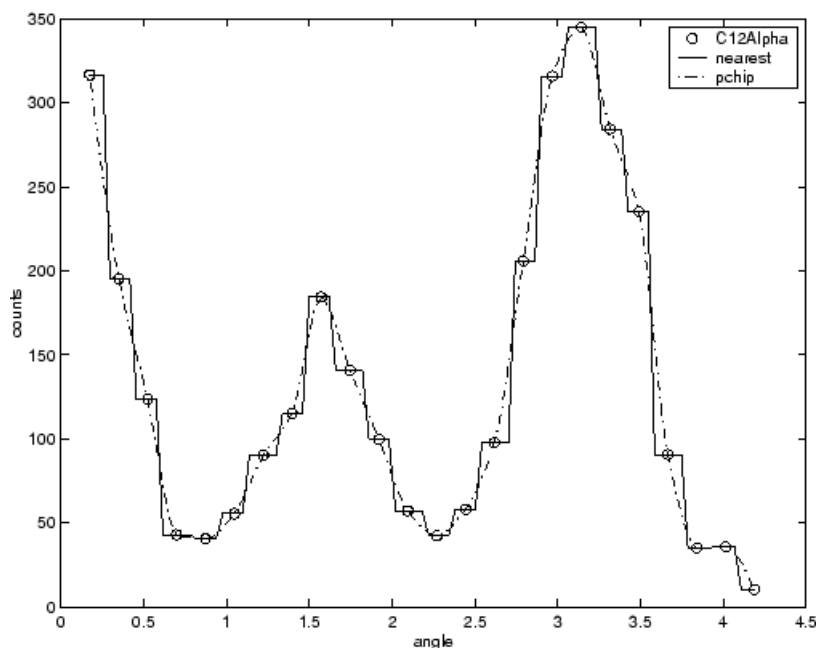
Note Goodness-of-fit statistics, prediction bounds, and weights are not defined for interpolants. Additionally, the fit residuals are always 0 (within computer precision) because interpolants pass through the data points.

Interpolants are defined as *piecewise polynomials* because the fitted curve is constructed from many "pieces" (except for `Biharmonic` for surfaces which is a radial basis function interpolant). For cubic spline and PCHIP interpolation, each piece is described by four coefficients, which the toolbox calculates using a cubic (third-degree) polynomial.

- Refer to the `spline` function for more information about cubic spline interpolation.

- Refer to the `pchip` function for more information about shape-preserving interpolation, and for a comparison of the two methods.

- Refer to the `TriScatteredInterp` and `griddata` functions for more information about surface interpolation.

It is possible to fit a single "global" polynomial interpolant to data, with a degree one less than the number of data points. However, such a fit can have wildly erratic behavior between data points. In contrast, the piecewise polynomials described here always produce a well-behaved fit, so they are more flexible than parametric polynomials and can be effectively used for a wider range of data sets.

Selecting an Interpolant Fit Interactively

In the Curve Fitting Tool, select `Interpolant` from the model type list.

The `Interpolant` fit category fits an interpolating curve or surface that passes through every data point. For surfaces, the Interpolant fit type uses the MATLAB `TriScatteredInterp` function for linear and nearest methods, and the MATLAB `griddata` function for cubic and biharmonic methods.

The settings are shown here.
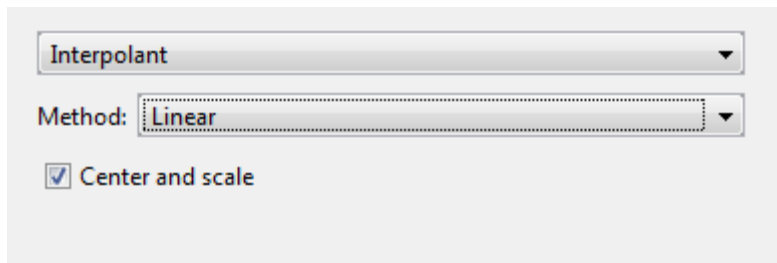
You can specify the **Method** setting: `Nearest neighbor`, `Linear`, `Cubic`,
`Shape-preserving (PCHIP)` (for curves) or `Biharmonic (v4)` (for surfaces).
For details, see "Interpolation Methods" on page 6-3.

Tip If you are fitting a surface and your input variables have different scales,
turn the **Center and scale** option on and off to see the difference in the
surface fit. Normalizing the inputs can strongly influence the results of the
triangle-based (i.e., piecewise `Linear` and `Cubic` interpolation) and `Nearest
neighbor` surface interpolation methods.

Selecting an Interpolant Fit at the Command Line

Specify the interpolant model method when you call the `fit` function using
one of these options.

Fitting Method	Description
`'nearestinterp'`	Nearest neighbor interpolation
`'linearinterp'`	Linear interpolation
`'cubicinterp'`	Cubic spline interpolation
`'pchipinterp'`	Piecewise cubic Hermite interpolation (curves only)
`'biharmonicinterp'`	Biharmonic (MATLAB 4 `griddata`) interpolation (surfaces only)

There are no additional fit option parameters for any of the interpolant
methods.

For example, to load some data and fit a linear interpolant model:

```
load census;
f = fit(cdate,  pop, 'linearinterp')
plot(f,cdate,pop)
```

To create and compare nearest neighbor and pchip interpolant fits on a plot:

```
load carbon12alpha
f1 = fit(angle, counts, 'nearestinterp')
f2 = fit(angle, counts, 'pchip')
p1 = plot(f1, angle, counts)
xlim( [min(angle), max(angle)])
hold on
p2 = plot(f2, 'b')
hold off
legend([p1; p2], 'Counts per Angle','Nearest', 'pchip')
```

For an alternative to 'cubicinterp' or 'pchipinterp', you can use other spline functions that allow greater control over what you can create. See "About Splines in Curve Fitting Toolbox" on page 8-2.

Smoothing Splines

In this section...
"About Smoothing Splines" on page 6-8
"Selecting a Smoothing Spline Fit Interactively" on page 6-9
"Selecting a Smoothing Spline Fit at the Command Line" on page 6-11
"Example: Nonparametric Fitting with Cubic and Smoothing Splines" on page 6-11

About Smoothing Splines

If your data is noisy, you might want to fit it using a smoothing spline. Alternatively, you can use one of the smoothing methods described in "Filtering and Smoothing Data" on page 6-30.

The smoothing spline s is constructed for the specified *smoothing parameter p* and the specified weights w_i. The smoothing spline minimizes

$$p\sum_i w_i \left(y_i - s(x_i)\right)^2 + (1-p)\int \left(\frac{d^2 s}{dx^2}\right)^2 dx$$

If the weights are not specified, they are assumed to be 1 for all data points.

p is defined between 0 and 1. $p = 0$ produces a least-squares straight-line fit to the data, while $p = 1$ produces a cubic spline interpolant. If you do not specify the smoothing parameter, it is automatically selected in the "interesting range." The interesting range of p is often near $1/(1+h^3/6)$ where h is the average spacing of the data points, and it is typically much smaller than the allowed range of the parameter. Because smoothing splines have an associated smoothing parameter, you might consider these fits to be parametric in that sense. However, smoothing splines are also piecewise polynomials like cubic spline or shape-preserving interpolants and are considered a nonparametric fit type in this guide.

Note The smoothing spline algorithm is based on the `csaps` function.

The nuclear reaction data from the file *carbon12alpha.mat* is shown here with three smoothing spline fits. The default smoothing parameter ($p = 0.99$) produces the smoothest curve. The cubic spline curve ($p = 1$) goes through all the data points, but is not quite as smooth. The third curve ($p = 0.95$) misses the data by a wide margin and illustrates how small the "interesting range" of p can be.
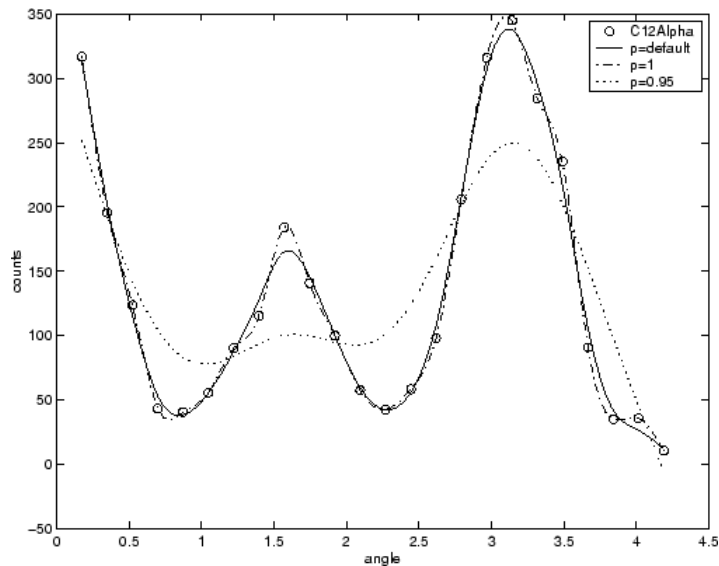
Selecting a Smoothing Spline Fit Interactively

In the Curve Fitting Tool, select `Smoothing Spline` from the model type list.
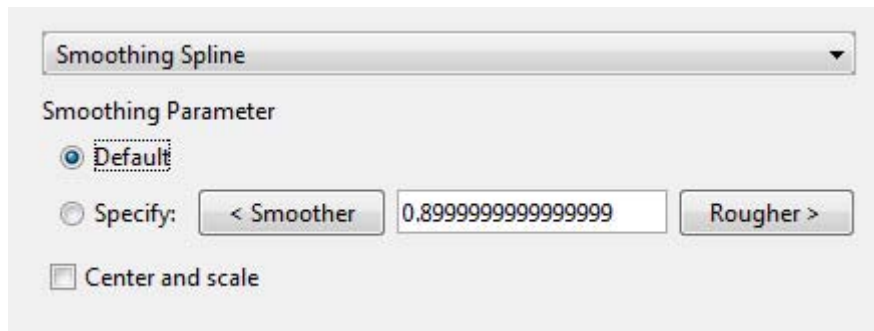
You can specify the following options:

- To make a smoother fit further from the data, click the **< Smoother** button repeatedly until the plot shows the smoothness you want.

- To make a rougher fit closer to the data, click the **Rougher >** button until you are satisfied with the plot.

- Alternatively, specify any value from 0 to 1 for the smoothing parameter. 0 produces a linear polynomial fit (a least-squares straight-line fit to the data), while 1 produces a piecewise cubic polynomial fit that passes through all the data points (a cubic spline interpolant).

- Click **Default** to return to the initial value. The toolbox attempts to select a default value appropriate for your data. See "About Smoothing Splines" on page 6-8.

For example:

1 Load the data in "About Smoothing Splines" on page 6-8 by entering:

```
load carbon12alpha
```

2 In the Curve Fitting Tool, select angle for **X data** and counts for **Y data**.

3 Select the Smoothing Spline fit type.

4 Try smoothing parameter values 1, 0.95, and the default value (0.99).

Selecting a Smoothing Spline Fit at the Command Line

Specify the model type `'smoothingspline'` when you call the `fit` function.

For example, to load some data and fit a smoothing spline model:

```
load enso
f = fit(month, pressure, 'smoothingspline')
 plot(f, month, pressure)
```

To view the smoothing parameter the toolbox calculates, create the fit using the third output argument that contains data-dependent fit options:

```
[f,gof,out] = fit( month, pressure, 'smoothingspline')
```

The smoothing parameter is the `p` value in the `out` structure: `out.p = 0.9`. The default value depends on the data set.

You can specify the smoothing parameter for a new fit with the `SmoothingParam` property. Its value must be between 0 and 1.

For example, to specify a smoothing parameter:

```
f = fit(month, pressure,  'smoothingspline', 'SmoothingParam', 0.6)
 plot(f, month, pressure)
```

Alternatively, use `fitoptions` to specify a smoothing parameter before fitting:

```
 options = fitoptions('Method','Smooth','SmoothingParam',0.07)
 [f,gof,out] = fit(month,pressure,'smooth',options)
```

For an alternative to `'smoothingspline'`, you can use the `csaps` cubic smoothing spline function or other spline functions that allow greater control over what you can create. See "About Splines in Curve Fitting Toolbox" on page 8-2.

Example: Nonparametric Fitting with Cubic and Smoothing Splines

This example fits some data using a cubic spline interpolant and several smoothing splines.

1 Create the variables in your workspace:

```
x = (4*pi)*[0 1 rand(1,25)];
y = sin(x) + .2*(rand(size(x))-.5);
```

2 Open the Curve Fitting Tool by entering:

```
cftool
```

3 Select x and y from the **X data** and **Y data** lists.

The Curve Fitting Tool fits and plots the data.

4 Fit the data with a cubic spline interpolant by selecting **Interpolant** fit type and the **Method** `Cubic`.
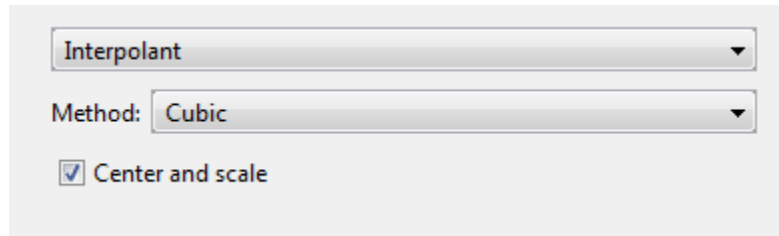
The Curve Fitting Tool fits and plots the cubic spline interpolant.

5 Enter the **Fit name** `cubicsp`.

6 View the **Results** pane. Goodness-of-fit statistics such as RMSE are not defined (shown as `NaN`) for interpolants.

Results

Cubic spline interpolant:
  f(x) = piecewise polynomial computed from p
  where x is normalized by mean 5.39 and std 3.998
Coefficients:
  p = coefficient structure

Goodness of fit:
 SSE: 7.704e-034
 R-square: 1
 Adjusted R-square: NaN
 RMSE: NaN

A cubic spline interpolation is defined as a piecewise polynomial that results in a structure of coefficients (p). The number of "pieces" in the structure is one less than the number of fitted data points, and the number of coefficients for each piece is four because the polynomial degree is three. You can examine the coefficient structure p if you export your fit to the workspace (e.g., enter *fitname*.p). For information on the structure of coefficients, see "Constructing and Working with ppform Splines" on page 10-12.

7 Create another fit to compare. Right-click your fit in the **Table of Fits** and select **Duplicate 'cubicsp'**.

8 Fit the data with a smoothing spline by selecting **Smoothing Spline**.

Smoothing Spline

Smoothing Parameter

⦿ Default

○ Specify: [<]   0.9994820481360743   [>]

☐ Center and scale
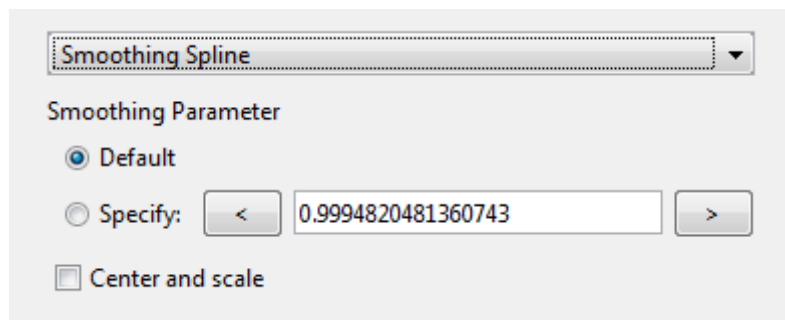
The level of smoothness is given by the **Smoothing Parameter**. The default smoothing parameter value depends on the data set, and is automatically calculated by the toolbox.

For this data set, the default smoothing parameter is close to 1, indicating that the smoothing spline is nearly cubic and comes very close to passing through each data point.

9 Name the default smoothing parameter fit `Smooth1`. If you do not like the level of smoothing produced by the default smoothing parameter, you can specify any value from 0 to 1. 0 produces a linear polynomial fit, while 1 produces a piecewise cubic polynomial fit that passes through all the data points.

The numerical results for the smoothing spline fit are shown here.

Results

Smoothing spline:
    f(x) = piecewise polynomial computed from p
Smoothing parameter:
    p = 0.99986514

Goodness of fit:
  SSE: 0.01494
  R-square: 0.9987
  Adjusted R-square: 0.994
  RMSE: 0.05128

10 For comparison purposes, create another smoothing spline fit. Right-click your fit in the **Table of Fits** and select **Duplicate 'smooth1'**. Change the smoothing parameter to `0.5` and name the fit `Smooth2`.

11 Compare the plots for your three fits. Explore the fit behavior beyond the data limits by increasing the default abscissa scale. You change the axes limits with **Tools > Axes Limit Control** menu item.

Note Your results depend on random start points and may vary from those described.

Note that the default smoothing parameter produces a curve that is smoother than the interpolant, but is a good fit to the data. In this case, decreasing the smoothing parameter from the default value produces a curve that is smoother still, but is not a good fit to the data. As the smoothing parameter increases beyond the default value, the associated curve approaches the cubic spline interpolant. The cubic spline and default smoothing spline are similar for interior points, but diverge at the end points.

Lowess Smoothing

In this section...
"About Lowess Smoothing" on page 6-16
"Selecting a Lowess Fit Interactively" on page 6-16
"Selecting a Lowess Fit at the Command Line" on page 6-18
"Fitting Automotive Fuel Efficiency Surfaces at the Command Line" on page 6-18

About Lowess Smoothing

Use Lowess models to fit smooth surfaces to your data. The names "lowess" and "loess" are derived from the term "locally weighted scatter plot smooth," as both methods use locally weighted linear regression to smooth data. The process is *weighted* because the toolbox defines a regression weight function for the data points contained within the span. In addition to the regression weight function, the **Robust** option is a weight function that can make the process resistant to outliers.

For more information on these two types of smoothing fit, see "Local Regression Smoothing" on page 6-34.

Selecting a Lowess Fit Interactively

In the Curve Fitting Tool, select Lowess from the model type list.

You can use the Lowess model type to fit smooth surfaces to your data with either lowess or loess methods. The Lowess fits use locally weighted linear regression to smooth data.
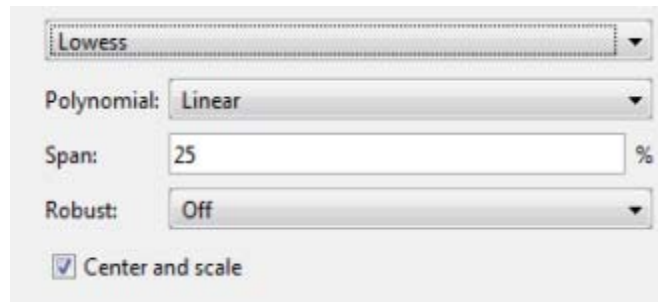
You can specify the following options:

- Select Linear or Quadratic from the list to specify the type of **Polynomial** model to use in the regression. In Curve Fitting Toolbox, lowess fitting uses a linear polynomial, while loess fitting uses a quadratic polynomial.

- Use **Span** to specify the span as a percentage of the total number of data points in the data set. The toolbox uses neighboring data points defined within the span to determine each smoothed value. This role of neighboring points is the reason why the smoothing process is called "local."

Tip Increase the span to make the surface smoother. Reduce the span to make the surface follow the data more closely.

- The **Robust** linear least-squares fitting method you want to use (Off, LAR, or Bisquare). The local regression uses the **Robust** option. Using the **Robust** weight function can make the process resistant to outliers. For details, see Robust on the fitoptions reference page.

Tip If your input variables have very different scales, turn the **Center and scale** option on and off to see the difference in the surface fit. Normalizing the inputs can strongly influence the results of a Lowess fitting.

For an interactive example using Lowess, see "Surface Fitting to Franke Data" on page 2-36.

Selecting a Lowess Fit at the Command Line

Specify the model type `'lowess'` when you call the `fit` function. For example, to load and fit some data with a lowess model and plot the results:

```
load franke
f=fit([x,y],z, 'lowess')
plot(f,[x,y], z)
```

For a command-line example using Lowess, see the next section, "Fitting Automotive Fuel Efficiency Surfaces at the Command Line" on page 6-18.

Fitting Automotive Fuel Efficiency Surfaces at the Command Line

- "Load and Preprocess Data" on page 6-19
- "Fit and Plot Surfaces of Fuel Efficiency" on page 6-20
- "Create a Table from the Surface" on page 6-24

The toolbox provides sample data generated from a GTPOWER predictive combustion engine model. The model emulates a naturally aspirated spark-ignition, 2-liter, inline 4-cylinder engine. You can use surface fitting methods to fit a response surface to this data to investigate fuel efficiency.

The data set includes the required variables to model response surfaces:

- *Speed* is in revolutions per minute (rpm) units.
- *Load* is the normalized cylinder air mass (the ratio of cylinder aircharge to maximum naturally aspirated cylinder aircharge at standard temperature and pressure).
- *BSFC* is the brake-specific fuel consumption in g/KWh. That is, the energy flow in, divided by mechanical power out (fuel efficiency).

The aim is to model a response surface to find the minimum BSFC as a function of speed and load. You can use this surface as a table, included as part of a hybrid vehicle optimization algorithm combining the use of a motor and your engine. To operate the engine as fuel efficiently as possible, the table must operate the engine near the bottom of the BSFC bowl.

Load and Preprocess Data

1 Load the data from the XLS spreadsheet. Use the `'basic'` command option for non- Windows® platforms.

Create a variable n that has all the numeric data in one array.

```
n = xlsread( 'Engine_Data_SI_NA_2L_I4.xls', 'SI NA 2L I4',...
 '', 'basic' );
```

2 Extract from the variable n the columns of interest:

```
SPEED = n(:,2);
LOAD_CMD = n(:,3);
LOAD = n(:,8);
BSFC = n(:,22);
```

3 Process the data before fitting, to pick out the `min(BSFC)` values from each sweep. The data points are organized insweeps on speed/load.

a Get a list of the speed/load sites:

```
SL = unique( [SPEED, LOAD_CMD], 'rows' );
nRuns = size( SL, 1 );
```

b For each speed/load site, find the data at the site and extract the actual measured load and the minimum BSFC.

```
minBSFC  = zeros( nRuns, 1 );
Load     = zeros( nRuns, 1 );
Speed    = zeros( nRuns, 1 );
for i = 1:nRuns
    idx = SPEED == SL(i,1) & LOAD_CMD == SL(i,2);

    minBSFC(i) = min(  BSFC(idx) );
    Load(i)    = mean( LOAD(idx) );
    Speed(i)   = mean( SPEED(idx) );
end
```

Fit and Plot Surfaces of Fuel Efficiency

Follow these steps to fit and plot some surfaces:

1 Fit a surface to the preprocessed data.

```
f1 = fit( [Speed, Load], minBSFC, 'Lowess', 'Normalize', 'on' )
```

This command results in the following output:

```
Locally weighted smoothing linear regression:
  f1(x,y) = lowess (linear) smoothing regression
  computed from p
  where x is normalized by mean 3407 and std 1214
  and where y is normalized by mean 0.5173 and std 0.1766
Coefficients:
  p = coefficient structure
```

2 Plot your fit:

```
plot( f1, [Speed, Load], minBSFC );
xlabel( 'Speed [RPM]' );
ylabel( 'Load [%]' );
zlabel( 'Minimum BSFC [g/Kwh]' );
```
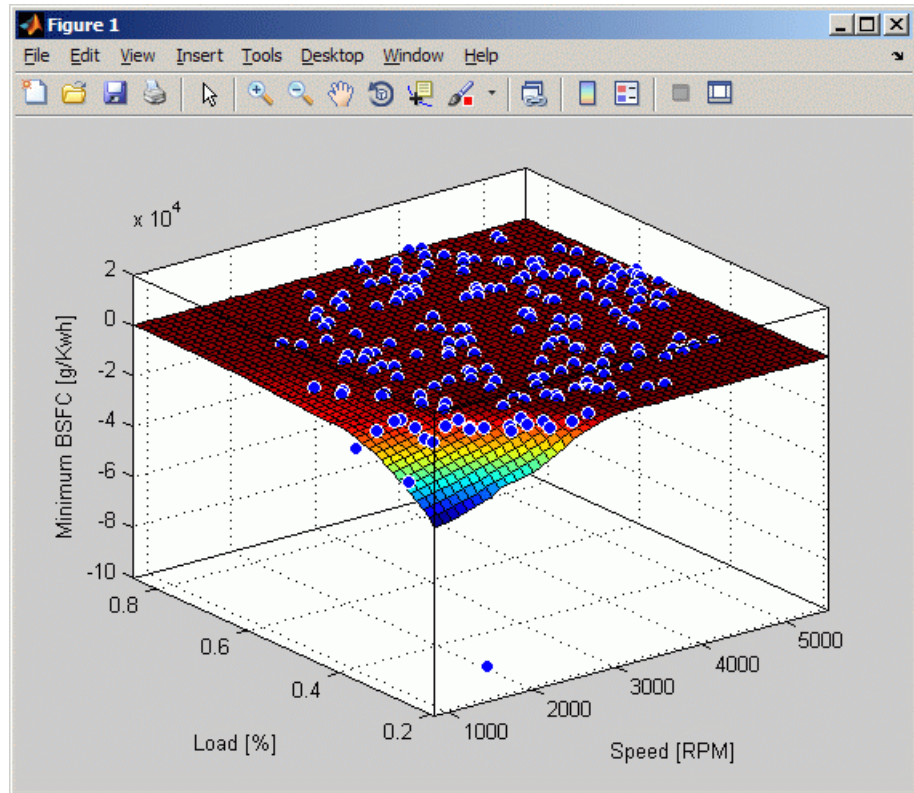
3 Review the resulting plot:

- There are points where BSFC is negative because this data is generated by an engine simulation.

- Remove those problem data points and keep points in the range [0, Inf].

```
out = excludedata( Speed, minBSFC, 'Range', [O, Inf] );
f2 = fit( [Speed, Load], minBSFC, 'Lowess', ...
'Normalize', 'on', 'Exclude', out )
```

Examine the following output:

```
Locally weighted smoothing linear regression:
       f2(x,y) = lowess (linear) smoothing regression
```

```
                    computed from p
                    where x is normalized by mean 3443 and std 1187
                    and where y is normalized by mean 0.521 and std 0.175
                Coefficients:
                    p = coefficient structure
```

4 Plot the new fit:

```
plot( f2, [Speed, Load], minBSFC, 'Exclude', out );
xlabel( 'Speed [RPM]' );
ylabel( 'Load [%]' );
zlabel( 'Minimum BSFC [g/Kwh]' );
```

5 Zoom in on the part of the z-axis of interest:

```
set( gca, 'ZLim', [0, max( minBSFC )] );
```
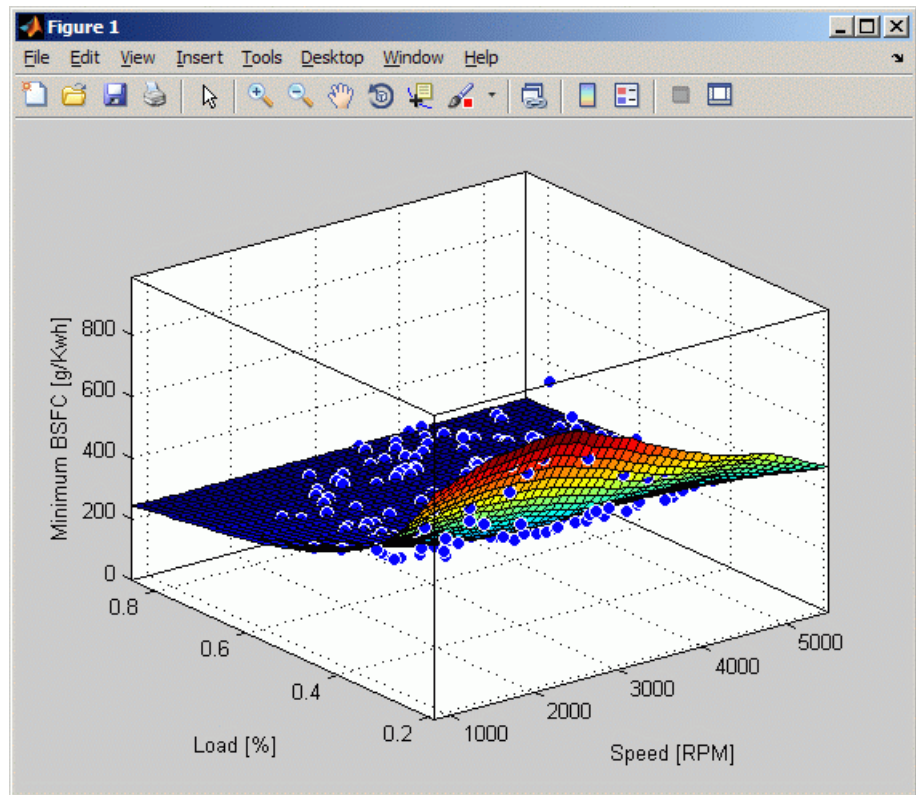
6 Because you want to operate the engine efficiently, create a contour plot to see the region where the BSFC is low. Use the plot command, and specify the parameter/value pair 'style''Contour'.

```
plot( f2, [Speed, Load], minBSFC, 'Exclude', out,...
  'Style', 'Contour' );
xlabel( 'Speed [RPM]' );
ylabel( 'Load [%]' );
colorbar
```
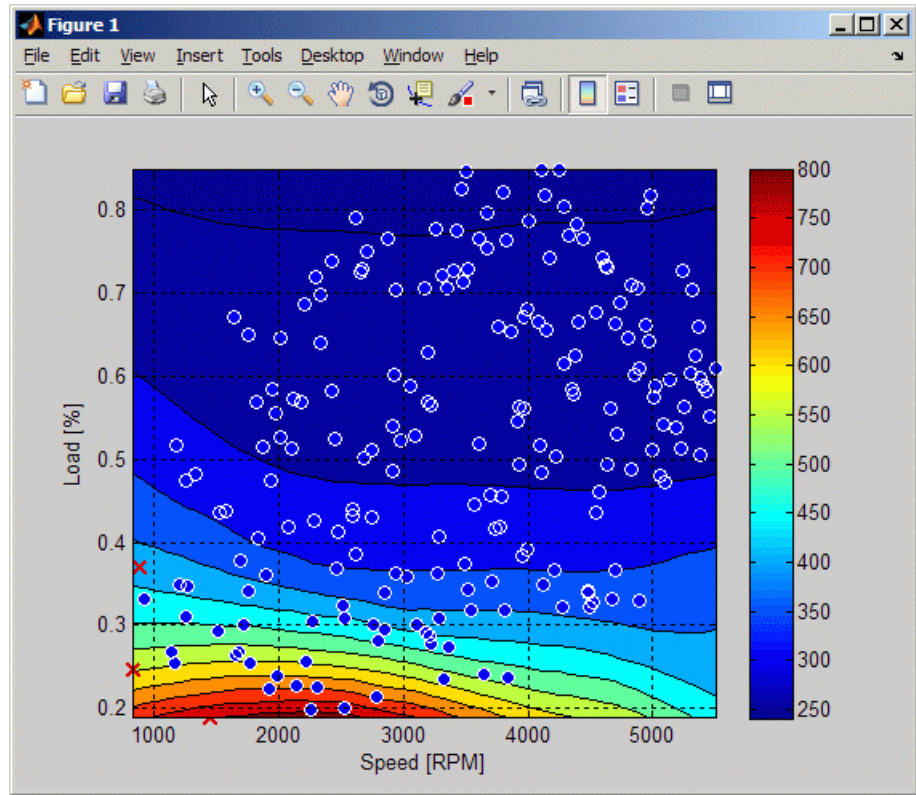
Create a Table from the Surface

In this exercise, you generate a table from the original data using model `f2`.

1 Create variables for the table breakpoints.

```
speedbreakpoints = linspace( 1000, 5500, 17 );
loadbreakpoints =  linspace( 0.2, 0.8, 13 );
```

2 To generate values for the table, evaluate the model over a grid of points.

```
[tSpeed, tLoad] = meshgrid( speedbreakpoints,...
 loadbreakpoints );
tBSFC = f2( tSpeed, tLoad );
```

3 Examine the rows and columns of the table at the command line.

```
tBSFC(1:2:end,1:2:end)
```

4 Plot the table against the original model. The grid on the model surface shows the table breakpoints.

```
h = plot( f2 );
set( h, 'EdgeColor', 'none' );
hold on
mesh( tSpeed, tLoad, tBSFC, ...
    'LineStyle', '-', 'LineWidth', 2, 'EdgeColor', 'k', ...
'FaceColor', 'none', 'FaceAlpha', 1 );
hold off
xlabel( 'Speed [RPM]' );
ylabel( 'Load [%]' );
zlabel( 'Minimum BSFC [g/Kwh]' );
```
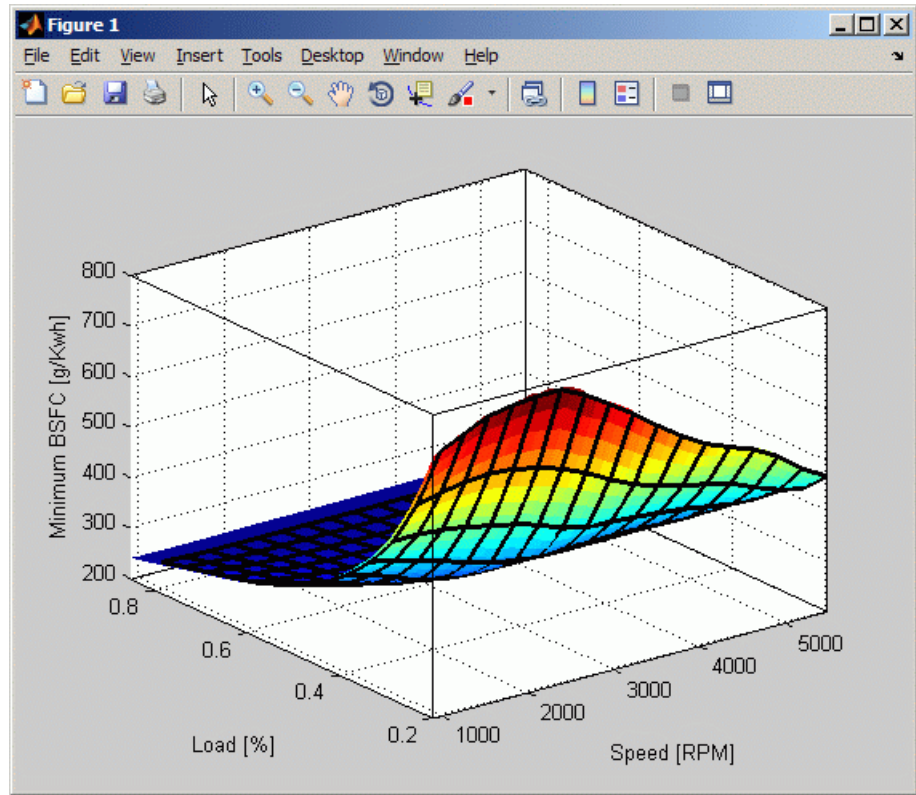
5 Check the table accuracy:

  a View the difference between the model and the table by plotting the difference between them on a finer grid.

  b Then, use this difference in prediction accuracy between the table and the model to determine the most efficient table size for your accuracy requirements.

The following code evaluates the model over a finer grid and plots the difference between the model and the table:

```
[tfSpeed, tfLoad] = meshgrid( linspace( 1000, 5500,...
 8*17+1 ), linspace( 0.2, 0.8, 8*13+1 ) );
tfBSFC_model = f2( tfSpeed, tfLoad );
```

```
tfBSFC_table = interp2( tSpeed, tLoad, tBSFC, tfSpeed,...
tfLoad, 'linear' );
tfDiff = tfBSFC_model - tfBSFC_table;

surf( tfSpeed, tfLoad, tfDiff, 'LineStyle', 'none'  );
hold on
mesh( tSpeed, tLoad, zeros( size( tBSFC ) ), ...
    'LineStyle', '-', 'LineWidth', 2, 'EdgeColor', 'k', ...
'FaceColor', 'none', 'FaceAlpha', 1 );
hold off
axis tight
xlabel( 'Speed [RPM]' );
ylabel( 'Load [%]' );
zlabel( 'Difference between model and table [g/Kwh]' );
title( sprintf( 'Max difference: %g', max( abs( tfDiff(:) ) ) ) ) );
```
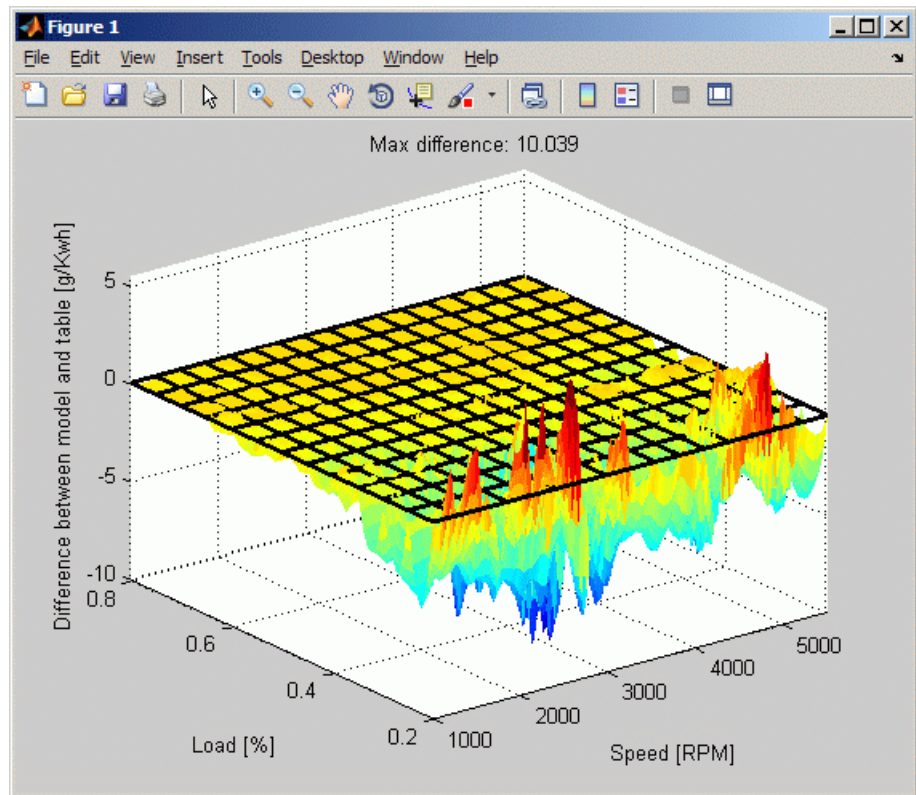
6 If you have Simulink® software, you can create a Look Up Table block.

   a Create a model with a 2-D Lookup Table block.

```
simulink
new_system('my_model')
open_system('my_model')
add_block('Simulink/Lookup Tables/2-D Lookup Table',...
'my_model/surfaceblock')
```

   b Populate the Lookup Table with speed breakpoints, load breakpoints, and a lookup table.

```
set_param('my_model/surfaceblock',...
'BreakpointsForDimension1','loadbreakpoints',...
```

```
'BreakpointsForDimension2','speedbreakpoints',...
'Table','tBSFC');
```

c Examine the populated Lookup Table block.

Filtering and Smoothing Data

In this section...
"About Data Smoothing and Filtering" on page 6-30
"Moving Average Filtering" on page 6-30
"Savitzky-Golay Filtering" on page 6-32
"Local Regression Smoothing" on page 6-34
"Example: Smoothing Data" on page 6-40
"Example: Smoothing Data Using Loess and Robust Loess" on page 6-42

About Data Smoothing and Filtering

You can use the smooth function to smooth response data. You can use optional methods for moving average, Savitzky-Golay filters, and local regression with and without weights and robustness (lowess, loess, rlowess and rloess). These smoothing methods are discussed in the next five sections.

Moving Average Filtering

A moving average filter smooths data by replacing each data point with the average of the neighboring data points defined within the span. This process is equivalent to lowpass filtering with the response of the smoothing given by the difference equation

$$y_s(i) = \frac{1}{2N+1}\big(y(i+N) + y(i+N-1) + ... + y(i-N)\big)$$

where $y_s(i)$ is the smoothed value for the ith data point, N is the number of neighboring data points on either side of $y_s(i)$, and $2N+1$ is the span.

The moving average smoothing method used by Curve Fitting Toolbox follows these rules:

- The span must be odd.

- The data point to be smoothed must be at the center of the span.

- The span is adjusted for data points that cannot accommodate the specified number of neighbors on either side.

- The end points are not smoothed because a span cannot be defined.

Note that you can use `filter` function to implement difference equations such as the one shown above. However, because of the way that the end points are treated, the toolbox moving average result will differ from the result returned by `filter`. Refer to Difference Equations and Filtering in the MATLAB documentation for more information.

For example, suppose you smooth data using a moving average filter with a span of 5. Using the rules described above, the first four elements of y_s are given by

```
ys(1) = y(1)
ys(2) = (y(1)+y(2)+y(3))/3
ys(3) = (y(1)+y(2)+y(3)+y(4)+y(5))/5
ys(4) = (y(2)+y(3)+y(4)+y(5)+y(6))/5
```

Note that $y_s(1)$, $y_s(2)$, ... ,$y_s($`end`$)$ refer to the order of the data after sorting, and not necessarily the original order.

The smoothed values and spans for the first four data points of a generated data set are shown below.

Moving Average Smoothing
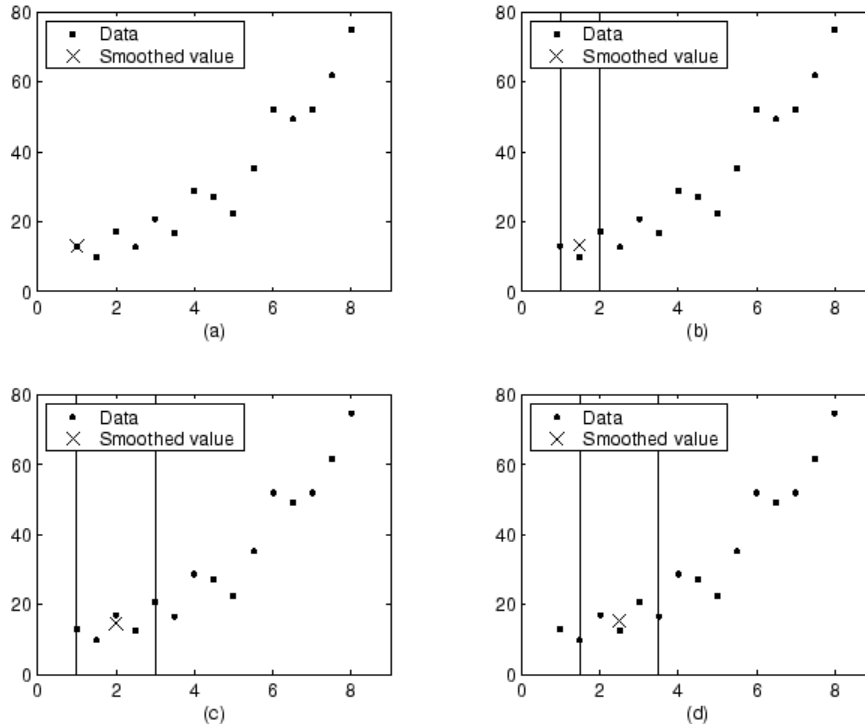
Plot (a) indicates that the first data point is not smoothed because a span cannot be constructed. Plot (b) indicates that the second data point is smoothed using a span of three. Plots (c) and (d) indicate that a span of five is used to calculate the smoothed value.

Savitzky-Golay Filtering

Savitzky-Golay filtering can be thought of as a generalized moving average. You derive the filter coefficients by performing an unweighted linear least-squares fit using a polynomial of a given degree. For this reason, a Savitzky-Golay filter is also called a digital smoothing polynomial filter or a least-squares smoothing filter. Note that a higher degree polynomial makes

it possible to achieve a high level of smoothing without attenuation of data features.

The Savitzky-Golay filtering method is often used with frequency data or with spectroscopic (peak) data. For frequency data, the method is effective at preserving the high-frequency components of the signal. For spectroscopic data, the method is effective at preserving higher moments of the peak such as the line width. By comparison, the moving average filter tends to filter out a significant portion of the signal's high-frequency content, and it can only preserve the lower moments of a peak such as the centroid. However, Savitzky-Golay filtering can be less successful than a moving average filter at rejecting noise.

The Savitzky-Golay smoothing method used by Curve Fitting Toolbox software follows these rules:

• The span must be odd.

• The polynomial degree must be less than the span.

• The data points are not required to have uniform spacing.

  Normally, Savitzky-Golay filtering requires uniform spacing of the predictor data. However, the Curve Fitting Toolbox algorithm supports nonuniform spacing. Therefore, you are not required to perform an additional filtering step to create data with uniform spacing.

The plot shown below displays generated Gaussian data and several attempts at smoothing using the Savitzky-Golay method. The data is very noisy and the peak widths vary from broad to narrow. The span is equal to 5% of the number of data points.
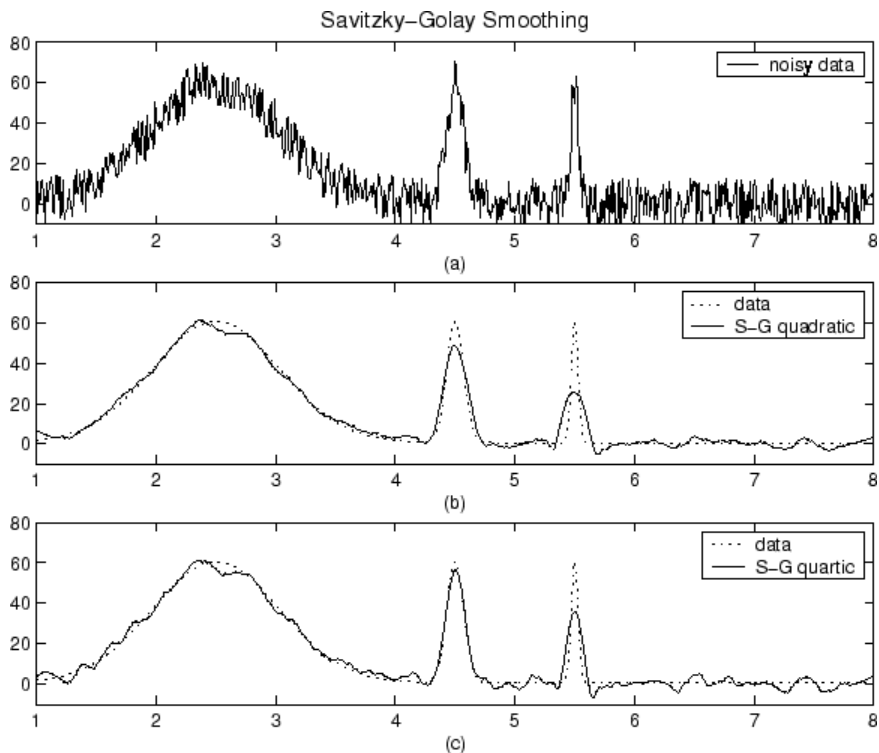
Plot (a) shows the noisy data. To more easily compare the smoothed results, plots (b) and (c) show the data without the added noise.

Plot (b) shows the result of smoothing with a quadratic polynomial. Notice that the method performs poorly for the narrow peaks. Plot (c) shows the result of smoothing with a quartic polynomial. In general, higher degree polynomials can more accurately capture the heights and widths of narrow peaks, but can do poorly at smoothing wider peaks.

Local Regression Smoothing

- "Lowess and Loess" on page 6-35
- "The Local Regression Method" on page 6-35
- "Robust Local Regression" on page 6-38

Lowess and Loess

The names "lowess" and "loess" are derived from the term "locally weighted scatter plot smooth," as both methods use locally weighted linear regression to smooth data.

The smoothing process is considered local because, like the moving average method, each smoothed value is determined by neighboring data points defined within the span. The process is weighted because a regression weight function is defined for the data points contained within the span. In addition to the regression weight function, you can use a robust weight function, which makes the process resistant to outliers. Finally, the methods are differentiated by the model used in the regression: lowess uses a linear polynomial, while loess uses a quadratic polynomial.

The local regression smoothing methods used by Curve Fitting Toolbox software follow these rules:

- The span can be even or odd.

- You can specify the span as a percentage of the total number of data points in the data set. For example, a span of 0.1 uses 10% of the data points.

The Local Regression Method

The local regression smoothing process follows these steps for each data point:

1 Compute the *regression weights* for each data point in the span. The weights are given by the tricube function shown below.

$$w_i = \left(1 - \left|\frac{x - x_i}{d(x)}\right|^3\right)^3$$

x is the predictor value associated with the response value to be smoothed, x_i are the nearest neighbors of x as defined by the span, and $d(x)$ is the distance along the abscissa from x to the most distant predictor value within the span. The weights have these characteristics:

- The data point to be smoothed has the largest weight and the most influence on the fit.

- Data points outside the span have zero weight and no influence on the fit.

2 A weighted linear least-squares regression is performed. For lowess, the regression uses a first degree polynomial. For loess, the regression uses a second degree polynomial.

3 The smoothed value is given by the weighted regression at the predictor value of interest.

If the smooth calculation involves the same number of neighboring data points on either side of the smoothed data point, the weight function is symmetric. However, if the number of neighboring points is not symmetric about the smoothed data point, then the weight function is not symmetric. Note that unlike the moving average smoothing process, the span never changes. For example, when you smooth the data point with the smallest predictor value, the shape of the weight function is truncated by one half, the leftmost data point in the span has the largest weight, and all the neighboring points are to the right of the smoothed value.

The weight function for an end point and for an interior point is shown below
for a span of 31 data points.

Local Regression Weight Function

The weight function for
the leftmost data point

The weight function for
an interior data point
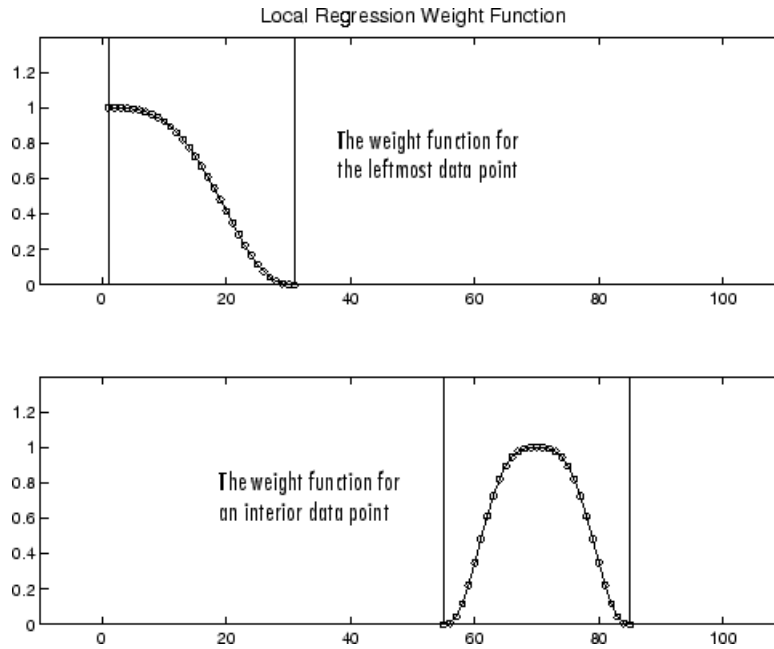
Using the lowess method with a span of five, the smoothed values and
associated regressions for the first four data points of a generated data set
are shown below.

Lowess Smoothing
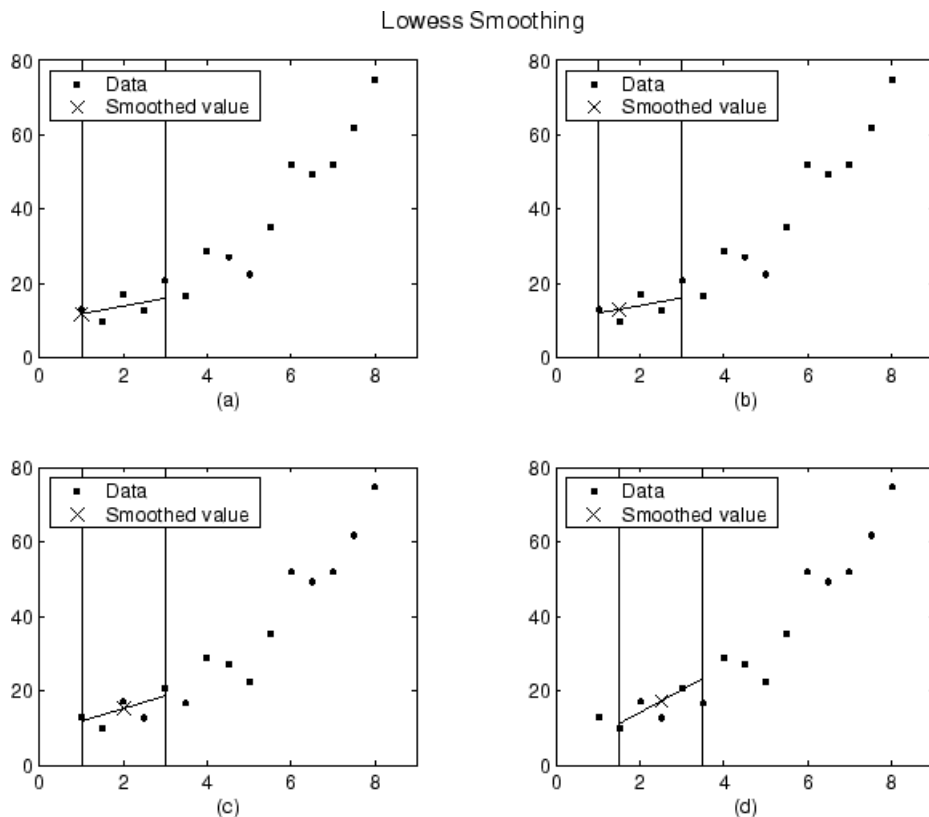
Notice that the span does not change as the smoothing process progresses from data point to data point. However, depending on the number of nearest neighbors, the regression weight function might not be symmetric about the data point to be smoothed. In particular, plots (a) and (b) use an asymmetric weight function, while plots (c) and (d) use a symmetric weight function.

For the loess method, the graphs would look the same except the smoothed value would be generated by a second-degree polynomial.

Robust Local Regression

If your data contains outliers, the smoothed values can become distorted, and not reflect the behavior of the bulk of the neighboring data points. To overcome this problem, you can smooth the data using a robust procedure that

is not influenced by a small fraction of outliers. For a description of outliers, refer to "Residual Analysis" on page 7-21.

Curve Fitting Toolbox software provides a robust version for both the lowess and loess smoothing methods. These robust methods include an additional calculation of robust weights, which is resistant to outliers. The robust smoothing procedure follows these steps:

1 Calculate the residuals from the smoothing procedure described in the previous section.

2 Compute the *robust weights* for each data point in the span. The weights are given by the bisquare function,

$$w_i = \begin{cases} \left(1 - (r_i / 6MAD)^2)\right)^2, & |ri| < 6MAD, \\ 0, & |ri| \geq 6MAD, \end{cases}$$

where r_i is the residual of the ith data point produced by the regression smoothing procedure, and *MAD* is the median absolute deviation of the residuals,

$$MAD = \text{median}\left(|r|\right).$$

The median absolute deviation is a measure of how spread out the residuals are. If r_i is small compared to $6MAD$, then the robust weight is close to 1. If r_i is greater than $6MAD$, the robust weight is 0 and the associated data point is excluded from the smooth calculation.

3 Smooth the data again using the robust weights. The final smoothed value is calculated using both the local regression weight and the robust weight.

4 Repeat the previous two steps for a total of five iterations.

The smoothing results of the lowess procedure are compared below to the results of the robust lowess procedure for a generated data set that contains a single outlier. The span for both procedures is 11 data points.

Robust Lowess Smoothing
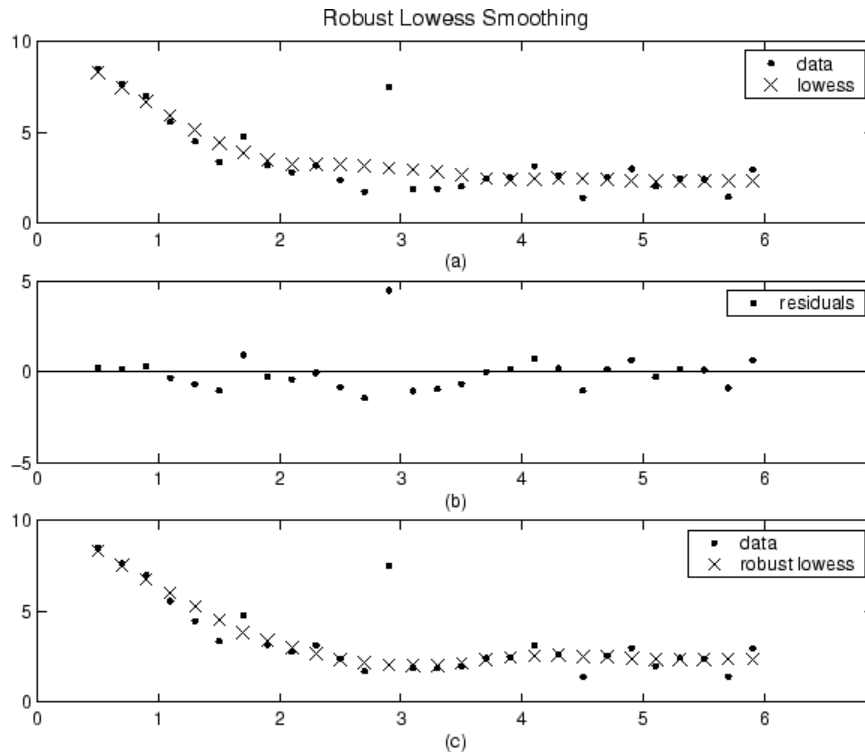
Plot (a) shows that the outlier influences the smoothed value for several nearest neighbors. Plot (b) suggests that the residual of the outlier is greater than six median absolute deviations. Therefore, the robust weight is zero for this data point. Plot (c) shows that the smoothed values neighboring the outlier reflect the bulk of the data.

Example: Smoothing Data

Load the data in count.dat:

```
load count.dat
```

The 24-by-3 array count contains traffic counts at three intersections for each hour of the day.

First, use a moving average filter with a 5-hour span to smooth all of the data at once (by linear index) :

```
c = smooth(count(:));
C1 = reshape(c,24,3);
```

Plot the original data and the smoothed data:

```
subplot(3,1,1)
plot(count,':');
hold on
plot(C1,'-');
title('Smooth C1 (All Data)')
```

Second, use the same filter to smooth each column of the data separately:

```
C2 = zeros(24,3);
for I = 1:3,
    C2(:,I) = smooth(count(:,I));
end
```

Again, plot the original data and the smoothed data:

```
subplot(3,1,2)
plot(count,':');
hold on
plot(C2,'-');
title('Smooth C2 (Each Column)')
```

Plot the difference between the two smoothed data sets:

```
subplot(3,1,3)
plot(C2 - C1,'o-')
title('Difference C2 - C1')
```
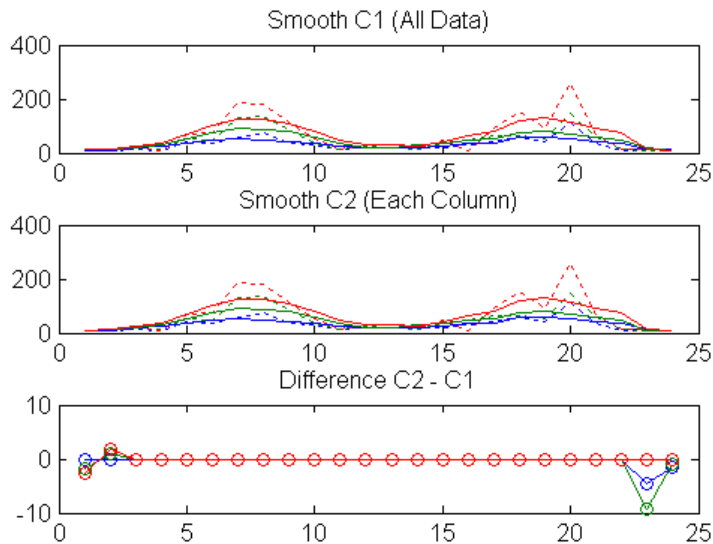
Note the additional end effects from the 3-column smooth.

Example: Smoothing Data Using Loess and Robust Loess

Create noisy data with outliers:

```
x = 15*rand(150,1);
y = sin(x) + 0.5*(rand(size(x))-0.5);
y(ceil(length(x)*rand(2,1))) = 3;
```

Smooth the data using the loess and rloess methods with a span of 10%:

```
yy1 = smooth(x,y,0.1,'loess');
yy2 = smooth(x,y,0.1,'rloess');
```

Plot original data and the smoothed data.

```
[xx,ind] = sort(x);
subplot(2,1,1)
plot(xx,y(ind),'b.',xx,yy1(ind),'r-')
set(gca,'YLim',[-1.5 3.5])
legend('Original Data','Smoothed Data Using ''loess''',...
```

```
        'Location','NW')
subplot(2,1,2)
plot(xx,y(ind),'b.',xx,yy2(ind),'r-')
set(gca,'YLim',[-1.5 3.5])
legend('Original Data','Smoothed Data Using ''rloess''',...
        'Location','NW')
```


Note that the outliers have less influence on the robust method.

7
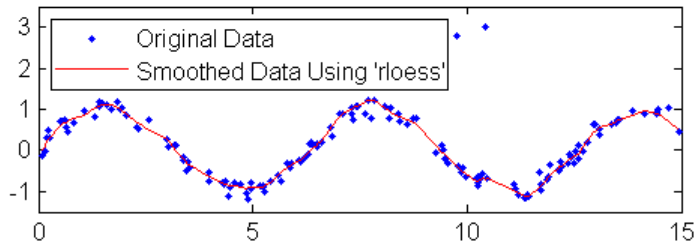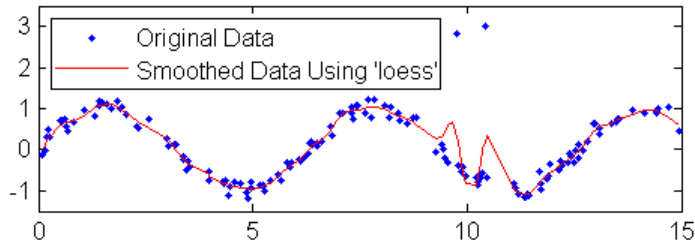
Fit Postprocessing

Exploring and Customizing Plots

Displaying Fit and Residual Plots

Within each fit figure, you can display up to three plots simultaneously to examine the fit. Use the toolbar or **View** menu to select the type of plot to display:

- **Main Plot** shows the curve or surface fit.

- **Residuals Plot** shows the errors between your fit and your data

- **Contour Plot** shows a contour map of a surface fit (not available for curve fits).

The next example shows a main plot with a curve fit and prediction bounds, and the residuals plot.
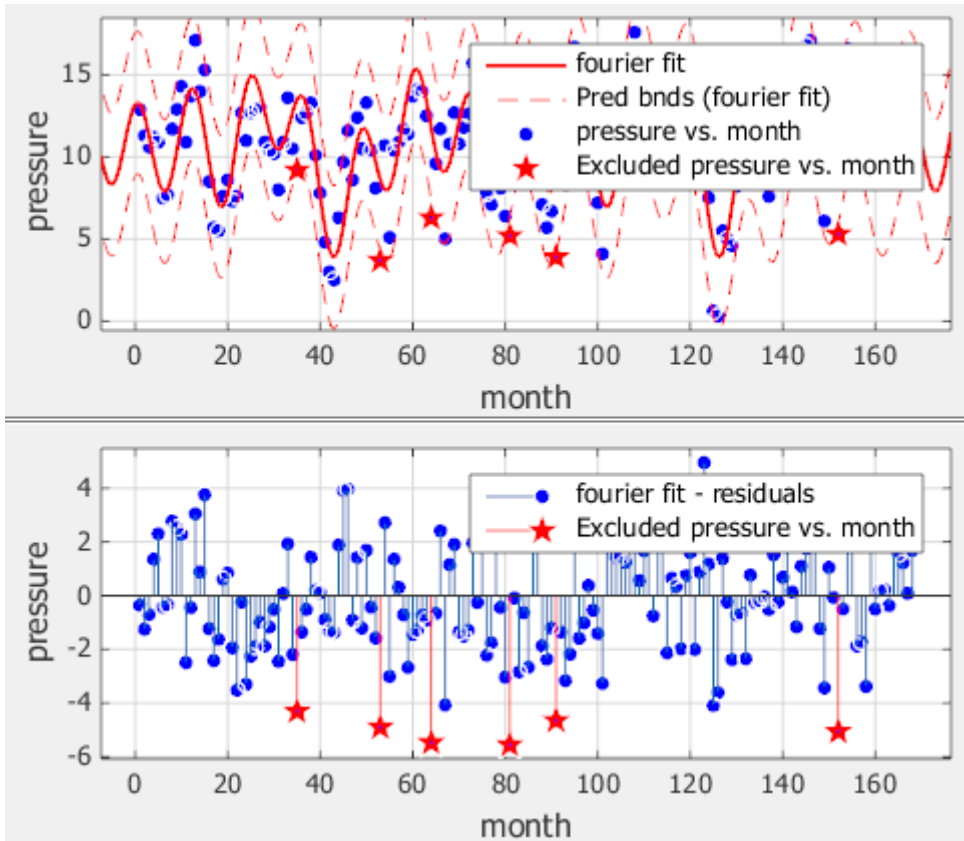
When you select **Tools > Prediction Bounds**, two additional curves (or surfaces) are plotted to show the prediction bounds on both sides of your model fit.

Choose which bounds to display: None, 90%, 95%, 99%, or Custom. The custom option opens a dialog box where you can enter the required confidence level.

See also "Customizing the Fit Display" on page 7-6.

Residuals Plot

On the residuals plot, you can view the errors between your fit and your data, and you can remove outliers. See "Removing Outliers" on page 7-10. This example shows a residuals plot with some excluded outliers.
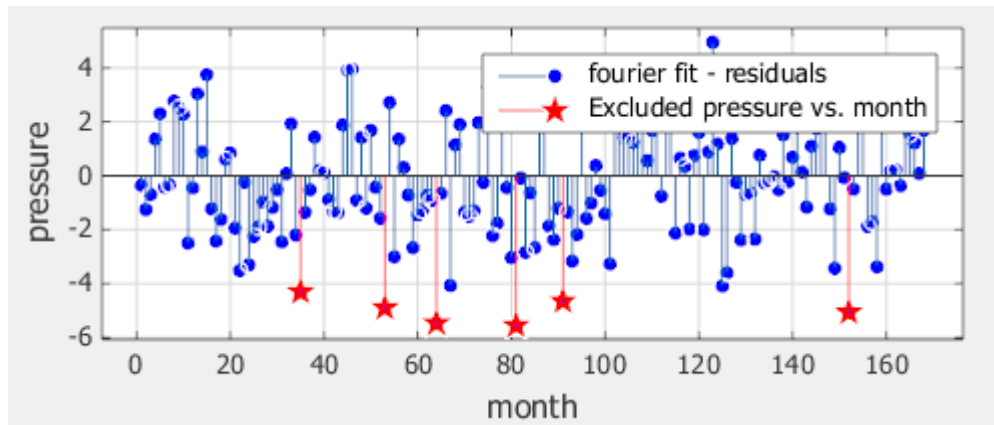
Viewing Surface Plots and Contour Plots

If you fit a surface, then the main plot shows your surface fit. Click-and-drag rotation or **Rotate 3D** is the default mouse mode for surface plots in the Curve Fitting Tool. Rotate mode in the Curve Fitting Tool is the same as **Rotate 3D** in MATLAB figures. You can change the mouse mode for manipulating plots just as for curve plots. See "Using Zoom, Pan, Data Cursor, and Outlier Exclusion" on page 7-6.

Tip To return to rotate mode, turn off any other mouse mode.

If you turn on a mouse mode for zoom, pan, data cursor or exlude outliers, turn the mode off again to return to rotate mode. For example, click the Zoom in toolbar button a second time to clear it and return to rotate mode.

If you have a surface fit, use the contour plot to examine a contour map of your surface. Contour plots are not available for curve fits. On a surface fit,

a contour plot makes it easier to see points that have the same height. An example follows.
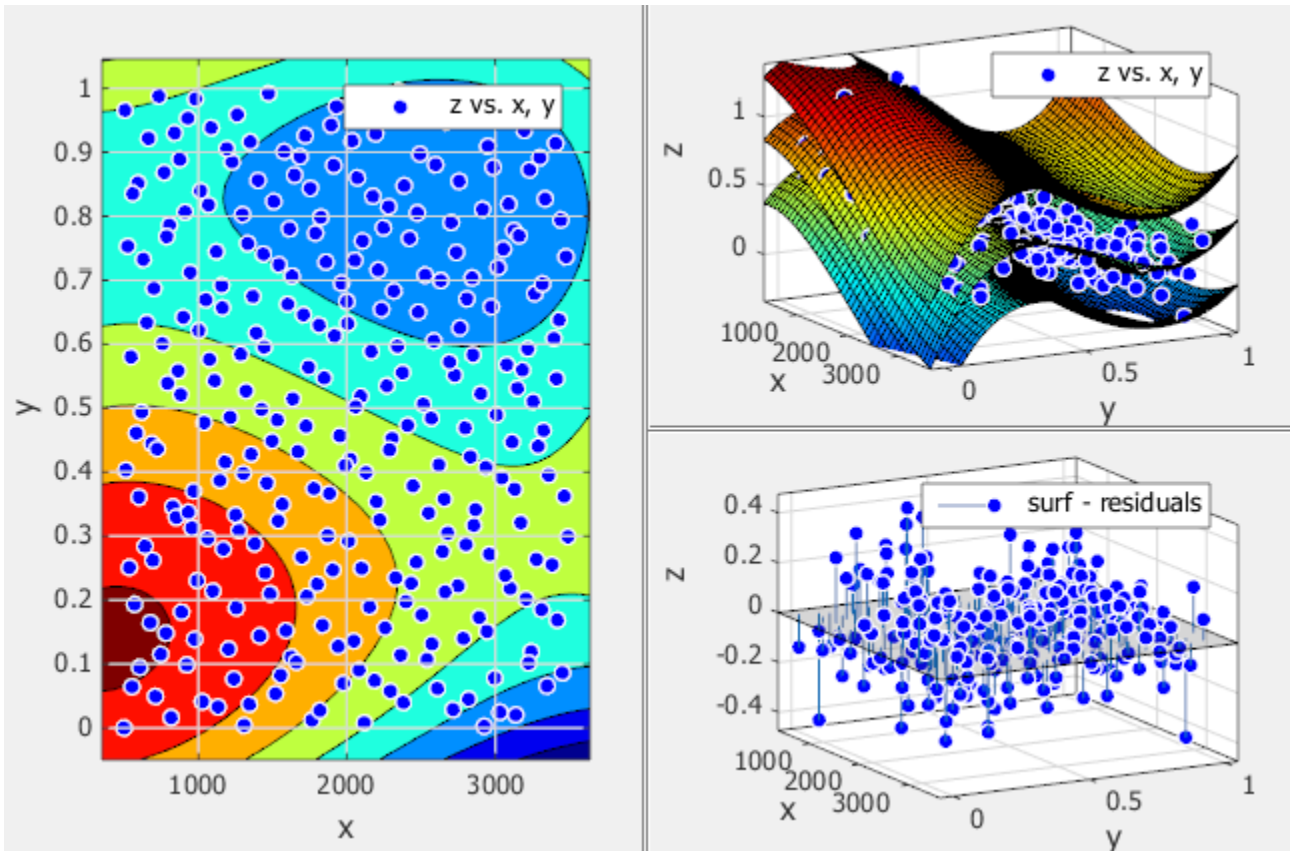
For polynomial and custom fits, you also can use the **Tools** menu to display prediction bounds. When you display prediction bounds, two additional surfaces are plotted to show the prediction bounds on both sides of your model fit. The previous example shows prediction bounds. You can see three surfaces on the plot. The top and bottom surfaces show the prediction bounds at the specified confidence level on either side of your model fit surface.

You can also switch your surface plot to a 2-D plot if desired. Your plot cursor must be in rotate mode. Clear any other mouse mode if necessary.

Then, right-click the plot to select **X-Y**, **X-Z**, or **Y-Z** view, or to select **Rotate Options**. All these context menu options are standard MATLAB 3-D plot tools. See "Rotate in 3-D" in the MATLAB Graphics documentation.

Using Zoom, Pan, Data Cursor, and Outlier Exclusion

You can change mouse mode for manipulating plots. Use the toolbar or **Tools** menu to switch to **Zoom**, **Pan**, **Data Cursor**, or **Exclude Outliers** modes.

The Curve Fitting Tool remembers your selected mouse mode in each fit figure within a session.

Use the toolbar or **Tools** menu to toggle mouse mode in your plots:

- **Zoom In**, **Zoom Out**, **Pan**, and **Data Cursor** are standard MATLAB plot tools.

  On surfaces, turn all these modes off to return to **Rotate 3D** mode. For surface plots, rotation is the default mouse mode in the Curve Fitting Tool. See "Viewing Surface Plots and Contour Plots" on page 7-4.

- — **Data Cursor** selects data cursor mode, where you can click points to display input and output values.

- — **Exclude Outliers** selects outlier mode, where you can click points to remove or include in your fit. Exclude outliers is a mouse mode for graphically excluding data from your fit. See "Removing Outliers" on page 7-10.

Customizing the Fit Display

To customize your plot display, use the toolbar, **Tools** menu, or the **View** menu. See also "Interactive Fit Comparison" on page 2-13.

Tools Menu and Toolbar

- — **Legend** toggles display of the legend on all plots in the currently selected fit tab.

- — **Grid** toggles display of the grid on all plots in the currently selected fit tab.

- **Tools > Prediction Bounds** lets you choose which bounds to display: None, 90%, 95%, 99%, or Custom. The custom option opens a dialog box where you can enter the required confidence level.

- — **Axes Limits** opens a dialog box where you can specify upper and lower bounds for the X- and Y-axes of plots. Click **Reset** to return to the default axes limits.

View Menu and Toolbar

Use the View controls to customize the display to show or hide the plots, fit settings, results and table of fits.

- Available in the **View** menu and the toolbar:

  - — **Main Plot** toggles the display of the main fit plot in the currently selected fit figure. This item is disabled if only the main plot is displayed.

  - — **Residuals Plot** toggles the display of the residuals plot in the currently selected fit tab. This item is disabled if only the residuals plot is displayed.

  - — **Contour Plot** toggles the display of the contour plot in the currently selected fit tab. This item is disabled if only the contour plot is displayed.

- **View > Fit Settings** toggles the display of the fit controls pane in the currently selected fit tab (**Fit name**, inputs, fit type, and so on).

- **View > Fit Results** toggles the display of the **Results** pane in the currently selected fit tab. When you display the **Results** pane, you can see model terms and coefficients, goodness-of-fit statistics, and information messages about the fit.

- **View > Table of Fits** toggles the display of the **Table of Fits** pane in the Curve Fitting Tool.

Tip For more space to view and compare plots, as shown next, use the **View** menu to hide or show the **Fit Settings**, **Fit Results**, or **Table of Fits** panes.
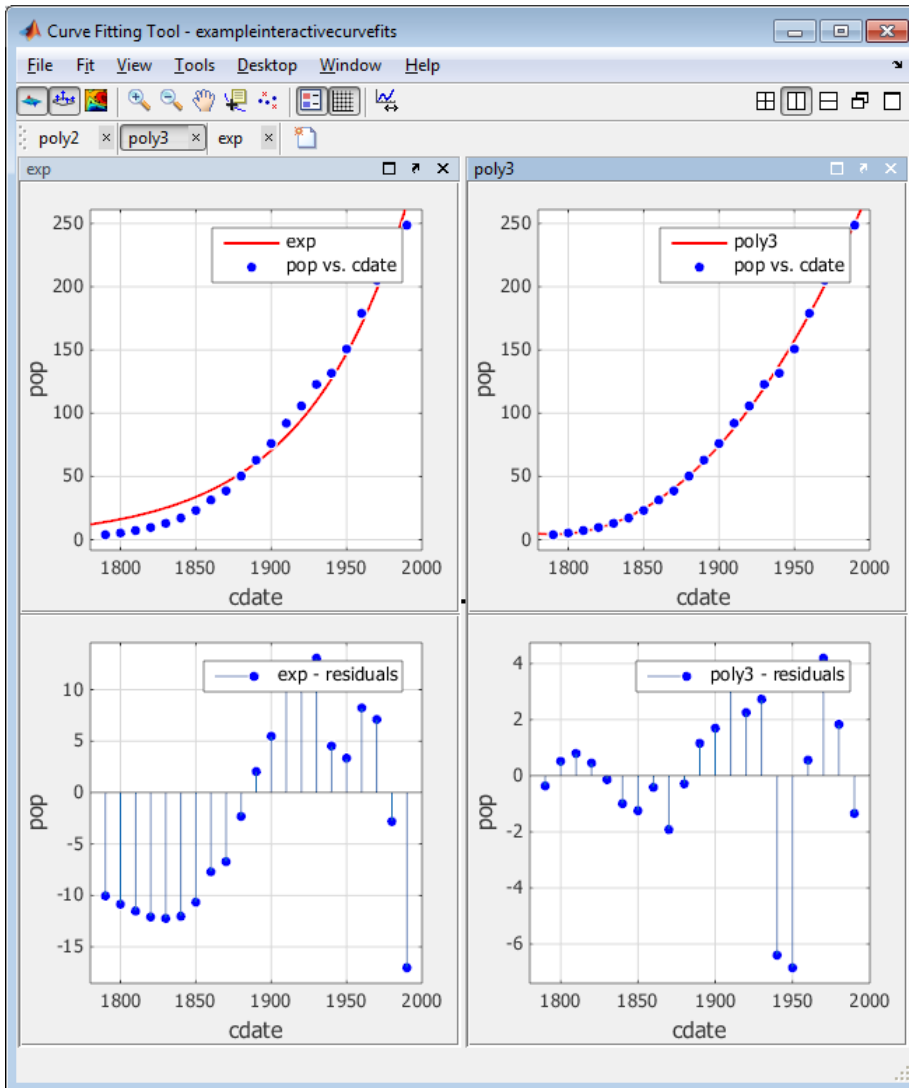
See also "Displaying Multiple Fits Simultaneously" on page 2-14.

Print to MATLAB Figures

In Curve Fitting Tool, select **File > Print to Figure** to produce MATLAB figures from the results of curve fitting. **Print to Figure** creates a figure containing all plots for the current fit. You can then use the interactive plotting tools to edit the figures showing fitting results for presentation purposes, and export these in different formats.

Removing Outliers

To remove outliers, follow these steps:

1 Select **Tools > Exclude Outliers** or click the toolbar button .

   When you move the mouse cursor to the plot, it changes to a cross-hair to show you are in outlier selection mode.

2 Click a point that you want to exclude in the main plot or residuals plot. Alternatively, click and drag to define a rectangle and remove all enclosed points.

   A removed plot point becomes a red star in the plots. If you have **Auto-fit** selected, the Curve Fitting Tool refits the surface without the point. Otherwise, you can click **Fit** to refit.

3 Repeat for all points you want to exclude.

When removing outliers from surface fits, it can be helpful to display a 2-D residuals plot for examining and removing outliers. With your plot cursor in rotation mode, right-click the plot to select **X-Y**, **X-Z**, or **Y-Z** view.

To replace individual excluded points in the fit, click an excluded point again in **Exclude Outliers** mode. To replace all excluded points in the fit, right-click and select **Clear all exclusions**.

In surface plots, to return to rotation mode, click the **Exclude outliers** toolbar button again to turn off outlier selection mode.

Selecting Validation Data

To specify validation data for the currently selected fit, follow these steps:

1 Select **Fit > Specify Validation Data**. The Specify Validation Data dialog box opens.

2 Select variables for **X data** and **Y data** (and **Z data** for surfaces).

When you select two or three variables, depending on whether your fit data is for a curve or a surface, the tool calculates validation statistics (SSE and RMSE) and displays them in the **Results** pane and the **Table of Fits**. For definitions of these statistics, see "Using the Statistics in the Table of Fits" on page 2-18. Your validation data points display on the main plot and residual plot along with the original data.

3 Close the dialog box.

Generating Code and Exporting Fits to the Workspace

In this section...
"Generating Code from the Curve Fitting Tool" on page 7-12
"Exporting a Fit to the Workspace" on page 7-13

Generating Code from the Curve Fitting Tool

You can generate and use MATLAB code from an interactive session in the Curve Fitting Tool. In this way, you can transform your interactive analysis into reusable functions for batch processing of multiple data sets. You can use the generated file without modification, or you can edit and customize the file as needed.

To generate code for all fits and plots in your Curve Fitting Tool session follow these steps:

1 Select **File > Generate Code**.

The Curve Fitting Tool generates code from your session and displays the file in the MATLAB Editor. The file includes all fits and plots in your current session. The file captures the following information:

- Names of fits and their variables

- Fit settings and options

- Plots

- Curve and surface fitting objects and methods used to create the fits:

    - A cell-array of cfit or sfit objects representing the fits

    - A structure array with goodness-of fit information.

2 Save the file.

To recreate your fits and plots, call the file from the command line with your original data as input arguments. You also can call the file with new data.

For example, enter:

```
[fitresult, gof] = myFileName(a, b, c)
```

where a, b, and c are your variable names, and *myFileName* is the file name.

Calling the file from the command line does *not* recreate your Curve Fitting Tool GUI and session. When you call the file, you get the same plots you had in your Curve Fitting Tool session in standard MATLAB figure windows. There is one window for each fit. For example, if your fit in the Curve Fitting Tool session displayed main, residual and contour plots, all three plots appear in a single figure window.

Curve Fitting Objects and Methods

The curve and surface fit objects (cfit and sfit) store the results from a fitting operation, making it easy to plot and analyze fits at the command line.

cfit and sfit objects are a subclass of fittype objects, so they inherit all the same methods of fittype and provide additional methods. To learn about available methods, see cfit, sfit and fittype.

Exporting a Fit to the Workspace

To export a fit to the MATLAB workspace, follow these steps:

1 Select a fit and save it to the MATLAB workspace using one of these methods:

- Right-click the fit listed in the Table of Fits and select **Save Fit** *myfitname* **to Workspace**

- Select a fit figure in the Curve Fitting Tool and select **Fit > Save to Workspace**.
  The Save Fit to MATLAB Workspace dialog box opens.
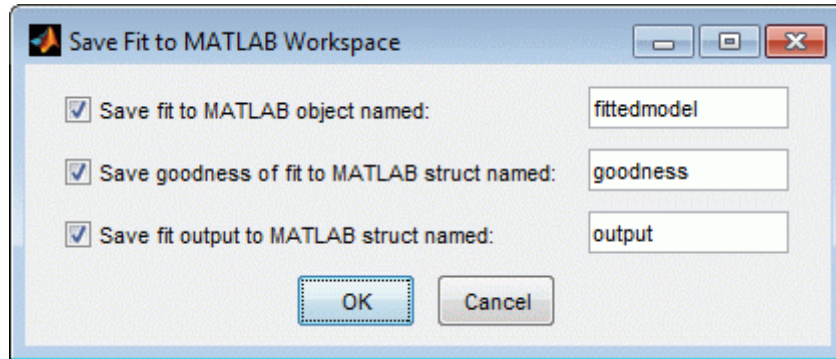
2 Edit the names as appropriate. If you previously exported fits, the toolbox automatically adds a numbered suffix to the default names so there is no danger of overwriting them.

3 Choose which options you want to export by selecting the check boxes. Check box options are as follows:

- **Save fit to MATLAB object named *fittedmodel*** — This option creates a cfit or sfit object, that encapsulates the result of fitting a curve or surface to data. You can examine the fit coefficients at the command line, for example:

```
fittedmodel
   Linear model Poly22:
      fittedmodel1(x,y) = p00 + p10*x + p01*y + p20*x^2...
                             + p11*x*y + p02*y^2
    Coefficients (with 95% confidence bounds):
      p00 =       302.1  (247.3, 356.8)
      p10 =        -1395  (-1751, -1039)
      p01 =      0.03525  (0.01899, 0.05151)
      p20 =         1696  (1099, 2293)
      p11 =      -0.1119  (-0.1624, -0.06134)
      p02 =   2.36e-006  (-8.72e-007, 5.593e-006)
```

  You also can treat the cfit or sfit object as a function to make predictions or evaluate the fit at values of X (or X and Y). See the cfit and sfit reference page.

- **Save goodness of fit to MATLAB struct named *goodness*** — This option creates a structure array that contains statistical information about the fit, for example:

```
goodness =
          sse: 0.0234
      rsquare: 0.9369
          dfe: 128
    adjrsquare: 0.9345
         rmse: 0.0135
```

- **Save fit output to MATLAB struct named *output*** — This option creates a structure array that contains information such as numbers of observations and parameters, residuals, and so on. For example:

```
output =
         numobs: 134
       numparam: 6
      residuals: [134x1 double]
       Jacobian: [134x6 double]
       exitflag: 1
      algorithm: 'QR factorization and solve'
     iterations: 1
```

Note **Goodness of fit** and **Output** arrays are outputs of the fit function. See the fit reference page.

4 Click **OK** to save the fit options to the workspace.

After you save your fit to the workspace, you can use fit postprocessing functions. For an example, see "Analyzing Your Best Fit in the Workspace" on page 2-32. For more information and a list of functions, see "Fit Postprocessing".

Evaluating Goodness of Fit

How to Evaluate Goodness of Fit

After fitting data with one or more models, you should evaluate the goodness of fit. A visual examination of the fitted curve displayed in Curve Fitting Tool should be your first step. Beyond that, the toolbox provides these methods to assess goodness of fit for both linear and nonlinear parametric fits:

- "Goodness-of-Fit Statistics" on page 7-17
- "Residual Analysis" on page 7-21
- "Confidence and Prediction Bounds" on page 7-28

As is common in statistical literature, the term *goodness of fit* is used here in several senses: A "good fit" might be a model

- that your data could reasonably have come from, given the assumptions of least-squares fitting
- in which the model coefficients can be estimated with little uncertainty
- that explains a high proportion of the variability in your data, and is able to predict new observations with high certainty

A particular application might dictate still other aspects of model fitting that are important to achieving a good fit, such as a simple model that is easy to interpret. The methods described here can help you determine goodness of fit in all these senses.

These methods group into two types: graphical and numerical. Plotting residuals and prediction bounds are graphical methods that aid visual interpretation, while computing goodness-of-fit statistics and coefficient confidence bounds yield numerical measures that aid statistical reasoning.

Generally speaking, graphical measures are more beneficial than numerical measures because they allow you to view the entire data set at once, and they can easily display a wide range of relationships between the model and the data. The numerical measures are more narrowly focused on a particular aspect of the data and often try to compress that information into a single number. In practice, depending on your data and analysis requirements, you might need to use both types to determine the best fit.

Note that it is possible that none of your fits can be considered suitable for your data, based on these methods. In this case, it might be that you need to select a different model. It is also possible that all the goodness-of-fit measures indicate that a particular fit is suitable. However, if your goal is to extract fitted coefficients that have physical meaning, but your model does not reflect the physics of the data, the resulting coefficients are useless. In this case, understanding what your data represents and how it was measured is just as important as evaluating the goodness of fit.

Goodness-of-Fit Statistics

After using graphical methods to evaluate the goodness of fit, you should examine the goodness-of-fit statistics. Curve Fitting Toolbox software supports these goodness-of-fit statistics for parametric models:

- The sum of squares due to error (SSE)

- R-square

- Adjusted R-square

- Root mean squared error (RMSE)

For the current fit, these statistics are displayed in the **Results** list box in the **Fit Editor**. For all fits in the current curve-fitting session, you can compare the goodness-of-fit statistics in the **Table of fits**.

Sum of Squares Due to Error

This statistic measures the total deviation of the response values from the fit to the response values. It is also called the summed square of residuals and is usually labeled as *SSE*.

$$SSE = \sum_{i=1}^{n} w_i \left(y_i - \hat{y}_i \right)^2$$

A value closer to 0 indicates that the model has a smaller random error component, and that the fit will be more useful for prediction.

R-Square

This statistic measures how successful the fit is in explaining the variation of the data. Put another way, R-square is the square of the correlation between the response values and the predicted response values. It is also called the square of the multiple correlation coefficient and the coefficient of multiple determination.

R-square is defined as the ratio of the sum of squares of the regression (*SSR*) and the total sum of squares (*SST*). *SSR* is defined as

$$SSR = \sum_{i=1}^{n} w_i \left(\hat{y}_i - \bar{y} \right)^2$$

SST is also called the sum of squares about the mean, and is defined as

$$SST = \sum_{i=1}^{n} w_i \left(y_i - \bar{y} \right)^2$$

where *SST* = *SSR* + *SSE*. Given these definitions, R-square is expressed as

$$\text{R-square} = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

R-square can take on any value between 0 and 1, with a value closer to 1 indicating that a greater proportion of variance is accounted for by the model. For example, an R-square value of 0.8234 means that the fit explains 82.34% of the total variation in the data about the average.

If you increase the number of fitted coefficients in your model, R-square will increase although the fit may not improve in a practical sense. To avoid this situation, you should use the degrees of freedom adjusted R-square statistic described below.

Note that it is possible to get a negative R-square for equations that do not contain a constant term. Because R-square is defined as the proportion of variance explained by the fit, if the fit is actually worse than just fitting a horizontal line then R-square is negative. In this case, R-square cannot be interpreted as the square of a correlation. Such situations indicate that a constant term should be added to the model.

Degrees of Freedom Adjusted R-Square

This statistic uses the R-square statistic defined above, and adjusts it based on the residual degrees of freedom. The residual degrees of freedom is defined as the number of response values n minus the number of fitted coefficients m estimated from the response values.

$$v = n - m$$

v indicates the number of independent pieces of information involving the n data points that are required to calculate the sum of squares. Note that if parameters are bounded and one or more of the estimates are at their bounds, then those estimates are regarded as fixed. The degrees of freedom is increased by the number of such parameters.

The adjusted R-square statistic is generally the best indicator of the fit quality when you compare two models that are *nested* — that is, a series of models each of which adds additional coefficients to the previous model.

$$\text{adjusted R-square} = 1 - \frac{SSE(n-1)}{SST(v)}$$

The adjusted R-square statistic can take on any value less than or equal to 1, with a value closer to 1 indicating a better fit. Negative values can occur when the model contains terms that do not help to predict the response.

Root Mean Squared Error

This statistic is also known as the fit standard error and the standard error of the regression. It is an estimate of the standard deviation of the random component in the data, and is defined as

$$RMSE = s = \sqrt{MSE}$$

where *MSE* is the mean square error or the residual mean square

$$MSE = \frac{SSE}{v}$$

Just as with *SSE*, an *MSE* value closer to 0 indicates a fit that is more useful for prediction.

Residual Analysis

Plotting and Analysing Residuals

The residuals from a fitted model are defined as the differences between the response data and the fit to the response data at each predictor value.

residual = data – fit

You display the residuals in Curve Fitting Tool by selecting the toolbar button or menu item **View > Residuals Plot**.

Mathematically, the residual for a specific predictor value is the difference between the response value y and the predicted response value $\hat{y}$.

$r = y - \hat{y}$

Assuming the model you fit to the data is correct, the residuals approximate the random errors. Therefore, if the residuals appear to behave randomly, it suggests that the model fits the data well. However, if the residuals display a systematic pattern, it is a clear sign that the model fits the data poorly. Always bear in mind that many results of model fitting, such as confidence bounds, will be invalid should the model be grossly inappropriate for the data.

A graphical display of the residuals for a first degree polynomial fit is shown below. The top plot shows that the residuals are calculated as the vertical distance from the data point to the fitted curve. The bottom plot displays the residuals relative to the fit, which is the zero line.
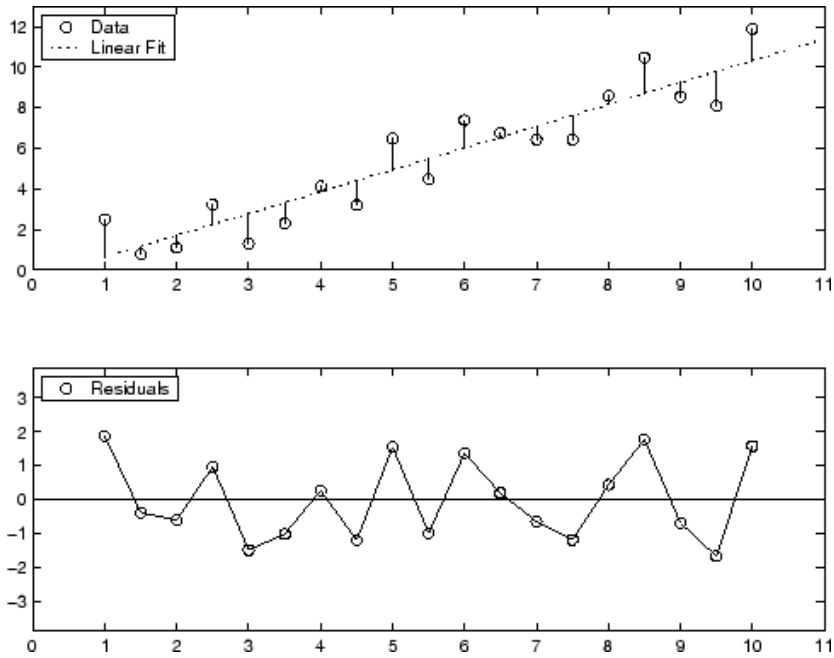
The residuals appear randomly scattered around zero indicating that the model describes the data well.

A graphical display of the residuals for a second-degree polynomial fit is shown below. The model includes only the quadratic term, and does not include a linear or constant term.
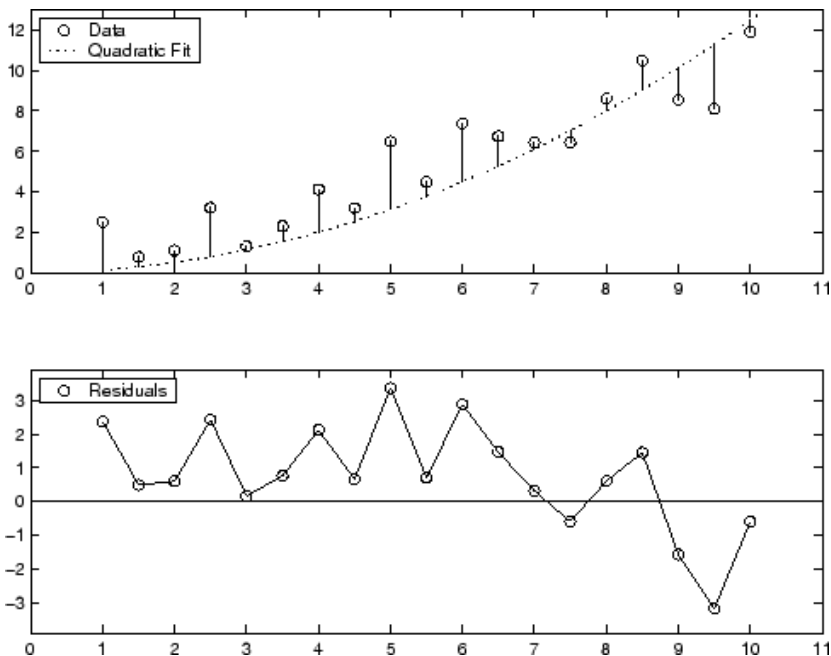
The residuals are systematically positive for much of the data range indicating that this model is a poor fit for the data.

Example: Residual Analysis

This example fits several polynomial models to generated data and evaluates how well those models fit the data and how precisely they can predict. The data is generated from a cubic curve, and there is a large gap in the range of the *x* variable where no data exist.

```
x = [1:0.1:3 9:0.1:10]';
c = [2.5 -0.5 1.3 -0.1];
y = c(1) + c(2)*x + c(3)*x.^2 + c(4)*x.^3 + (rand(size(x))-0.5);
```

Fit the data in the Curve Fitting Tool using a cubic polynomial and a fifth degree polynomial. The data, fits, and residuals are shown below. Display the residuals in the Curve Fitting Tool by selecting **View > Residuals Plot**.
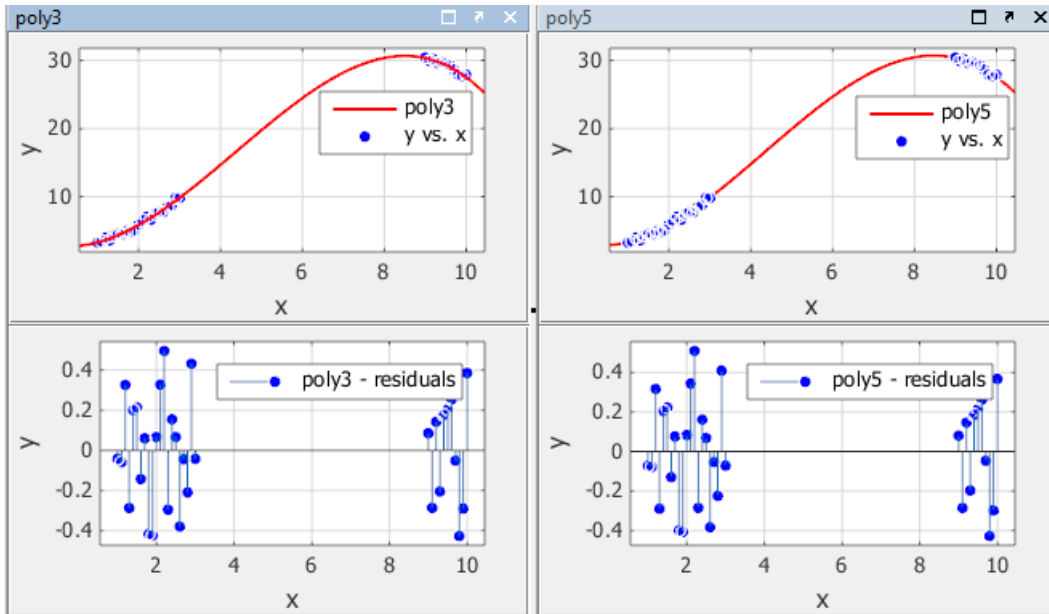
Both models appear to fit the data well, and the residuals appear to be randomly distributed around zero. Therefore, a graphical evaluation of the fits does not reveal any obvious differences between the two equations.

Look at the numerical fit results in the **Results** pane and compare the confidence bounds for the coefficients.

The results show that the cubic fit coefficients are accurately known (bounds are small), while the quintic fit coefficients are not accurately known. As expected, the fit results for `poly3` are reasonable because the generated data follows a cubic curve. The 95% confidence bounds on the fitted coefficients indicate that they are acceptably precise. However, the 95% confidence bounds for `poly5` indicate that the fitted coefficients are not known precisely.

The goodness-of-fit statistics are shown in the **Table of Fits**. By default, the adjusted R-square and RMSE statistics are displayed in the table. The statistics do not reveal a substantial difference between the two equations. To choose statistics to display or hide, right-click the column headers.

The 95% nonsimultaneous prediction bounds for new observations are shown below. To display prediction bounds in Curve Fitting Tool, select **Tools > Prediction Bounds > 95%**.
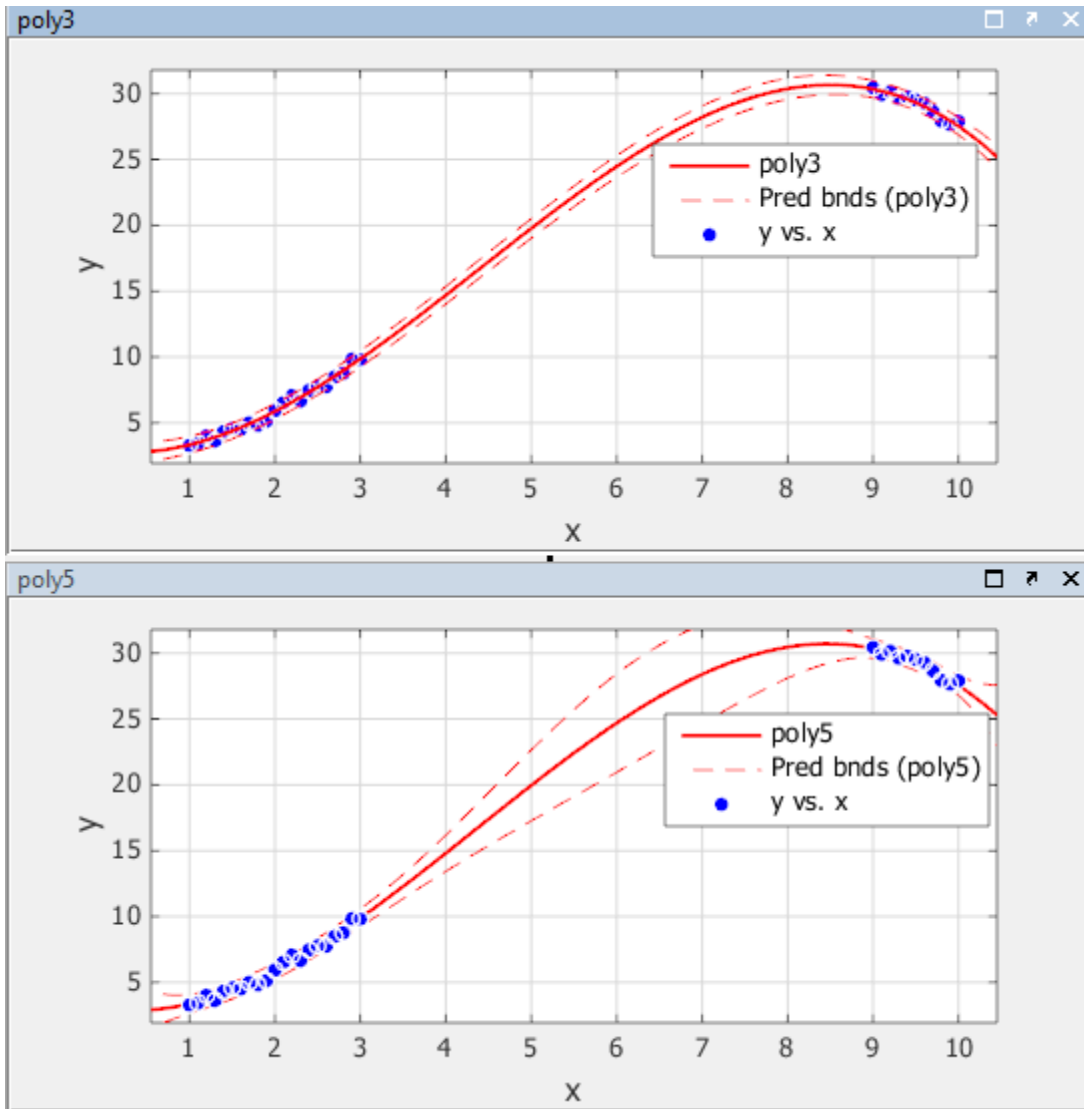
The prediction bounds for `poly3` indicate that new observations can be predicted with a small uncertainty throughout the entire data range. This is not the case for `poly5`. It has wider prediction bounds in the area where no data exist, apparently because the data does not contain enough information

to estimate the higher degree polynomial terms accurately. In other words, a fifth-degree polynomial overfits the data. You can confirm this by using the Analysis GUI to compute bounds for the functions themselves.

The 95% prediction bounds for the fitted function using `poly5` are shown below. As you can see, the uncertainty in predicting the function is large in the center of the data. Therefore, you would conclude that more data must be collected before you can make precise predictions using a fifth-degree polynomial.

In conclusion, you should examine all available goodness-of-fit measures before deciding on the fit that is best for your purposes. A graphical examination of the fit and residuals should always be your initial approach. However, some fit characteristics are revealed only through numerical fit results, statistics, and prediction bounds.

Confidence and Prediction Bounds

About Confidence and Prediction Bounds

Curve Fitting Toolbox software lets you calculate confidence bounds for the fitted coefficients, and prediction bounds for new observations or for the fitted function. Additionally, for prediction bounds, you can calculate simultaneous bounds, which take into account all predictor values, or you can calculate nonsimultaneous bounds, which take into account only individual predictor values. The coefficient confidence bounds are presented numerically, while the prediction bounds are displayed graphically and are also available numerically.

The available confidence and prediction bounds are summarized below.

Types of Confidence and Prediction Bounds

Interval Type	Description
Fitted coefficients	Confidence bounds for the fitted coefficients
New observation	Prediction bounds for a new observation (response value)
New function	Prediction bounds for a new function value

Note Prediction bounds are also often described as confidence bounds because you are calculating a confidence interval for a predicted response.

Confidence and prediction bounds define the lower and upper values of the associated interval, and define the width of the interval. The width of the interval indicates how uncertain you are about the fitted coefficients, the predicted observation, or the predicted fit. For example, a very wide interval for the fitted coefficients can indicate that you should use more data when fitting before you can say anything very definite about the coefficients.

The bounds are defined with a level of certainty that you specify. The level of certainty is often 95%, but it can be any value such as 90%, 99%, 99.9%, and so on. For example, you might want to take a 5% chance of being incorrect

about predicting a new observation. Therefore, you would calculate a 95% prediction interval. This interval indicates that you have a 95% chance that the new observation is actually contained within the lower and upper prediction bounds.

Confidence Bounds on Coefficients

The confidence bounds for fitted coefficients are given by

$$C = b \pm t\sqrt{S}$$

where b are the coefficients produced by the fit, t depends on the confidence level, and is computed using the inverse of Student's t cumulative distribution function, and S is a vector of the diagonal elements from the estimated covariance matrix of the coefficient estimates, $(X^TX)^{-1}s^2$. In a linear fit, X is the design matrix, while for a nonlinear fit X is the Jacobian of the fitted values with respect to the coefficients. X^T is the transpose of X, and s^2 is the mean squared error.

The confidence bounds are displayed in the **Results** pane in the Curve Fitting Tool using the following format.

```
p1 = 1.275  (1.113, 1.437)
```

The fitted value for the coefficient p1 is 1.275, the lower bound is 1.113, the upper bound is 1.437, and the interval width is 0.324. By default, the confidence level for the bounds is 95%.

You can calculate confidence intervals at the command line with the confint function.

Prediction Bounds on Fits

As mentioned previously, you can calculate prediction bounds for the fitted curve. The prediction is based on an existing fit to the data. Additionally, the bounds can be simultaneous and measure the confidence for all predictor values, or they can be nonsimultaneous and measure the confidence only for a single predetermined predictor value. If you are predicting a new observation, nonsimultaneous bounds measure the confidence that the new observation lies within the interval given a single predictor value. Simultaneous bounds

measure the confidence that a new observation lies within the interval regardless of the predictor value.

The nonsimultaneous prediction bounds for a new observation at the predictor value x are given by

$$P_{n,o} = y \pm t\sqrt{s^2 + xSx^T}$$

where s^2 is the mean squared error, t depends on the confidence level, and is computed using the inverse of Student's t cumulative distribution function, and S is the covariance matrix of the coefficient estimates, $(X^TX)^{-1}s^2$. Note that x is defined as a row vector of the design matrix or Jacobian evaluated at a specified predictor value.

The simultaneous prediction bounds for a new observation and for all predictor values are given by

$$P_{s,o} = y \pm f\sqrt{s^2 + xSx^T}$$

where f depends on the confidence level, and is computed using the inverse of the F cumulative distribution function.

The nonsimultaneous prediction bounds for the function at a single predictor value x are given by

$$P_{n,f} = y \pm t\sqrt{xSx^T}$$

The simultaneous prediction bounds for the function and for all predictor values are given by

$$P_{s,f} = y \pm f\sqrt{xSx^T}$$

You can graphically display prediction bounds using Curve Fitting Tool. With Curve Fitting Tool, you can display nonsimultaneous prediction bounds for new observations with **Tools > Prediction Bounds**. By default, the

confidence level for the bounds is 95%. You can change this level to any value with **Tools > Prediction Bounds > Custom**.

You can display numerical prediction bounds of any type at the command line with the `predint` function.

To understand the quantities associated with each type of prediction interval, recall that the data, fit, and residuals are related through the formula

data = fit + residuals

where the fit and residuals terms are estimates of terms in the formula

data = model + random error

Suppose you plan to take a new observation at the predictor value x_{n+1}. Call the new observation $y_{n+1}(x_{n+1})$ and the associated error ε_{n+1}. Then

$$y_{n+1}(x_{n+1}) = f(x_{n+1}) + \varepsilon_{n+1}$$

where $f(x_{n+1})$ is the true but unknown function you want to estimate at x_{n+1}. The likely values for the new observation or for the estimated function are provided by the nonsimultaneous prediction bounds.

If instead you want the likely value of the new observation to be associated with any predictor value, the previous equation becomes

$$y_{n+1}(x) = f(x) + \varepsilon$$

The likely values for this new observation or for the estimated function are provided by the simultaneous prediction bounds.

The types of prediction bounds are summarized below.

Types of Prediction Bounds

Type of Bound	Simultaneous or Non-simultaneous	Associated Equation
Observation	Non-simultaneous	$y_{n+1}(x_{n+1})$
	Simultaneous	$y_{n+1}(x)$, for all x
Function	Non-simultaneous	$f(x_{n+1})$
	Simultaneous	$f(x)$, for all x

The nonsimultaneous and simultaneous prediction bounds for a new observation and the fitted function are shown below. Each graph contains three curves: the fit, the lower confidence bounds, and the upper confidence bounds. The fit is a single-term exponential to generated data and the bounds reflect a 95% confidence level. Note that the intervals associated with a new observation are wider than the fitted function intervals because of the additional uncertainty in predicting a new response value (the curve plus random errors).
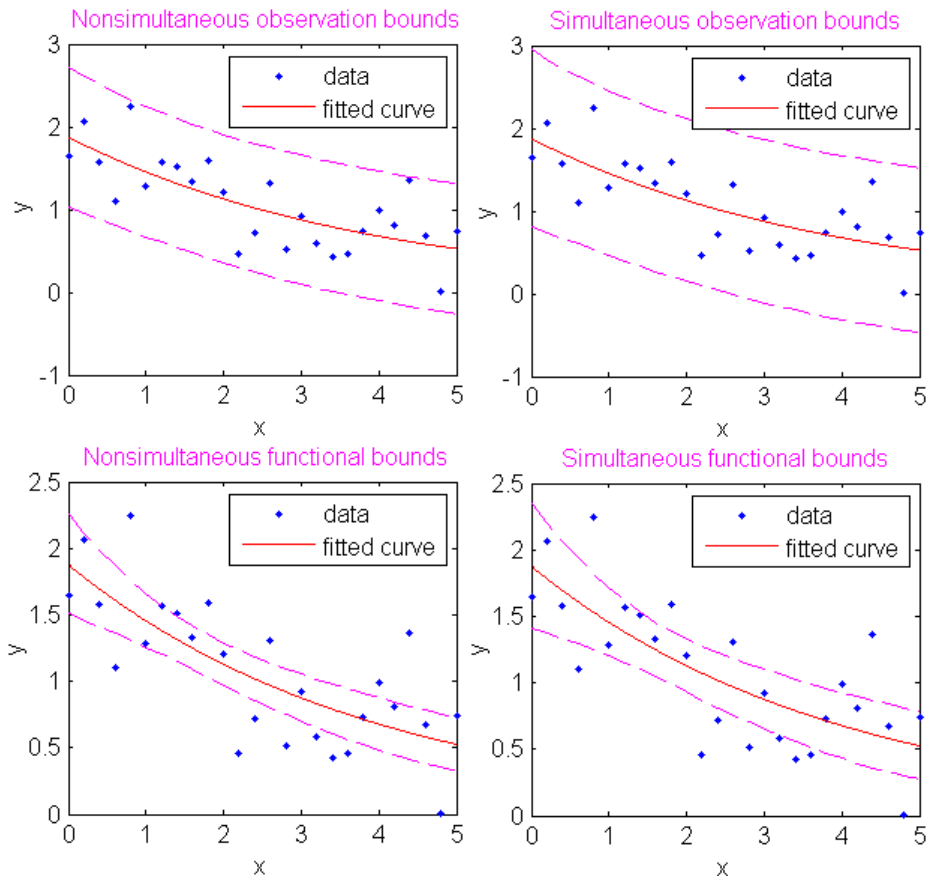
Example: Prediction Intervals

The following example computes and plots prediction intervals at the command line.

Generate data with an exponential trend:

```
x = (0:0.2:5)';
y = 2*exp(-0.2*x) + 0.5*randn(size(x));
```

Fit the data using a single-term exponential:

```
fitresult = fit(x,y,'exp1');
```

Compute prediction intervals:

```
p11 = predint(fitresult,x,0.95,'observation','off');
p12 = predint(fitresult,x,0.95,'observation','on');
p21 = predint(fitresult,x,0.95,'functional','off');
p22 = predint(fitresult,x,0.95,'functional','on');
```

Plot the data, fit, and prediction intervals:

```
subplot(2,2,1)
plot(fitresult,x,y), hold on, plot(x,p11,'m--'), xlim([O 5])
title('Nonsimultaneous observation bounds','Color','m')
subplot(2,2,2)
plot(fitresult,x,y), hold on, plot(x,p12,'m--'), xlim([O 5])
title('Simultaneous observation bounds','Color','m')
subplot(2,2,3)
plot(fitresult,x,y), hold on, plot(x,p21,'m--'), xlim([O 5])
title('Nonsimultaneous functional bounds','Color','m')
subplot(2,2,4)
plot(fitresult,x,y), hold on, plot(x,p22,'m--'), xlim([O 5])
title('Simultaneous functional bounds','Color','m')
```

Differentiating and Integrating a Fit

Create a baseline sinusoidal signal:

```
xdata = (0:.1:2*pi)';
y0 = sin(xdata);
```

Add noise to the signal:

```
noise = 2*y0.*randn(size(y0)); % Response-dependent noise
ydata = y0 + noise;
```

Fit the noisy data with a custom sinusoidal model:

```
f = fittype('a*sin(b*x)');
fit1 = fit(xdata,ydata,f,'StartPoint',[1 1]);
```

Find the derivatives of the fit at the predictors:

```
[d1,d2] = differentiate(fit1,xdata);
```

Plot the data, the fit, and the derivatives:

```
subplot(3,1,1)
plot(fit1,xdata,ydata) % cfit plot method
subplot(3,1,2)
plot(xdata,d1,'m') % double plot method
grid on
legend('1st derivative')
subplot(3,1,3)
plot(xdata,d2,'c') % double plot method
grid on
legend('2nd derivative')
```
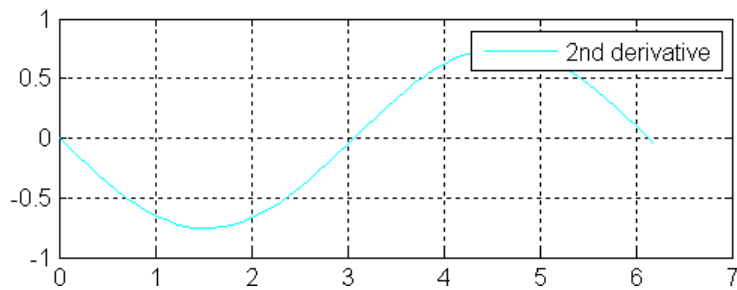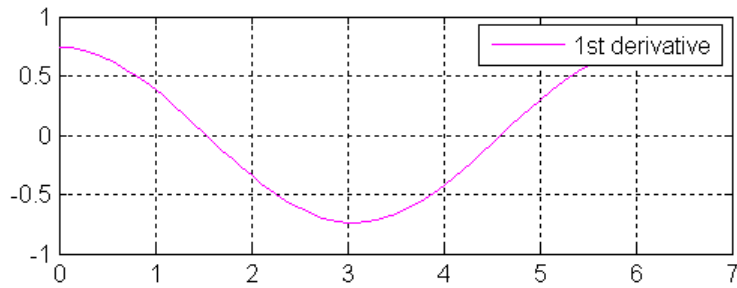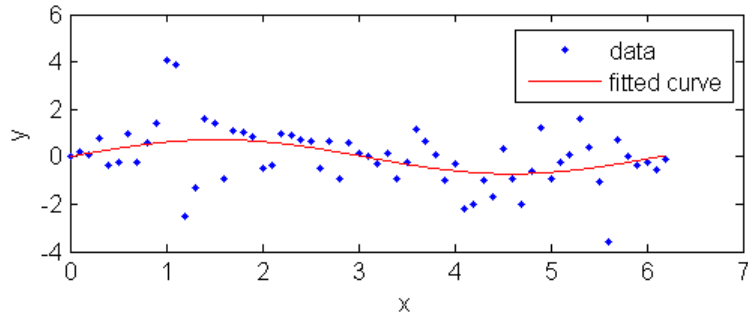
Note that derivatives can also be computed and plotted directly with the cfit plot method, as follows:

```
plot(fit1,xdata,ydata,{'fit','deriv1','deriv2'})
```

The plot method, however, does not return data on the derivatives.

Find the integral of the fit at the predictors:

```
int = integrate(fit1,xdata,0);
```

Plot the data, the fit, and the integral:

```
subplot(2,1,1)
plot(fit1,xdata,ydata) % cfit plot method
subplot(2,1,2)
plot(xdata,int,'m') % double plot method
grid on
legend('integral')
```


Note that integrals can also be computed and plotted directly with the `cfit` `plot` method, as follows:

```
plot(fit1,xdata,ydata,{'fit','integral'})
```

The `plot` method, however, does not return data on the integral.

Spline Fitting

About Splines

Introducing Spline Fitting

In this section...
"About Splines in Curve Fitting Toolbox" on page 8-2
"Spline Overview" on page 8-3
"Interactive Spline Fitting" on page 8-3
"Programmatic Spline Fitting" on page 8-4

About Splines in Curve Fitting Toolbox

You can work with splines in Curve Fitting Toolbox in several ways.

Using Curve Fitting Tool or the `fit` function you can:

- Fit cubic spline interpolants to curves or surfaces

- Fit smoothing splines and shape-preserving cubic spline interpolants to curves (but not surfaces)

The toolbox also contains specific splines functions to allow greater control over what you can create. For example, you can use the `csapi` function for cubic spline interpolation. Why would you use `csapi` instead of the `fit` function `'cubicinterp'` option? You might require greater flexibility to work with splines for the following reasons:

- You want to combine the results with other splines, e.g., by addition.

- You want vector-valued splines. You can use `csapi` with scalars, vectors, matrices, and ND-arrays. The `fit` function only allows scalar-valued splines.

- You want other types of splines such as ppform, B-form, tensor-product, rational, and stform thin-plate splines.

- You want to create splines without data.

- You want to specify breaks, optimize knot placement, and use specialized functions for spline manipulation such as differentiation and integration.

If you require specialized spline functions, see the following sections for an overview of splines, and interactive and programmatic spline fitting.

Spline Overview

The Curve Fitting Toolbox spline functions are a collection of tools for creating, viewing, and analyzing spline approximations of data. *Splines* are smooth piecewise polynomials that can be used to represent functions over large intervals, where it would be impractical to use a single approximating polynomial.

The spline functionality includes a graphical user interface (GUI) that provides easy access to functions for creating, visualizing, and manipulating splines. The toolbox also contains functions that enable you to evaluate, plot, combine, differentiate, and integrate splines. Because all toolbox functions are implemented in the open MATLAB language, you can inspect the algorithms, modify the source code, and create your own custom functions.

Key spline features:

- GUIs that let you create, view, and manipulate splines and manage and compare spline approximations

- Functions for advanced spline operations, including differentiation, integration, break/knot manipulation, and optimal knot placement

- Support for piecewise polynomial form (ppform) and basis form (B-form) splines

- Support for tensor-product splines and rational splines (including NURBS)

Interactive Spline Fitting

You can access all spline functions from the `splinetool` GUI. You can use the GUI to:

- Vary spline parameters and tolerances

- View and modify data, breaks, knots, and weights

- View the error of the spline, or the spline's first or second derivative

- Observe the toolbox commands that generated your spline

- Create and import data, including built-in instructive data sets, and save splines to the workspace

See `splinetool`.

Programmatic Spline Fitting

To programmatically fit splines, see "Construction" for descriptions of types of splines and numerous examples.

Curve Fitting Toolbox Splines and MATLAB Splines

In this section...

Curve Fitting Toolbox Splines

Curve Fitting Toolbox spline functions contain versions of the essential MATLAB programs of the B-spline package (extended to handle also *vector*-valued splines) as described in *A Practical Guide to Splines,* (Applied Math. Sciences Vol. 27, Springer Verlag, New York (1978), xxiv + 392p; revised edition (2001), xviii+346p), hereafter referred to as *PGS*. The toolbox makes it easy to create and work with piecewise-polynomial functions.

The typical use envisioned for this toolbox involves the construction and subsequent use of a piecewise-polynomial approximation. This construction would involve data fitting, but there is a wide range of possible data that could be fit. In the simplest situation, one is given points (t_i, y_i) and is looking for a piecewise-polynomial function f that satisfies $f(t_i) = y_i$, all i, more or less. An exact fit would involve *interpolation*, an approximate fit might involve *least-squares approximation* or the *smoothing spline*. But the function to be approximated may also be described in more implicit ways, for example as the solution of a differential or integral equation. In such a case, the data would be of the form $(Af)(t_i)$, with A some differential or integral operator. On the other hand, one might want to construct a spline *curve* whose exact location is less important than is its overall shape. Finally, in all of this, one might be looking for functions of more than one variable, such as *tensor product splines*.

Care has been taken to make this work as painless and intuitive as possible. In particular, the user need not worry about just how splines are constructed

or stored for later use, nor need the casual user worry about such items as "breaks" or "knots" or "coefficients". It is enough to know that each function constructed is just another variable that is freely usable as input (where appropriate) to many of the commands, including all commands beginning with `fn,` which stands for `function`. At times, it may be also useful to know that, internal to the toolbox, splines are stored in different forms, with the command `fn2fm` available to convert between forms.

At present, the toolbox supports two major forms for the representation of piecewise-polynomial functions, because each has been found to be superior to the other in certain common situations. The B-form is particularly useful during the construction of a spline, while the ppform is more efficient when the piecewise-polynomial function is to be evaluated extensively. These two forms are almost exactly the B-representation and the pp representation used in *A Practical Guide to Splines*.

But, over the years, the Curve Fitting Toolbox spline functions have gone beyond the programs in *A Practical Guide to Splines*. The toolbox now supports the 'scattered translates' form, or stform, in order to handle the construction and use of bivariate thin-plate splines, and also two ways to represent rational splines, the rBform and the rpform, in order to handle NURBS.

Splines can be very effective for data fitting because the linear systems to be solved for this are banded, hence the work needed for their solution, done properly, grows only linearly with the number of data points. In particular, the MATLAB sparse matrix facilities are used in the Curve Fitting Toolbox spline functions when that is more efficient than the toolbox's own equation solver, `slvblk`, which relies on the fact that some of the linear systems here are even almost block diagonal.

All polynomial spline construction commands are equipped to produce bivariate (or even multivariate) piecewise-polynomial functions as tensor products of the univariate functions used here, and the various `fn...` commands also work for these multivariate functions.

There are various examples, all accessible through the Demos tab in the MATLAB Help browser. You are strongly urged to have a look at some of them, or at the GUI `splinetool`, before attempting to use this toolbox, or even before reading on.

MATLAB Splines

The MATLAB technical computing environment provides spline approximation via the command `spline`. If called in the form `cs = spline(x,y)`, it returns the ppform of the cubic spline with break sequence `x` that takes the value `y(i)` at `x(i)`, all `i`, and satisfies the not-a-knot end condition. In other words, the command `cs = spline(x,y)` gives the same result as the command `cs = csapi(x,y)` available in the Curve Fitting Toolbox spline functions. But only the latter also works when `x,y` describe multivariate gridded data. In MATLAB, cubic spline interpolation to multivariate gridded data is provided by the command `interpn(x1,...,xd,v,y1,...,yd,'spline')` which returns values of the interpolating tensor product cubic spline at the grid specified by `y1,...,yd`.

Further, any of the Curve Fitting Toolbox spline `fn...` commands can be applied to the output of the MATLAB `spline(x,y)` command, with simple versions of the Curve Fitting Toolbox spline commands `fnval`, `ppmak`, `fnbrk` available directly in MATLAB, as the commands `ppval`, `mkpp`, `unmkpp`, respectively.

Expected Background

The Curve Fitting Toolbox spline functions started out as an extension of the MATLAB environment of interest to experts in spline approximation, to aid them in the construction and testing of new methods of spline approximation. Such people will have mastered the material in *A Practical Guide to Splines*.

However, the basic commands for constructing and using spline approximations are set up to be usable with no more knowledge than it takes to understand what it means to, say, construct an interpolant or a least squares approximant to some data, or what it means to differentiate or integrate a function.

With that in mind, there are sections, like "Cubic Spline Interpolation" on page 9-2, that are meant even for the novice, while sections devoted to a detailed example, like the one on constructing a Chebyshev spline or on constructing and using tensor products, are meant for users interested in developing their own spline commands.

A "Glossary" at the end of this guide provides definitions of almost all the mathematical terms used in this document.

Vector Data Type Support

The Curve Fitting Toolbox spline functions can handle *vector*-valued splines, i.e., splines whose values lie in R^d. Since MATLAB started out with just one variable type, that of a matrix, there is even now some uncertainty about how to deal with vectors, i.e., lists of numbers. MATLAB sometimes stores such a list in a matrix with just one row, and other times in a matrix with just one column. In the first instance, such a *1-row matrix* is called a row-vector; in the second instance, such a *1-column matrix* is called a column-vector. Either way, these are merely different ways for *storing* vectors, not different *kinds* of vectors.

In this toolbox, *vectors*, i.e., lists of numbers, may also end up stored in a 1-row matrix or in a 1-column matrix, but with the following agreements.

A point in R^d, i.e., a d-vector, is always stored as a column vector. In particular, if you want to supply an n-list of d-vectors to one of the commands, you are expected to provide that list as the n columns of a matrix of size `[d,n]`.

While other lists of numbers (e.g., a knot sequence or a break sequence) may be stored internally as row vectors, you may supply such lists as you please, as a row vector or a column vector.

Spline Function Naming Conventions

Most of the spline commands in this toolbox have names that follow one of the following patterns:

`cs...` commands construct cubic splines (in ppform)

`sp...` commands construct splines in B-form

`fn...` commands operate on spline functions

`..2..` commands convert something

`..api` commands construct an approximation by interpolation

`..aps` commands construct an approximation by smoothing

`..ap2` commands construct a least-squares approximation

...`knt` commands construct (part of) a particular knot sequence

...`dem` commands are examples.

Note See the "Glossary" for information about notation used in this book.

Arguments for Curve Fitting Toolbox Spline Functions

For ease of use, most Curve Fitting Toolbox spline functions have default arguments. In the reference entry under Syntax, we usually first list the function with all *necessary* input arguments and then with all *possible* input arguments. When there is more than one optional argument, then, sometimes, but not always, their exact order is immaterial. When their order does matter, you have to specify every optional argument preceding the one(s) you are interested in. In this situation, you can specify the default value for an optional argument by using `[]` (the empty matrix) as the input for it. The description in the reference page tells you the default value for each optional input argument.

As in MATLAB, only the output arguments explicitly specified are returned to the user.

Acknowledgments

MathWorks® would like to acknowledge the contributions of **Carl de Boor** to the Curve Fitting Toolbox spline functions. Professor de Boor authored the Spline Toolbox™ from its first release until Version 3.3.4 (2008).

Professor de Boor received the John von Neumann Prize in 1996 and the National Medal of Science in 2003. He is a member of both the American Academy of Arts and Sciences and the National Academy of Sciences. He is the author of *A Practical Guide to Splines* (Springer, 2001).

Some of the spline function naming conventions are the result of a discussion with Jörg Peters, then a graduate student in Computer Sciences at the University of Wisconsin-Madison.

9

Simple Spline Examples

Cubic Spline Interpolation

Cubic Spline Interpolant of Smooth Data

Suppose you want to interpolate some smooth data, e.g., to

```
rand('seed',6), x = (4*pi)*[0 1 rand(1,15)]; y = sin(x);
```

You can use the cubic spline interpolant obtained by

```
cs = csapi(x,y);
```

and plot the spline, along with the data, with the following code:

```
fnplt(cs);
hold on
plot(x,y,'o')
legend('cubic spline','data')
hold off
```

This produces a figure like the following.
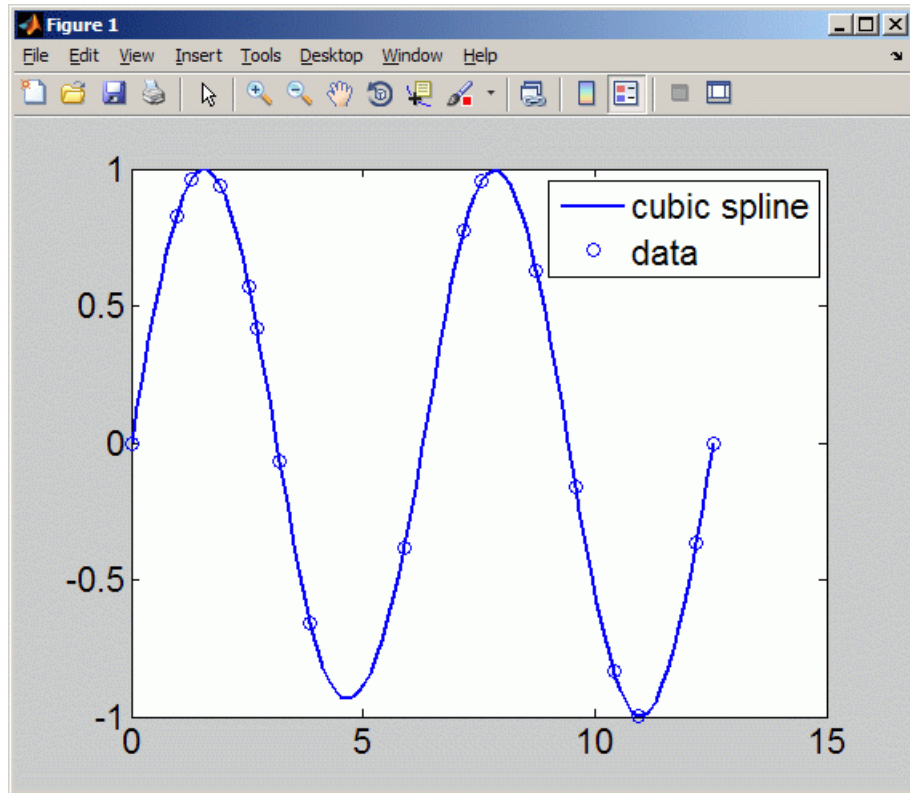
Cubic Spline Interpolant of Smooth Data

This is, more precisely, the cubic spline interpolant with the not-a-knot end conditions, meaning that it is the unique piecewise cubic polynomial with two continuous derivatives with breaks at all *interior* data sites except for the leftmost and the rightmost one. It is the same interpolant as produced by the MATLAB spline command, spline(x,y).

Periodic Data

The sine function is 2π-periodic. To check how well your interpolant does on that score, compute, e.g., the difference in the value of its first derivative at the two endpoints,

```
diff(fnval(fnder(cs),[0 4*pi]))
```

```
ans = -.0100
```

which is not so good. If you prefer to get an interpolant whose first and second derivatives at the two endpoints, 0 and 4*pi, match, use instead the command csape which permits specification of many different kinds of end conditions, including periodic end conditions. So, use instead

```
pcs = csape(x,y,'periodic');
```

for which you get

```
diff(fnval(fnder(pcs),[0 4*pi]))
```

Output is ans = 0 as the difference of end slopes. Even the difference in end second derivatives is small:

```
diff(fnval(fnder(pcs,2),[0 4*pi]))
```

Output is ans = -4.6074e-015.

Other End Conditions

Other end conditions can be handled as well. For example,

```
cs = csape(x,[3,y,-4],[1 2]);
```

provides the cubic spline interpolant with breaks at the $x(i)$ and with its slope at the leftmost data site equal to 3, and its second derivative at the rightmost data site equal to -4.

General Spline Interpolation

If you want to interpolate at sites other than the breaks and/or by splines other than cubic splines with simple knots, then you use the spapi command. In its simplest form, you would say sp = spapi(k,x,y); in which the first argument, k, specifies the *order* of the interpolating spline; this is the number of coefficients in each polynomial piece, i.e., 1 more than the nominal degree of its polynomial pieces. For example, the next figure shows a linear, a quadratic, and a quartic spline interpolant to your data, as obtained by the statements

```
sp2 = spapi(2,x,y); fnplt(sp2,2), hold on
```

```
sp3 = spapi(3,x,y); fnplt(sp3,2,'k--'), set(gca,'Fontsize',16)
sp5 = spapi(5,x,y); fnplt(sp5,2,'r-.'), plot(x,y,'o')
legend('linear','quadratic','quartic','data'), hold off
```
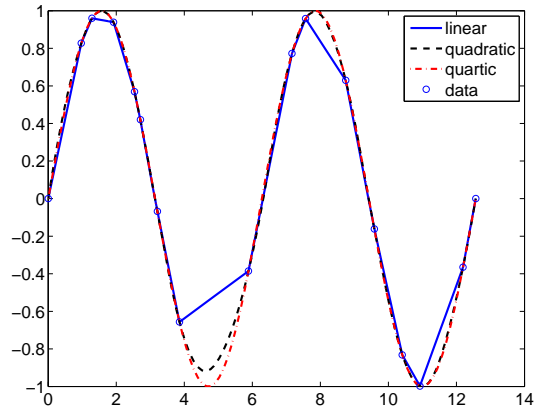

Spline Interpolants of Various Orders of Smooth Data

Even the cubic spline interpolant obtained from spapi is different from the one provided by csapi and spline. To emphasize their difference, compute and plot their second derivatives, as follows:

```
fnplt(fnder(spapi(4,x,y),2)), hold on, set(gca,'Fontsize',16)
fnplt(fnder(csapi(x,y),2),2,'k--'),plot(x,zeros(size(x)),'o')
legend('from spapi','from csapi','data sites'), hold off
```

This gives the following graph:
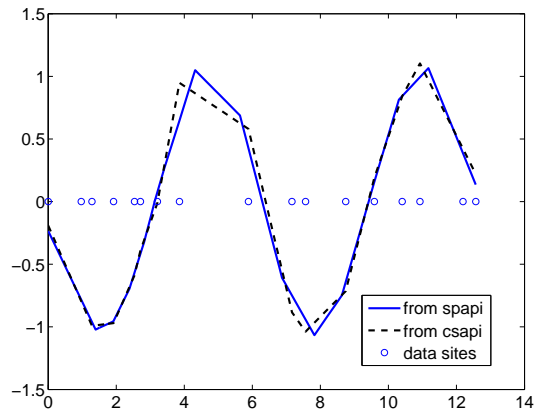
Second Derivative of Two Cubic Spline Interpolants of the Same Smooth Data

Since the second derivative of a cubic spline is a broken line, with vertices at the breaks of the spline, you can see clearly that csapi places breaks at the data sites, while spapi does not.

Knot Choices

It is, in fact, possible to specify explicitly just where the spline interpolant should have its breaks, using the command sp = spapi(knots,x,y); in which the sequence knots supplies, in a certain way, the breaks to be used. For example, recalling that you had chosen y to be sin(x), the command

```
ch = spapi(augknt(x,4,2),[x x],[y cos(x)]);
```

provides a cubic Hermite interpolant to the sine function, namely the piecewise cubic function, with breaks at all the x(i)'s, that matches the sine function in value *and* slope at all the x(i)'s. This makes the interpolant continuous with continuous first derivative but, in general, it has jumps across the breaks in its second derivative. Just how does this command know which part of the data value array [y cos(x)] supplies the values and which the slopes? Notice that the data site array here is given as [x x], i.e., each data site appears twice. Also notice that y(i) is associated with the first occurrence of x(i), and cos(x(i)) is associated with the second occurrence of x(i). The data value associated with the first appearance of a

data site is taken to be a function value; the data value associated with the second appearance is taken to be a slope. If there were a third appearance of that data site, the corresponding data value would be taken as the second derivative value to be matched at that site. See "Constructing and Working with B-form Splines" on page 10-22 for a discussion of the command `augknt` used here to generate the appropriate "knot sequence".

Smoothing

What if the data are noisy? For example, suppose that the given values are

```
noisy = y + .3*(rand(size(x))-.5);
```

Then you might prefer to approximate instead. For example, you might try the cubic smoothing spline, obtained by the command

```
scs = csaps(x,noisy);
```

and plotted by

```
fnplt(scs,2), hold on, plot(x,noisy,'o'), set(gca,'Fontsize',16)
legend('smoothing spline','noisy data'), hold off
```

This produces a figure like this:
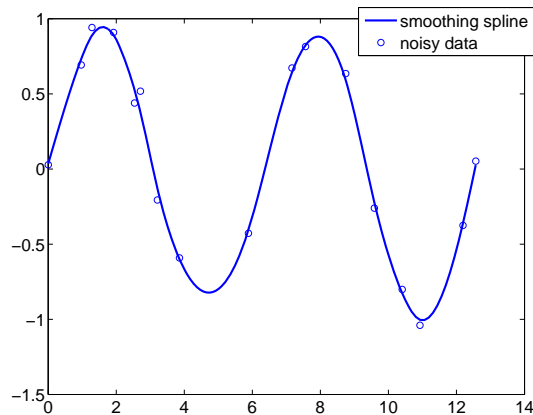
Cubic Smoothing Spline of Noisy Data

If you don't like the level of smoothing done by `csaps(x,y)`, you can change it by specifying the smoothing parameter, `p`, as an optional third argument. Choose this number anywhere between 0 and 1. As `p` changes from 0 to 1, the smoothing spline changes, correspondingly, from one extreme, the least squares straight-line approximation to the data, to the other extreme, the "natural" cubic spline interpolant to the data. Since `csaps` returns the smoothing parameter actually used as an optional second output, you could now experiment, as follows:

```
[scs,p] = csaps(x,noisy); fnplt(scs,2), hold on
fnplt(csaps(x,noisy,p/2),2,'k--'), set(gca,'Fontsize',16)
fnplt(csaps(x,noisy,(1+p)/2),2,'r:'), plot(x,noisy,'o')
legend('smoothing spline','more smoothed','less smoothed',...
'noisy data'), hold off
```

This produces the following picture.
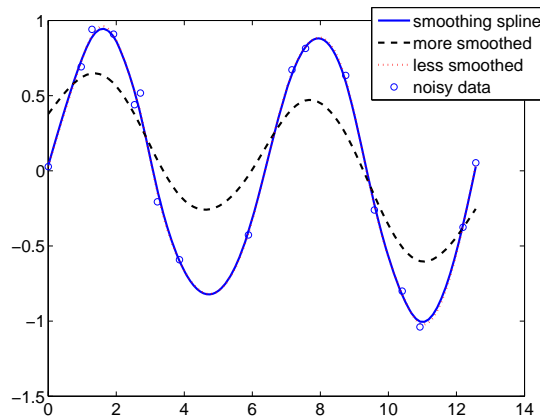
Noisy Data More or Less Smoothed

At times, you might prefer simply to get the smoothest cubic spline `sp` that is within a specified tolerance `tol` of the given data in the sense that `norm(noisy - fnval(sp,x))^2 <= tol`. You create this spline with the command `sp = spaps(x,noisy,tol)` for your defined tolerance `tol`.

Least Squares

If you prefer a least squares approximant, you can obtain it by the statement
`sp = spap2(knots,k,x,y);` in which both the knot sequence `knots` and the order `k` of the spline must be provided.

The popular choice for the order is 4, and that gives you a cubic spline. If you have no clear idea of how to choose the knots, simply specify the number of polynomial pieces you want used. For example,

```
sp = spap2(3,4,x,y);
```

gives a cubic spline consisting of three polynomial pieces. If the resulting error is uneven, you might try for a better knot distribution by using `newknt` as follows:

```
sp = spap2(newknt(sp),4,x,y);
```

Vector-Valued Functions

The toolbox supports *vector-valued* splines. For example, if you want a spline *curve* through given planar points $(x(i), y(i))$, $i = 1, \ldots, n$, then the following code defines some data and then creates and plots such a spline curve, using chord-length parametrization and cubic spline interpolation with the not-a-knot end condition.

```
x=[19 43 62 88 114 120 130 129 113 76 135 182 232 298 ...
 348 386 420 456 471 485 463 444 414 348 275 192 106 ...
 30 48 83 107 110 109 92 66 45 23 22 30 40 55 55 52 34 20 16];
y=[306 272 240 215 218 237 275 310 368 424 425 427 428 ...
 397 353 302 259 200 148 105 77 47 28 17 10 12 23 41 43 ...
 77 96 133 155 164 157 148 142 162 181 187 192 202 217 245 266 303];

xy = [x;y]; df = diff(xy,1,2);
t = cumsum([0, sqrt([1 1]*(df.*df))]);
cv = csapi(t,xy);
fnplt(cv), hold on, plot(x,y,'o'), hold off
```
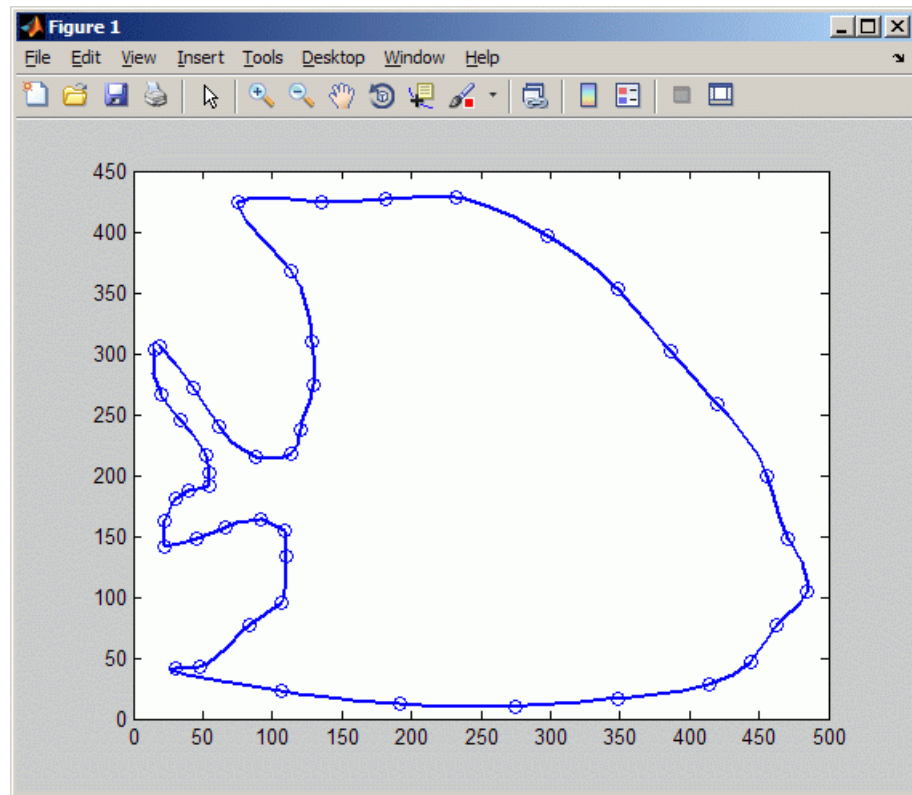
If you then wanted to know the area enclosed by this curve, you would want to evaluate the integral $\int y(t)\,\mathrm{d}x(t) = \int y(t)Dx(t)\,\mathrm{d}t$, with $(x(t),y(t))$ the point on the curve corresponding to the parameter value t. For the spline curve in cv just constructed, this can be done exactly in one (somewhat complicated) command:

```
area = diff(fnval(fnint( ...
       fncmb(fncmb(cv,[0 1]),'*',fnder(fncmb(cv,[1 0]))) ...
                       ),fnbrk(cv,'interval')));
```

To explain, y=fncmb(cv,[0 1]) picks out the second component of the curve in cv, Dx=fnder(fncmb(cv,[1 0])) provides the derivative of the first component, and yDx=fncmb(y,'*',Dx) constructs their pointwise product. Then IyDx=fnint(yDx) constructs the indefinite integral of yDx and, finally,

`diff(fnval(IyDx,fnbrk(cv,'interval')))` evaluates that indefinite integral at the endpoints of the basic interval and then takes the difference of the second from the first value, thus getting the definite integral of `yDx` over its basic interval. Depending on whether the enclosed area is to the right or to the left as the curve point travels with increasing parameter, the resulting number is either positive or negative.

Further, all the values `Y` (if any) for which the point `(X,Y)` lies on the spline curve in `cv` just constructed can be obtained by the following (somewhat complicated) command:

```
X=250; %Define a value of X
Y = fnval(fncmb(cv,[0 1]), ...
        mean(fnzeros(fncmb(fncmb(cv,[1 0]),'-',X))))
```

To explain: `x = fncmb(cv,[1 0])` picks out the first component of the curve in `cv`; `xmX = fncmb(x,'-',X)` translates that component by `X`; `t = mean(fnzeros(xmX))` provides all the parameter values for which `xmX` is zero, i.e., for which the first component of the curve equals `X`; `y = fncmb(cv,[0,1])` picks out the second component of the curve in `cv`; and, finally, `Y = fnval(y,t)` evaluates that second component at those parameter sites at which the first component of the curve in `cv` equals `X`.

As another example of the use of vector-valued functions, suppose that you have solved the equations of motion of a particle in some specified force field in the plane, obtaining, at discrete times $t_j = t(j), j = 1:n$, the position $(x(t_j), y(t_j))$ as well as the velocity $(\dot{x}(t_j), \dot{y}(t_j))$ stored in the 4-vector $z(:,j)$, as you would if, in the standard way, you had solved the equivalent first-order system numerically. Then the following statement, which uses cubic Hermite interpolation, will produce a plot of the particle path:`fnplt(spapi(augknt(t,4,2),t,reshape(z,2,2*n)))`.

Fitting Values at N-D Grid with Tensor-Product Splines

Vector-valued splines are also used in the approximation to *gridded data*, in any number of variables, using *tensor-product* splines. The same spline-construction commands are used, only the form of the input differs. For example, if x is an m-vector, y is an n-vector, and z is an array of size [m,n], then cs = csapi({x,y},z); describes a bicubic spline *f* satisfying $f(x(i),y(j))=z(i,j)$ for i=1:m, j=1:n. Such a multivariate spline can be vector-valued. For example,

```
x = 0:4; y=-2:2; s2 = 1/sqrt(2);
z(3,:,:) = [0 1 s2 0 -s2 -1 0].'*[1 1 1 1 1];
z(2,:,:) = [1 0 s2 1 s2 0 -1].'*[0 1 0 -1 0];
z(1,:,:) = [1 0 s2 1 s2 0 -1].'*[1 0 -1 0 1];
sph = csape({x,y},z,{'clamped','periodic'});
fnplt(sph), axis equal, axis off
```

gives a perfectly acceptable sphere. Its projection onto the (x,z)-plane is plotted by

```
fnplt(fncmb(sph,[1 0 0; 0 0 1])), axis equal, axis off
```

Both plots are shown below.
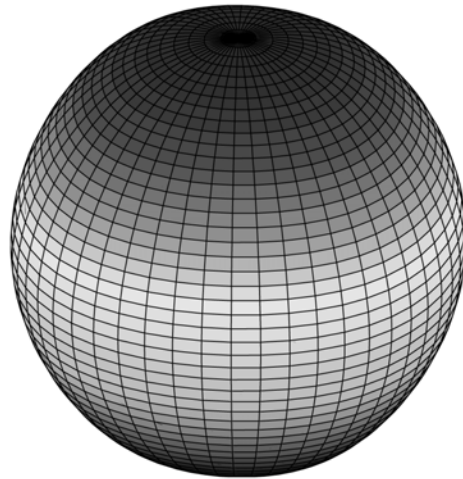
A Sphere Made by a 3-D-Valued Bivariate Tensor Product Spline
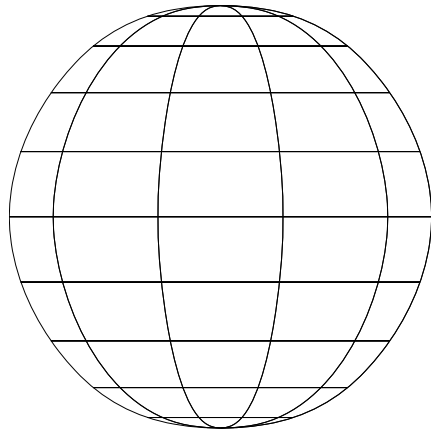
Planar Projection of Spline Sphere

Fitting Values at Scattered 2-D Sites with Thin-Plate Smoothing Splines

Tensor-product splines are good for gridded (bivariate and even multivariate) data. For work with scattered bivariate data, the toolbox provides the thin-plate smoothing spline. Suppose you have given data values y(j) at scattered data sites x(:,j), j=1:N, in the plane. To give a specific example,

```
n = 65; t = linspace(0,2*pi,n+1);
x = [cos(t);sin(t)]; x(:,end) = [0;0];
```

provides 65 sites, namely 64 points equally spaced on the unit circle, plus the center of that circle. Here are corresponding data values, namely noisy values of the very nice function $g(x) = (x(1)+1/2)^2 + (x(2)+1/2)^2$.

```
y = (x(1,:)+.5).^2 + (x(2,:)+.5).^2;
noisy = y + (rand(size(y))-.5)/3;
```

Then you can compute a reasonable approximation to these data by

```
st = tpaps(x,noisy);
```

and plot the resulting approximation along with the noisy data by

```
fnplt(st); hold on
plot3(x(1,:),x(2,:),noisy,'wo','markerfacecolor','k')
hold off
```

and so produce the following picture:
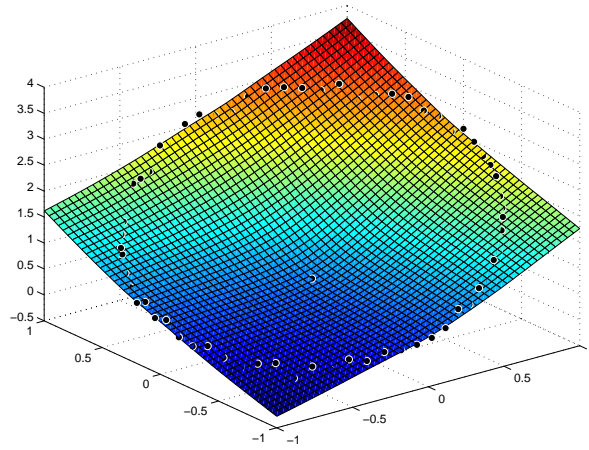
Thin-Plate Smoothing Spline Approximation to Noisy Data

Postprocessing Splines

You can use the following commands with any example spline, such as the cs, ch and sp examples constructed in the section "Cubic Spline Interpolation" on page 9-2.

First construct a spline, for example:

```
sp = spmak(1:6,0:2)
```

To display a plot of the spline:

```
fnplt(sp)
```

To get the value at a, use the syntax fnval(f,a), for example:

```
fnval(sp,4)
```

To construct the spline's second derivative:

```
DDf = fnder(fnder(sp))
```

An alternative way to construct the second derivative:

```
DDf = fnder(sp,2);
```

To obtain the spline's definite integral over an interval [a..b], in this example from 2 to 5:

```
diff(fnval(fnint(sp),[2;5]))
```

To compute the difference between two splines, use the form fncmb(sp1,'-',sp2), for example:

```
fncmb(sp,'-',DDf);
```

10

Types of Splines

Types of Splines: ppform and B-form

In this section...
"Polynomials vs. Splines" on page 10-2
"ppform" on page 10-3
"B-form" on page 10-3
"Knot Multiplicity" on page 10-3

Polynomials vs. Splines

Polynomials are the approximating functions of choice when a smooth function is to be approximated locally. For example, the truncated Taylor series

$$\sum_{i=0}^{n}(x-a)^{i}\, D^{i}f(a)\,/\,i!$$

provides a satisfactory approximation for *f(x)* if *f* is sufficiently smooth and *x* is sufficiently close to *a*. But if a function is to be approximated on a larger interval, the degree, *n*, of the approximating polynomial may have to be chosen unacceptably large. The alternative is to subdivide the interval *[a..b]* of approximation into sufficiently small intervals $[\xi_j..\xi_{j+1}]$, with $a = \xi_1 < \cdots < \xi_{l+1} = b$, so that, on each such interval, a polynomial p_j of relatively low degree can provide a good approximation to *f*. This can even be done in such a way that the polynomial pieces blend smoothly, i.e., so that the resulting patched or composite function *s(x)* that equals $p_j(x)$ for $x\epsilon[\xi_j\ \xi_{j+1}]$, all *j*, has several continuous derivatives. Any such smooth piecewise polynomial function is called a *spline*. I.J. Schoenberg coined this term because a twice continuously differentiable cubic spline with sufficiently small first derivative approximates the shape of a draftsman's spline.

There are two commonly used ways to represent a polynomial spline, the ppform and the B-form. In this toolbox, a spline in ppform is often referred to as a *piecewise polynomial*, while a piecewise polynomial in B-form is often referred to as a spline. This reflects the fact that piecewise polynomials and (polynomial) splines are just two different views of the same thing.

ppform

The *ppform* of a polynomial spline of *order k* provides a description in terms of its *breaks* $\xi_1..\xi_{l+1}$ and the *local polynomial coefficients* c_{ji} of its *l* pieces.

$$p_j(x) = \sum_{i=1}^{k}\left(x - \xi_j\right)^{k-i} c_{ji}, \quad j = 1:l$$

For example, a cubic spline is of order 4, corresponding to the fact that it requires four coefficients to specify a cubic polynomial. The ppform is convenient for the evaluation and other *uses* of a spline.

B-form

The *B-form* has become the standard way to represent a spline during its *construction*, because the B-form makes it easy to build in smoothness requirements across breaks and leads to banded linear systems. The B-form describes a spline as a weighted sum

$$\sum_{j=1}^{n} B_{j,k} a_j$$

of B-splines of the required order *k*, with their number, *n*, at least as big as *k*–1 plus the number of polynomial pieces that make up the spline. Here, $B_{j,k} = B\left(\cdot\,|\,t_j, ...,t_{j+k}\right)$ is the *j*th B-spline of *order k* for the *knot sequence* $t_1 \le t_2 \le \cdots \le t_{n+k}$. In particular, $B_{j,k}$ is piecewise-polynomial of degree $< k$, with breaks $t_j, ...,t_{j+k}$, is nonnegative, is zero outside the interval $[t_j\, ..t_{j+k}]$, and is so normalized that

$$\sum_{j=1}^{n} B_{j,k}(x) = 1 \quad on \quad \left[t_k..t_{n+1}\right]$$

Knot Multiplicity

The multiplicity of the knots governs the smoothness, in the following way: If the number τ occurs exactly *r* times in the sequence $t_j,...t_{j+k}$, then $B_{j,k}$ and its first *k-r-1* derivatives are continuous across the break τ, while the *(k-r)*th

derivative has a jump at τ. You can experiment with all these properties of the B-spline in a very visual and interactive way using the command `bspligui`.

B-Splines and Smoothing Splines

B-Spline Properties

Because $B_{j,k}$ is nonzero only on the interval $(t_j..t_{j+k})$, the linear system for the B-spline coefficients of the spline to be determined, by interpolation or least squares approximation, or even as the approximate solution of some differential equation, is *banded*, making the solving of that linear system particularly easy. For example, to construct a spline s of order k with knot sequence $t_1 \le t_2 \le \cdots \le t_{n+k}$ so that $s(x_i)=y_i$ for i=1, ..., n, use the linear system

$$\sum_{j=1}^{n} B_{j,k}(x_i) a_j = y_i \quad i = 1:n$$

for the unknown B-spline coefficients a_j in which each equation has at most k nonzero entries.

Also, many theoretical facts concerning splines are most easily stated and/or proved in terms of B-splines. For example, it is possible to match arbitrary data at sites $x_1 < \cdots < x_n$ uniquely by a spline of order k with knot sequence $(t_1, ..., t_{n+k})$ if and only if $B_{j,k}(x_j) \ne 0$ for all j (Schoenberg-Whitney Conditions). Computations with B-splines are facilitated by stable *recurrence relations*

$$B_{j,k}(x) = \frac{x - t_j}{t_{j+k-1} - t_j} B_{j,k-1}(x) + \frac{t_{j+k} - x}{t_{j+k} - t_{j+1}} B_{j+1,k-1}(x)$$

which are also of help in the conversion from B-form to ppform. The dual functional

$$a_j(s) := \sum_{i<k} (-D)^{k-i-1} \Psi_j(\tau) D^i s(\tau)$$

provides a useful expression for the jth B-spline coefficient of the spline s in terms of its value and derivatives at an arbitrary site τ between t_j and t_{j+k}, and with $\psi_j(t):=(t_{j+1}-t)\cdots(t_{j+k-1}-t)/(k-1)!$ It can be used to show that $a_j(s)$ is closely related to s on the interval $[t_j..t_{j+k}]$, and seems the most efficient means for converting from ppform to B-form.

Variational Approach and Smoothing Splines

The above *constructive* approach is not the only avenue to splines. In the *variational* approach, a spline is obtained as a *best interpolant*, e.g., as the function with smallest mth derivative among all those matching prescribed function values at certain sites. As it turns out, among the many such splines available, only those that are piecewise-polynomials or, perhaps, piecewise-exponentials have found much use. Of particular practical interest is the *smoothing spline* $s = s_p$ which, for given data (x_i, y_i) with $x\epsilon[a..b]$, all i, and given corresponding positive weights w_i, and for given *smoothing parameter p*, minimizes

$$p\sum_i w_i \left| y_i - f(x_i) \right|^2 + (1-p)\int_a^b \left| D^m f(t) \right|^2 dt$$

over all functions f with m derivatives. It turns out that the smoothing spline s is a spline of order $2m$ with a break at every data site. The smoothing parameter, p, is chosen artfully to strike the right balance between wanting the *error measure*

$$E(s) = \sum_i w_i \left| y_i - s(x_i) \right|^2$$

small and wanting the *roughness measure*

$$F\left(D^m s\right) = \int_a^b \left| D^m s(t) \right|^2 dt$$

small. The hope is that s contains as much of the information, and as little of the supposed noise, in the data as possible. One approach to this (used in spaps) is to make $F(D^m f)$ as small as possible subject to the condition that $E(f)$ be no bigger than a prescribed tolerance. For computational reasons, spaps

uses the (equivalent) smoothing parameter $\rho=p/(1-p)$, i.e., minimizes $\rho E(f) + F(D^m f)$. Also, it is useful at times to use the more flexible roughness measure

$$F\left(D^m s\right) = \int_a^b \lambda\left(t\right)\left|D^m s(t)\right|^2 dt$$

with λ a suitable positive weight function.

Multivariate and Rational Splines

In this section...
"Multivariate Splines" on page 10-8
"Rational Splines" on page 10-9

Multivariate Splines

Multivariate splines can be obtained from univariate splines by the tensor product construct. For example, a trivariate spline in B-form is given by

$$f(x,y,z) = \sum_{u=1}^{U} \sum_{v=1}^{V} \sum_{w=1}^{W} B_{u,k}(x) B_{v,l}(y) B_{w,m}(z) a_{u,v,w}$$

with $B_{u,k}, B_{v,l}, B_{w,m}$ univariate B-splines. Correspondingly, this spline is of order k in x, of order l in y, and of order m in z. Similarly, the ppform of a tensor-product spline is specified by break sequences in each of the variables and, for each hyper-rectangle thereby specified, a coefficient array. Further, as in the univariate case, the coefficients may be vectors, typically 2-vectors or 3-vectors, making it possible to represent, e.g., certain surfaces in $\Re^3$.

A very different bivariate spline is the *thin-plate spline*. This is a function of the form

$$f(x) = \sum_{j=1}^{n-3} \Psi(x - c_j) a_j + x(1) a_{n-2} + x(2) a_{n-1} + a_n$$

with $\psi(x) = |x|^2 \log|x|^2$ the thin-plate spline basis function, and $|x|$ denoting the Euclidean length of the vector x. Here, for convenience, denote the independent variable by x, but x is now a *vector* whose two components, $x(1)$ and $x(2)$, play the role of the two independent variables earlier denoted x and y. Correspondingly, the sites c_j are points in $\Re^2$.

Thin-plate splines arise as bivariate *smoothing splines*, meaning a thin-plate spline minimizes

$$p \sum_{i=1}^{n-3} \left| y_i - fc_i{}^2 \right| + (1-p) \int \left(\left| D_1 D_1 f \right|^2 + 2 \left| D_1 D_2 f \right|^2 + \left| D_2 D_2 f \right|^2 \right)$$

over all sufficiently smooth functions f. Here, the y_i are data values given at the data sites c_i, p is the smoothing parameter, and $D_j f$ denotes the partial derivative of f with respect to $x(j)$. The integral is taken over the entire $\Re^2$. The upper summation limit, n–3, reflects the fact that 3 degrees of freedom of the thin-plate spline are associated with its polynomial part.

Thin-plate splines are functions in stform, meaning that, up to certain polynomial terms, they are a weighted sum of arbitrary or scattered translates $\Psi(\cdot \text{-c})$ of one fixed function, Ψ. This so-called basis function for the thin-plate spline is special in that it is radially symmetric, meaning that $\Psi(x)$ only depends on the Euclidean length, $|x|$, of x. For that reason, thin-plate splines are also known as RBFs or radial basis functions. See "Constructing and Working with stform Splines" on page 10-36 for more information.

Rational Splines

A *rational spline* is any function of the form $r(x) = s(x)/w(x)$, with both s and w splines and, in particular, w a scalar-valued spline, while s often is vector-valued.

Rational splines are attractive because it is possible to describe various basic geometric shapes, like conic sections, exactly as the range of a rational spline. For example, a circle can so be described by a quadratic rational spline with just two pieces.

In this toolbox, there is the additional requirement that both s and w be of the same form and even of the same order, and with the same knot or break sequence. This makes it possible to store the rational spline r as the ordinary spline R whose value at x is the vector $[s(x);w(x)]$. Depending on whether the two splines are in B-form or ppform, such a representation is called here the rBform or the rpform of such a rational spline.

It is easy to obtain r from R. For example, if v is the value of R at x, then v(1:end-1)/v(end) is the value of r at x. There are corresponding ways to express derivatives of r in terms of derivatives of R.

The ppform

Introduction to ppform

A univariate *piecewise polynomial f* is specified by its *break sequence* breaks and the *coefficient array* coefs of the local power form (see equation in "Definition of ppform" on page 10-10) of its polynomial pieces; see "Multivariate Tensor Product Splines" on page 10-27 for a discussion of multivariate piecewise-polynomials. The coefficients may be (column-)vectors, matrices, even ND-arrays. For simplicity, the present discussion deals only with the case when the coefficients are scalars.

The break sequence is assumed to be strictly increasing,

```
breaks(1)
< breaks(2) < ... < breaks(l+1)
```

with l the number of polynomial pieces that make up *f*.

While these polynomials may be of varying degrees, they are all recorded as polynomials of the same *order* k, i.e., the coefficient array coefs is of size [l,k], with coefs(j,:) containing the k coefficients in the local power form for the jth polynomial piece, from the highest to the lowest power; see equation in "Definition of ppform" on page 10-10.

Definition of ppform

The items breaks, coefs, l, and k, make up the *ppform* of *f*, along with the dimension d of its coefficients; usually d equals 1. The *basic interval* of this form is the interval [breaks(1) .. breaks(l+1)]. It is the default interval over which a function in ppform is plotted by the plot command fnplt.

In these terms, the precise description of the piecewise-polynomial *f* is

$$f(t) = \text{polyval}(\text{coefs}(j,:), t - \text{breaks}(j))$$ **(10-1)**

for breaks(*j*)≤*t*<breaks(*j*+1).

Here, polyval(a,x) is the MATLAB function; it returns the number

$$\sum_{j=1}^{k} a(j) x^{k-j} = a(1) x^{k-1} + a(2) x^{k-2} + \ldots + a(k) x^{0}$$

This defines *f(t)* only for *t* in the half-open interval [breaks(1)..breaks(1+1)). For any other *t*, *f(t)* is defined by

$$f(t) = polyval\left(coefs(j,:), t - breaks(j)\right) \quad j = \begin{array}{l} 1, t < breaks(1) \\ l, t \geq breaks(l+1) \end{array}$$

i.e., by extending the first, respectively last, polynomial piece. In this way, a function in ppform has possible jumps, in its value and/or its derivatives, only across the interior breaks, breaks(2:1). The end breaks, breaks([1,1+1]), mainly serve to define the basic interval of the ppform.

Constructing and Working with ppform Splines

In this section...
"Constructing a ppform" on page 10-12
"Working With ppform Splines" on page 10-13
"Example ppform" on page 10-14

Constructing a ppform

A piecewise-polynomial is usually constructed by some command, through a process of interpolation or approximation, or conversion from some other form e.g., from the B-form, and is output as a variable. But it is also possible to make one up from scratch, using the statement

```
pp
= ppmak(breaks,coefs)
```

For example, if you enter `pp=ppmak(-5:-1,-22:-11)`, or, more explicitly,

```
breaks = -5:-1;
coefs = -22:-11; pp = ppmak(breaks,coefs);
```

you specify the uniform break sequence `-5:-1` and the coefficient sequence `-22:-11`. Because this break sequence has 5 entries, hence 4 break intervals, while the coefficient sequence has 12 entries, you have, in effect, specified a piecewise-polynomial of order 3 (= 12/4). The command

```
fnbrk(pp)
```

prints out all the constituent parts of this piecewise-polynomial, as follows:

```
breaks(1:l+1)
    -5 -4 -3 -2 -1
coefficients(d*l,k)
     -22 -21 -20
     -19 -18 -17
     -16 -15 -14
     -13 -12 -11
  pieces number l
```

```
    4
order k
  3
dimension d of target
    1
```

Further, `fnbrk` can be used to supply each of these parts separately. But the point of Curve Fitting Toolbox spline functionality is that you usually need not concern yourself with these details. You simply use `pp` as an argument to commands that evaluate, differentiate, integrate, convert, or plot the piecewise-polynomial whose description is contained in `pp`.

Working With ppform Splines

Here are some functions for operations you can perform on a piecewise-polynomial.

`v = fnval(pp,x)`	Evaluates
`dpp = fnder(pp)`	Differentiates
`dirpp = fndir(pp,dir)`	Differentiates in the direction `dir`
`ipp = fnint(pp)`	Integrates
`fnmin(pp,[a,b])`	Finds the minimum value in given interval
`fnzeros(pp,[a,b])`	Finds the zeros in the given interval
`pj = fnbrk(pp,j)`	Pulls out the `j`th polynomial piece
`pc = fnbrk(pp,[a b])`	Restricts/extends to the interval [a..b]
`po = fnxtr(pp,order)`	Extends outside its basic interval by polynomial of specified order
`fnplt(pp,[a,b])`	Plots on given interval
`sp = fn2fm(pp,'B-')`	Converts to B-form
`pr = fnrfn(pp,morebreaks)`	Inserts additional breaks

Inserting additional breaks comes in handy when you want to add two piecewise-polynomials with different breaks, as is done in the command `fncmb`.

Example ppform

Execute the following commands to create and plot the particular piecewise-polynomial (ppform) described in the "Constructing a ppform" on page 10-12 section.

1 Create the piecewise-polynomial with break sequence `-5:-1` and coefficient sequence `-22:-11`:

```
pp=ppmak(-5:-1,-22:-11)
```

2 Create the basic plot:

```
x = linspace(-5.5,-.5,101);
plot(x, fnval(pp,x),'x')
```

3 Add the break lines to the plot:

```
breaks=fnbrk(pp,'b'); yy=axis; hold on
for j=1:fnbrk(pp,'l')+1
   plot(breaks([j j]),yy(3:4))
end
```

4 Superimpose the plot of the polynomial that supplies the third polynomial piece:

```
plot(x,fnval(fnbrk(pp,3),x),'linew',1.3)
set(gca,'ylim',[-60 -10]), hold off
```
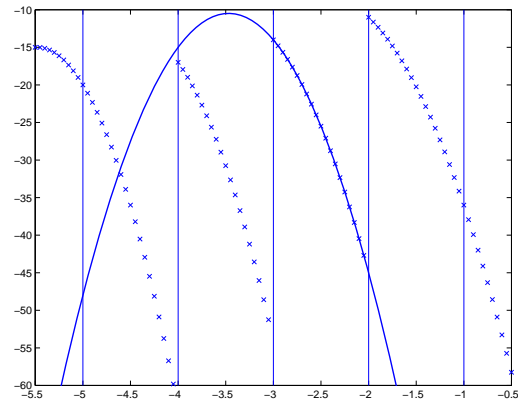
A Piecewise-Polynomial Function, Its Breaks, and the Polynomial Giving Its Third Piece

The figure above is the final picture. It shows the piecewise-polynomial as a sequence of points and, solidly on top of it, the polynomial from which its third polynomial piece is taken. It is quite noticeable that the value of a piecewise-polynomial at a break is its limit from the *right*, and that the value of the piecewise-polynomial outside its basic interval is obtained by extending its leftmost, respectively its rightmost, polynomial piece.

While the ppform of a piecewise-polynomial is efficient for evaluation, the *construction* of a piecewise-polynomial from some data is usually more efficiently handled by determining first its *B-form*, i.e., its representation as a linear combination of B-splines.

The B-form

Introduction to B-form

A univariate spline f is specified by its nondecreasing knot sequence t and by its B-spline coefficient sequence a. See "Multivariate Tensor Product Splines" on page 10-27 for a discussion of multivariate splines. The coefficients may be (column-)vectors, matrices, even ND-arrays. When the coefficients are 2-vectors or 3-vectors, f is a curve in R^2 or R^3 and the coefficients are called the *control points* for the curve.

Roughly speaking, such a spline is a piecewise-polynomial of a certain order and with breaks t(*i*). But knots are different from breaks in that they may be repeated, i.e., t need not be *strictly* increasing. The resulting knot *multiplicities* govern the smoothness of the spline across the knots, as detailed below.

With `[d,n] = size(a)`, and `n+k = length(t)`, the spline is of *order* k. This means that its polynomial pieces have degree < k. For example, a *cubic* spline is a spline of *order 4* because it takes four coefficients to specify a cubic polynomial.

Definition of B-form

These four items, t, a, n, and k, make up the B-form of the spline f.

This means, explicitly, that

$$f = \sum_{i=1}^{n} B_{i,k} a(:,i)$$

with $B_{i,k}=B(\cdot|t(i:i+k))$ the ith B-spline of order k for the given knot sequence t, i.e., the B-spline with knots $t(i),...,t(i+k)$. The basic interval of this B-form is the interval $[t(1)..t(n+k)]$. It is the default interval over which a spline in B-form is plotted by the command fnplt. Note that a spline in B-form is zero outside its basic interval while, after conversion to ppform via fn2fm, this is usually not the case because, outside its basic interval, a piecewise-polynomial is defined by extension of its first or last polynomial piece. In particular, a function in B-form may have jumps in value and/or one of its derivative not only across its interior knots, i.e., across $t(i)$ with $t(1)<t(i)<t(n+k)$, but also across its end knots, $t(1)$ and $t(n+k)$.

B-form and B-Splines

The building blocks for the B-form of a spline are the B-splines. A B-Spline of Order 4, and the Four Cubic Polynomials from Which It Is Made on page 10-18 shows a picture of such a B-spline, the one with the knot sequence [0 1.5 2.3 4 5], hence of order 4, together with the polynomials whose pieces make up the B-spline. The information for that picture could be generated by the command

```
bspline([0 1.5 2.3 4 5])
```
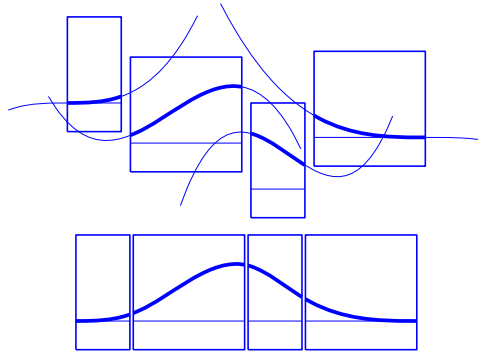
A B-Spline of Order 4, and the Four Cubic Polynomials from Which It Is Made

To summarize: The B-spline with knots $t(i) \le \cdots \le t(i+k)$ is positive on the interval $(t(i)..t(i+k))$ and is zero outside that interval. It is piecewise-polynomial of order k with breaks at the sites $t(i),...,t(i+k)$. These knots may coincide, and the precise *multiplicity* governs the smoothness with which the two polynomial pieces join there.

Definition of B-Splines

The shorthand

$$f \in S_{k,t}$$

is one of several ways to indicate that *f* is a spline of order k with knot sequence t, i.e., *a linear combination of the B-splines* of order k for the knot sequence t.

A word of caution: The term *B-spline* has been expropriated by the Computer-Aided Geometric Design (CAGD) community to mean what is called here a *spline in B-form*, with the unhappy result that, in any discussion between mathematicians/approximation theorists and people in CAGD, one now always has to check in what sense the term is being used.

B-Spline Knot Multiplicity

The rule is

knot multiplicity + condition multiplicity = order
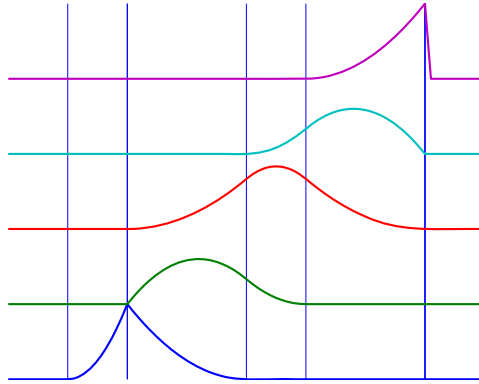
All Third-Order B-Splines for a Certain Knot Sequence with Various Knot Multiplicities

For example, for a B-spline of order 3, a simple knot would mean two smoothness conditions, i.e., continuity of function and first derivative, while a double knot would only leave one smoothness condition, i.e., just continuity, and a triple knot would leave no smoothness condition, i.e., even the function would be discontinuous.

All Third-Order B-Splines for a Certain Knot Sequence with Various Knot Multiplicities on page 10-19 shows a picture of all the third-order B-splines for a certain mystery knot sequence t. The breaks are indicated by vertical lines. For each break, try to determine its multiplicity in the knot sequence (it is 1,2,1,1,3), as well as its multiplicity as a knot in each of the B-splines. For example, the second break has multiplicity 2 but appears only with multiplicity 1 in the third B-spline and not at all, i.e., with multiplicity 0, in the last two B-splines. Note that only one of the B-splines shown has all its knots simple. It is the only one having three different nontrivial polynomial pieces. Note also that you can tell the knot-sequence multiplicity of a knot by the number of B-splines whose nonzero part begins or ends there. The

picture is generated by the following MATLAB statements, which use the command `spcol` from this toolbox to generate the function values of all these B-splines at a fine net `x`.

```
t=[0,1,1,3,4,6,6,6]; x=linspace(-1,7,81);
c=spcol(t,3,x);[l,m]=size(c);
c=c+ones(l,1)*[0:m-1];
axis([-1 7 0 m]); hold on
for tt=t, plot([tt tt],[0 m],'-'), end
plot(x,c,'linew',2), hold off, axis off
```

Further illustrated examples are provided by the example "Construct and Work with the B-form". You can also use the GUI `bspligui` to study the dependence of a B-spline on its knots experimentally.

Choice of Knots for B-form

The rule "knot multiplicity + condition multiplicity = order" has the following consequence for the process of choosing a knot sequence for the B-form of a spline approximant. Suppose the spline s is to be of order k, with basic interval $[a..b]$, and with interior breaks $\xi_2 < \cdots < \xi_l$. Suppose, further, that, at ξ_i, the spline is to satisfy μ_i smoothness conditions, i.e.,

$$jump_{\xi_i} D^j s := D^j s(\xi_{i+}) - D^j s(\xi_{i-}) = 0, \quad 0 \le j < \mu_i, \quad i = 2,...,l$$

Then, the appropriate knot sequence t should contain the break ξ_i exactly $k - \mu_i$ times, $i=2,...,l$. In addition, it should contain the two endpoints, a and b, of the basic interval exactly k times. This last requirement can be relaxed, but has become standard. With this choice, there is exactly one way to write each spline s with the properties described as a weighted sum of the B-splines of order k with knots a segment of the knot sequence t. This is the reason for the B in *B-spline*: B-splines are, in Schoenberg's terminology, *basic* splines.

For example, if you want to generate the B-form of a cubic spline on the interval [1 .. 3], with interior breaks 1.5, 1.8, 2.6, and with two continuous derivatives, then the following would be the appropriate knot sequence:

```
t = [1, 1, 1, 1, 1.5, 1.8, 2.6, 3, 3, 3, 3];
```

This is supplied by augknt([1, 1.5, 1.8, 2.6, 3], 4). If you wanted, instead, to allow for a corner at 1.8, i.e., a possible jump in the first derivative there, you would triple the knot 1.8, i.e., use

```
t = [1, 1, 1, 1, 1.5, 1.8, 1.8, 1.8, 2.6, 3, 3, 3, 3];
```

and this is provided by the statement

```
t = augknt([1, 1.5, 1.8, 2.6, 3], 4, [1, 3, 1] );
```

Constructing and Working with B-form Splines

In this section...
"Construction of B-form" on page 10-22
"Working With B-form Splines" on page 10-23
"Example: B-form Spline Approximation to a Circle" on page 10-24

Construction of B-form

Usually, a spline is constructed from some information, like function values and/or derivative values, or as the approximate solution of some ordinary differential equation. But it is also possible to make up a spline from scratch, by providing its knot sequence and its coefficient sequence to the command `spmak`.

For example, if you enter

```
sp = spmak(1:10,3:8);
```

you supply the uniform knot sequence `1:10` and the coefficient sequence `3:8`. Because there are 10 knots and 6 coefficients, the order must be $4(= 10 - 6)$, i.e., you get a cubic spline. The command

```
fnbrk(sp)
```

prints out the constituent parts of the B-form of this cubic spline, as follows:

```
knots(1:n+k)
   1 2 3 4 5 6 7 8 9 10
coefficients(d,n)
   3 4 5 6 7 8
number n of coefficients
   6
order k
   4
dimension d of target
   1
```

Further, `fnbrk` can be used to supply each of these parts separately.

But the point of the Curve Fitting Toolbox spline functionality is that there shouldn't be any need for you to look up these details. You simply use `sp` as an argument to commands that evaluate, differentiate, integrate, convert, or plot the spline whose description is contained in `sp`.

Working With B-form Splines

The following commands are available for spline work. There is `spmak` and `fnbrk` to make up a spline and take it apart again. Use `fn2fm` to convert from B-form to ppform. You can evaluate, differentiate, integrate, minimize, find zeros of, plot, refine, or selectively extrapolate a spline with the aid of `fnval`, `fnder`, `fndir`, `fnint`, `fnmin`, `fnzeros`, `fnplt`, `fnrfn`, and `fnxtr`.

There are five commands for generating knot sequences:

- `augknt` for providing boundary knots and also controlling the multiplicity of interior knots

- `brk2knt` for supplying a knot sequence with specified multiplicities

- `aptknt` for providing a knot sequence for a spline space of given order that is suitable for interpolation at given data sites

- `optknt` for providing an *optimal* knot sequence for interpolation at given sites

- `newknt` for a knot sequence perhaps more suitable for the function to be approximated

In addition, there is:

- `aveknt` to supply certain knot averages (the Greville sites) as recommended sites for interpolation

- `chbpnt` to supply such sites

- `knt2brk` and `knt2mlt` for extracting the breaks and/or their multiplicities from a given knot sequence

To display a spline *curve* with given two-dimensional coefficient sequence and a uniform knot sequence, use `spcrv`.

You can also write your own spline construction commands, in which case you will need to know the following. The construction of a spline satisfying some interpolation or approximation conditions usually requires a *collocation matrix*, i.e., the matrix that, in each row, contains the sequence of numbers $D^r B_{j,k}(\tau)$, i.e., the rth derivative at τ of the jth B-spline, for all j, for some r and some site τ. Such a matrix is provided by `spcol`. An optional argument allows for this matrix to be supplied by `spcol` in a space-saving spline-almost-block-diagonal-form or as a MATLAB sparse matrix. It can be fed to `slvblk`, a command for solving linear systems with an almost-block-diagonal coefficient matrix. If you are interested in seeing how `spcol` and `slvblk` are used in this toolbox, have a look at the commands `spapi`, `spap2`, and `spaps`.

In addition, there are routines for constructing *cubic* splines. `csapi` and `csape` provide the cubic spline interpolant at knots to given data, using the not-a-knot and various other end conditions, respectively. A parametric cubic spline curve through given points is provided by `cscvn`. The cubic *smoothing* spline is constructed in `csaps`.

Example: B-form Spline Approximation to a Circle

As another simple example,

```
points = .95*[0 -1 0 1;1 0 -1 0];
sp = spmak(-4:8,[points points]);
```

provides a planar, quartic, spline curve whose middle part is a pretty good approximation to a circle, as the plot on the next page shows. It is generated by a subsequent

```
plot(points(1,:),points(2,:),'x'), hold on
fnplt(sp,[0,4]), axis equal square, hold off
```

Insertion of additional control points $(\pm 0.95, \pm 0.95)/\sqrt{1.9}$ would make a visually perfect circle.

Here are more details. The spline curve generated has the form $\Sigma^8_{j=1} B_{j,5} a(:, j)$, with `-4:8` the uniform knot sequence, and with its control points $a(:,j)$ the sequence $(0,\alpha),(-\alpha,0),(0,-\alpha),(\alpha,0),(0,\alpha),(-\alpha,0),(0,-\alpha),(\alpha,0)$ with $\alpha = 0.95$. Only the curve part between the parameter values 0 and 4 is actually plotted.

To get a feeling for how close to circular this part of the curve actually is, compute its unsigned curvature. The curvature $\kappa(t)$ at the curve point $\gamma(t) = (x(t), y(t))$ of a space curve γ can be computed from the formula

$$\kappa = \frac{\left| x'\, y'' - y'\, x'' \right|}{(x'^2 + y'^2)^{3/2}}$$

in which x', x″, y', and y″ are the first and second derivatives of the curve with respect to the parameter used (*t*). Treat the planar curve as a space curve in the (*x*,*y*)-plane, hence obtain the maximum and minimum of its curvature at 21 points as follows:

```
t = linspace(0,4,21);zt = zeros(size(t));
dsp = fnder(sp); dspt = fnval(dsp,t); ddspt = fnval(fnder(dsp),t);
kappa = abs(dspt(1,:).*ddspt(2,:)-dspt(2,:).*ddspt(1,:))./...
   (sum(dspt.^2)).^(3/2);
[min(kappa),max(kappa)]

ans =
    1.6747    1.8611
```

So, while the curvature is not quite constant, it is close to 1/radius of the circle, as you see from the next calculation:

```
1/norm(fnval(sp,0))

ans =
    1.7864
```
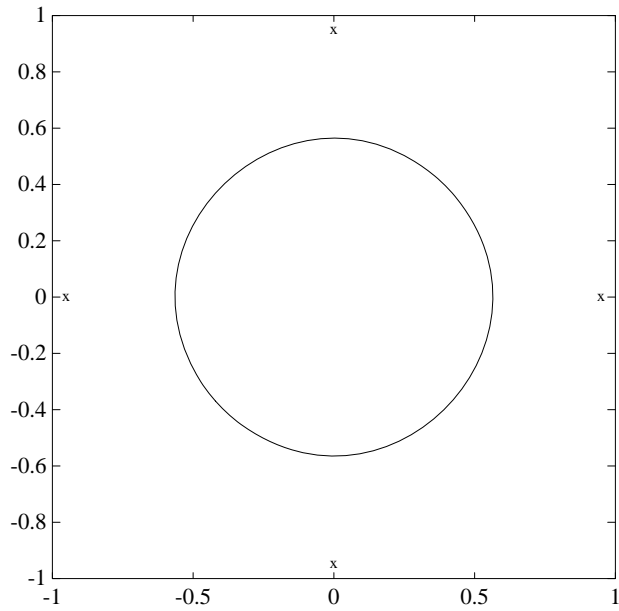
Spline Approximation to a Circle; Control Points Are Marked x

Multivariate Tensor Product Splines

In this section...
"Introduction to Multivariate Tensor Product Splines" on page 10-27
"B-form of Tensor Product Splines" on page 10-27
"Construction With Gridded Data" on page 10-28
"ppform of Tensor Product Splines" on page 10-28
"Example: The Mobius Band" on page 10-28

Introduction to Multivariate Tensor Product Splines

The toolbox provides (polynomial) spline functions in any number of variables, as tensor products of univariate splines. These multivariate splines come in both standard forms, the B-form and the ppform, and their construction and use parallels entirely that of the univariate splines discussed in previous sections, "Constructing and Working with ppform Splines" on page 10-12 and "Constructing and Working with B-form Splines" on page 10-22. The same commands are used for their construction and use.

For simplicity, the following discussion deals just with bivariate splines.

B-form of Tensor Product Splines

The tensor-product idea is very simple. If f is a function of x, and g is a function of y, then their tensor-product $p\,(x,y) := f\,(x)g(y)$ is a function of x and y, i.e., a bivariate function. More generally, with $s=(s_1,...,s_{m+h})$ and $t=(t_1,...,t_{n+k})$ knot sequences and $a_{ji}:i=1,...,m;j=1,...n)$ a corresponding coefficient array, you obtain a bivariate spline as

$$f(x,y) = \sum_{i=1}^{m} \sum_{j=1}^{n} B\big(x \,|\, s_i,...,s_{i+h}\big) B\big(y \,|\, t_j,...,t_{j+k}\big) a_{ij}$$

The B-form of this spline comprises the cell array $\{s,t\}$ of its knot sequences, the coefficient array a, the numbers vector $[m,n]$, and the orders vector $[h,k]$. The command

```
sp = spmak({s,t},a);
```

constructs this form. Further, `fnplt`, `fnval`, `fnder`, `fndir`, `fnrfn`, and `fn2fm` can be used to plot, evaluate, differentiate and integrate, refine, and convert this form.

Construction With Gridded Data

You are most likely to construct such a form by looking for an interpolant or approximant to gridded data. For example, if you know the values $z(i,j)=g(x(i),y(j))$, $i=1{:}m$, $j=1{:}n$, of some function g at all the points in a rectangular grid, then, assuming that the strictly increasing sequence `x` satisfies the Schoenberg-Whitney conditions with respect to the above knot sequence s, and that the strictly increasing sequence `y` satisfies the Schoenberg-Whitney conditions with respect to the above knot sequence t, the command

```
sp=spapi({s,t},[h,k],{x,y},z);
```

constructs the unique bivariate spline of the above form that matches the given values. The command `fnplt(sp)` gives you a quick plot of this interpolant. The command `pp = fn2fm(sp,'pp')` gives you the ppform of this spline, which is probably what you want when you want to evaluate the spline at a fine grid (`(xx(i),yy(j))` for `i=1:M`, `j=1:N`), by the command:

```
values = fnval(pp,{xx,yy});
```

ppform of Tensor Product Splines

The ppform of such a bivariate spline comprises, analogously, a cell array of break sequences, a multidimensional coefficient array, a vector of number pieces, and a vector of polynomial orders. Fortunately, the toolbox is set up in such a way that there is usually no reason for you to concern yourself with these details of either form. You use interpolation, approximation, or smoothing to construct splines, and then use the `fn...` commands to make use of them.

Example: The Mobius Band

Here is an example of a surface constructed as a 3-D-valued bivariate spline. The surface is the famous Möbius band, obtainable by taking a longish strip

of paper and gluing its narrow ends together, but with a twist. The figure is obtained by the following commands:

```
x = 0:1; y = 0:4; h = 1/4; o2 = 1/sqrt(2); s = 2; ss = 4;
v(3,:,:) = h*[0, -1, -o2, 0, o2, 1, 0;0, 1, o2, 0, -o2, -1, 0];
v(2,:,:) = [ss, 0, s-h*o2, 0, -s-h*o2, 0, ss;...
            ss, 0, s+h*o2, 0,-s+h*o2, 0, ss];
v(1,:,:) = s*[0, 1, 0, -1+h, 0, 1, 0; 0, 1, 0, -1-h, 0, 1, 0];
cs = csape({x,y},v,{'variational','clamped'});
fnplt(cs), axis([-2 2 -2.5 2.5 -.5 .5]), shading interp
axis off, hold on
values = squeeze(fnval(cs,{1,linspace(y(1),y(end),51)}));
plot3(values(1,:), values(2,:), values(3,:),'k','linew',2)
view(-149,28), hold off
```
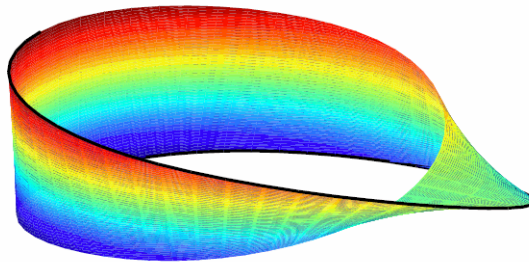

A Möbius Band Made by Vector-Valued Bivariate Spline Interpolation

NURBS and Other Rational Splines

Introduction to Rational Splines

A rational spline is, by definition, any function that is the ratio of two splines:

$$r(x) = s(x)/w(x)$$

This requires w to be scalar-valued, but s is often chosen to be vector-valued. Further, it is desirable that $w(x)$ be not zero for any x of interest.

Rational splines are popular because, in contrast to ordinary splines, they can be used to describe certain basic design shapes, like conic sections, exactly.

rsform: rpform, rBform

The two splines, s and w, in the rational spline $r(x)=s(x)/w(x)$ need not be related to one another. They could even be of different forms. But, in the context of this toolbox, it is convenient to restrict them to be of the same form, and even of the same order and with the same breaks or knots. For, under that assumption, you can represent such a rational spline by the (vector-valued) spline function

$$R(x) = \left[s(x); w(x) \right]$$

whose values are vectors with one more entry than the values of the rational spline r, and call this the *rsform* of the rational spline, or, more precisely, the *rpform* or *rBform*, depending on whether s and w are in ppform or in B-form. Internally, the only thing that distinguishes these rational forms from their corresponding ordinary spline forms, rpform and B-form, is their form part, i.e., the string obtained via fnbrk(r,'form'). This is enough to alert the fn... commands to act appropriately on a function in one of the rsforms.

For example, as is done in `fnval`, it is very easy to obtain $r(x)$ from $R(x)$. If `v` is the value of R at x, then `v(1:end-1)/v(end)` is the value of r at x. If, in addition, `dv` is $DR(x)$, then `(dv(1:end-1)-dv(end)*v(1:end-1))/v(end)` is $Dr(x)$. More generally, by Leibniz's formula,

$$D^j s = D^j (wr) = \sum_{i=0}^{j} \binom{j}{i} D^i w D^{j-i} r$$

Therefore,

$$D^j r = \left(D^j s - \sum_{i=1}^{j} \binom{j}{i} D^i w D^{j-i} r \right) / w$$

This shows that you can compute the derivatives of r inductively, using the derivatives of s and w (i.e., the derivatives of R) along with the derivatives of r of order less than j to compute the jth derivative of r. This inductive scheme is used in `fntlr` to provide the first so many derivatives of a rational spline. There is a corresponding formula for partial and directional derivatives for multivariate rational splines.

Constructing and Working with Rational Splines

In this section...
"Rational Spline Example: Circle" on page 10-32
"Rational Spline Example: Sphere" on page 10-33
"Functions for Working With Rational Splines" on page 10-34

Rational Spline Example: Circle

For example,

```
circle = rsmak('circle');
```

provides a rational spline whose values on its basic interval trace out the unit circle, i.e., the circle of radius 1 with center at the origin, as the command

```
fnplt(circle), axis square
```

readily shows; the resulting output is the circle in the figure A Circle and an Ellipse, Both Given by a Rational Spline on page 10-33.

It is easy to manipulate this circle to obtain related shapes. For example, the next commands stretch the circle into an ellipse, rotate the ellipse 45 degrees, and translate it by (1,1), and then plot it on top of the circle.

```
ellipse = fncmb(circle,[2 0;0 1]);
s45 = 1/sqrt(2);
rtellipse = fncmb(fncmb(ellipse, [s45 -s45;s45 s45]), [1;1] );
hold on, fnplt(rtellipse), hold off
```

As a further example, the "circle" just constructed is put together from four pieces. To highlight the first such piece, use the following commands:

```
quarter = fnbrk(fn2fm(circle,'rp'),1);
hold on, fnplt(quarter,3), hold off
```

In the first command, fn2fm is used to change forms, from one based on the B-form to one based on the ppform, and then fnbrk is used to extract the first piece, and this piece is then plotted on top of the circle in A Circle and

an Ellipse, Both Given by a Rational Spline on page 10-33, with linewidth 3 to make it stand out.
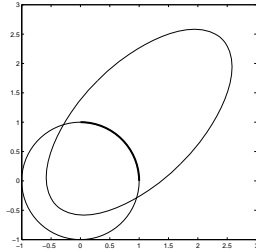
A Circle and an Ellipse, Both Given by a Rational Spline

Rational Spline Example: Sphere

As a surface example, the command rsmak('southcap') provides a 3-vector valued rational bicubic polynomial whose values on the unit square [-1 .. 1]^2 fill out a piece of the unit sphere. Adjoin to it five suitable rotates of it and you get the unit sphere exactly. For illustration, the following commands generate two-thirds of that sphere, as shown in Part of a Sphere Formed by Four Rotates of a Quartic Rational on page 10-34.

```
southcap = rsmak('southcap'); fnplt(southcap)
xpcap = fncmb(southcap,[0 0 -1;0 1 0;1 0 0]);
ypcap = fncmb(xpcap,[0 -1 0; 1 0 0; 0 0 1]);
northcap = fncmb(southcap,-1);
hold on, fnplt(xpcap), fnplt(ypcap), fnplt(northcap)
axis equal, shading interp, view(-115,10), axis off, hold off
```

Part of a Sphere Formed by Four Rotates of a Quartic Rational

Functions for Working With Rational Splines

Having chosen to represent the rational spline $r = s/w$ in this way by the ordinary spline $R=[s;w]$ makes it is easy to apply to a rational spline all the fn... commands in the Curve Fitting Toolbox spline functions, with the following exceptions. The integral of a rational spline need not be a rational spline, hence there is no way to extend fnint to rational splines. The derivative of a rational spline *is* again a rational spline but one of roughly twice the order. For that reason, fnder and fndir will not touch rational splines. Instead, there is the command fntlr for computing the value at a given x of all derivatives up to a given order of a given function. If that function is rational, the needed calculation is based on the considerations given in the preceding paragraph.

The command r = rsmak(shape) provides rational splines in rBform that describe exactly certain standard geometric shapes , like 'circle', 'arc', 'cylinder', 'sphere', 'cone', 'torus'. The command fncmb(r,trans) can be used to apply standard transformations to the resulting shape. For example, if trans is a column-vector of the right length, the shape would be translated by that vector while, if trans is a suitable matrix like a rotation, the shape would be transformed by that matrix.

The command r = rscvn(p) constructs the quadratic rBform of a tangent-continuous curve made up of circular arcs and passing through the given sequence, p, of points in the plane.

A special rational spline form, called a NURBS, has become a standard tool in CAGD. A NURBS is, by definition, any rational spline for which both s and w are in the same B-form, with each coefficient for s containing explicitly the corresponding coefficient for w as a factor:

$$s = \sum_i B_i v(i) a(:,i), \quad w = \sum_i B_i v(i)$$

The normalized coefficients $a(:,i)$ for the numerator spline are more readily used as control points than the unnormalized coefficients $v(i)a(:,i)$ used in the rBform. Nevertheless, this toolbox provides no special NURBS form, but only the more general rational spline, but in both B-form (called `rBform` internally) and in ppform (called `rpform` internally).

The rational spline `circle` used earlier is put together in `rsmak` by code like the following.

```
x = [1 1 0 -1 -1 -1  0  1 1]; y = [0 1 1  1  0 -1 -1 -1 0];
s45 = 1/sqrt(2); w =[1 s45 1 s45 1 s45 1 s45 1];
circle = rsmak(augknt(0:4,3,2), [w.*x;w.*y;w]);
```

Note the appearance of the denominator spline as the last component. Also note how the coefficients of the denominator spline appear here explicitly as factors of the corresponding coefficients of the numerator spline. The normalized coefficient sequence [x;y] is very simple; it consists of the vertices and midpoints, in proper order, of the "unit square". The resulting control polygon is tangent to the circle at the places where the four quadratic pieces that form the circle abut.

For a thorough discussion of NURBS, see [G. Farin, *NURBS*, 2nd ed., AKPeters Ltd, 1999] or [Les Piegl and Wayne Tiller, *The NURBS Book*, 2nd ed., Springer-Verlag, 1997].

Constructing and Working with stform Splines

Introduction to the stform

A multivariate function form quite different from the tensor-product construct is the scattered translates form, or stform for short. As the name suggests, it uses arbitrary or scattered translates $\psi(\cdot - c_j)$ of one fixed function ψ, in addition to some polynomial terms. Explicitly, such a form describes a function

$$f(x) = \sum_{j=1}^{n-k} \psi\left(x - c_j\right) a_j + p(x)$$

in terms of the *basis function* ψ, a sequence (c_j) of sites called *centers* and a corresponding sequence (a_j) of n coefficients, with the final k coefficients, $a_{n-k+1},...,a_n$, involved in the *polynomial part*, p.

When the basis function is radially symmetric, meaning that $\psi(x)$ depends only on the Euclidean length $|x|$ of its argument, x, then ψ is called a *radial basis function*, and, correspondingly, f is then often called an RBF.

At present, the toolbox works with just one kind of stform, namely a bivariate thin-plate spline and its first partial derivatives. For the thin-plate spline, the basis function is $\psi(x) = \varphi(|x|^2)$, with $\varphi(t) = t\log t$, i.e., a radial basis function. Its polynomial part is a linear polynomial, i.e., $p(x)=x(1)a_{n-2}+x(2)a_{n-1}+a_n$. The first partial derivative with respect to its first argument uses, correspondingly, the basis function $\psi(x)=\varphi(|x|^2)$, with $\varphi(t) = (D_1 t) \cdot (\log t+1)$ and $D_1 t = D_1 t(x) = 2x(1)$, and $p(x) = a_n$.

Construction and Properties of the stform

A function in stform can be put together from its center sequence `centers` and its coefficient sequence `coefs` by the command

```
f = stmak(centers, coefs, type);
```

with the string `type` one of `'tp00'`, `'tp10'`, `'tp01'`, to indicate, respectively, a thin-plate spline, a first partial of a thin-plate spline with respect to the first argument, and a first partial of a thin-plate spline with respect to the second argument. There is one other choice, `'tp'`; it denotes a thin-plate spline without any polynomial part and is likely to be used only during the construction of a thin-plate spline, as in `tpaps`.

A function *f* in stform depends linearly on its coefficients, meaning that

$$f(x) = \sum_{j=1}^{n} \psi_j(x) a_j$$

with ψ_j either a translate of the basis function Ψ or else some polynomial. Suppose you wanted to determine these coefficients a_j so that the function *f* matches prescribed values at prescribed sites x_i. Then you would need the collocation matrix $(\psi_j(x_i))$. You can obtain this matrix by the command `stcol(centers,x,type)`. In fact, because the stform has a_j as the *j*th *column*, `coefs(:,j)`, of its coefficient array, it is worth noting that `stcol` can also supply the *transpose* of the collocation matrix. Thus, the command

```
values = coefs*stcol(centers,x,type,'tr');
```

would provide the values at the entries of `x` of the st function specified by `centers` and `type`.

The stform is attractive because, in contrast to piecewise polynomial forms, its complexity is the same in any number of variables. It is quite simple, yet, because of the complete freedom in the choice of centers, very flexible and adaptable.

On the negative side, the most attractive choices for a radial basis function share with the thin-plate spline that the evaluation at any site involves all coefficients. For example, plotting a scalar-valued thin-plate spline via

`fnplt` involves evaluation at a 51-by-51 grid of sites, a nontrivial task when there are 1000 coefficients or more. The situation is worse when you want to determine these 1000 coefficients so as to obtain the stform of a function that matches function values at 1000 data sites, as this calls for solving a full linear system of order 1000, a task requiring $O(10^9)$ flops if done by a direct method. Just the construction of the collocation matrix for this linear system (by `stcol`) takes $O(10^6)$ flops.

The command `tpaps`, which constructs thin-plate spline interpolants and approximants, uses iterative methods when there are more than 728 data points, but convergence of such iteration may be slow.

Working with the stform

After you have constructed an approximating or interpolating thin-plate spline `st` with the aid of `tpaps` (or directly via `stmak`), you can use the following commands:

- `fnbrk` to obtain its parts or change its basic interval,

- `fnval` to evaluate it

- `fnplt` to plot it

- `fnder` to construct its two first partial derivatives, but no higher order derivatives as they become infinite at the centers.

  This is just one indication that the stform is quite different in nature from the other forms in this toolbox, hence other `fn...` commands by and large don't work with stforms. For example, it makes no sense to use `fnjmp`, and `fnmin` or `fnzeros` only work for univariate functions. It also makes no sense to use `fnint` on a function in stform because such functions cannot be integrated in closed form.

- The command `Ast = fncmb(st,A)` can be used on `st`, provided `A` is something that can be applied to the values of the function described by `st`. For example, `A` might be `'sin'`, in which case `Ast` is the stform of the function whose coefficients are the sine of the coefficients of `st`. In effect, `Ast` describes the function obtained by composing `A` with `st`. But, because of the singularities in the higher-order derivatives of a thin-plate spline, there seems little point to make `fndir` or `fntlr` applicable to such a `st`.

11

Advanced Spline Examples

Least-Squares Approximation by Natural Cubic Splines

The construction of a least-squares approximant usually requires that one have in hand a basis for the space from which the data are to be approximated. As the example of the space of "natural" cubic splines illustrates, the explicit construction of a basis is not always straightforward.

This section makes clear that an explicit basis is not actually needed; it is sufficient to have available some means of interpolating in some fashion from the space of approximants. For this, the fact that the Curve Fitting Toolbox spline functions support work with vector-valued functions is essential.

This section discusses these aspects of least-squares approximation by "natural" cubic splines.

- "Problem" on page 11-2
- "General Resolution" on page 11-2
- "Need for a Basis Map" on page 11-3
- "A Basis Map for "Natural" Cubic Splines" on page 11-3
- "The One-line Solution" on page 11-4
- "The Need for Proper Extrapolation" on page 11-4
- "The Correct One-Line Solution" on page 11-6
- "Least-Squares Approximation by Cubic Splines" on page 11-7

Problem

You want to construct the least-squares approximation to given data (x,y) from the space S of "natural" cubic splines with given breaks b(1) < ...< b(1+1).

General Resolution

If you know a basis, (f1,f2,...,fm), for the linear space S of all "natural" cubic splines with break sequence b, then you have learned to find the least-squares approximation in the form c(1)f1+ c(2)f2+ ... + c(m)fm, with the vector c the least-squares solution to the linear system A*c = y, whose coefficient matrix is given by

```
A(i,j) = fj(x(i)),   i=1:length(x),   j=1:m .
```

In other words, `c = A\y`.

Need for a Basis Map

The general solution seems to require that you know a basis. However, in order to construct the coefficient sequence `c`, you only need to know the matrix `A`. For this, it is sufficient to have at hand a *basis map*, namely a function `F` say, so that `F(c)` returns the spline given by the particular weighted sum `c(1)f1+c(2)f2+... +c(m)fm`. For, with that, you can obtain, for `j=1:m`, the `j`-th column of `A` as `fnval(F(ej),x)`, with `ej` the `j`-th column of `eye(m)`, the identity matrix of order `m`.

Better yet, the Curve Fitting Toolbox spline functions can handle *vector-valued* functions, so you should be able to construct the basis map `F` to handle vector-valued coefficients `c(i)` as well. However, by agreement, in this toolbox, a vector-valued coefficient is a *column* vector, hence the sequence `c` is necessarily a row vector of column vectors, i.e., a *matrix*. With that, `F(eye(m))` is the vector-valued spline whose `i`-th component is the basis element `fi`, `i=1:m`. Hence, assuming the vector `x` of data sites to be a row vector, `fnval(F(eye(m)),x)` is the matrix whose `(i,j)`-entry is the value of `fi` at `x(j)`, i.e., the *transpose* of the matrix `A` you are seeking. On the other hand, as just pointed out, your basis map `F` expects the coefficient sequence `c` to be a row vector, i.e., the transpose of the vector `A\y`. Hence, assuming, correspondingly, the vector `y` of data values to be a row vector, you can obtain the least-squares approximation from *S* to data (x,y) as

```
F(y/fnval(F(eye(m)),x))
```

To be sure, if you wanted to be prepared for `x` and `y` to be arbitrary vectors (of the same length), you would use instead

```
F(y(:).'/fnval(F(eye(m)),x(:).'))
```

A Basis Map for "Natural" Cubic Splines

What exactly is required of a basis map `F` for the linear space *S* of "natural" cubic splines with break sequence `b(1) < ...  < b(l+1)`? Assuming the dimension of this linear space is `m`, the map `F` should set up a linear one-to-one

correspondence between m-vectors and elements of *S*. But that is exactly what csape(b, . ,'var') does.

To be explicit, consider the following function F:

```
function s = F(c)
s = csape(b,c,'var');
```

For given vector c (of the same length as b), it provides the *unique* "natural" cubic spline with break sequence b that takes the value c(i) at b(i), i=1:l+1. The uniqueness is key. It ensures that the correspondence between the vector c and the resulting spline F(c) is one-to-one. In particular, m equals length(b). More than that, because the value *f(t)* of a function *f* at a point *t* depends linearly on *f*, this uniqueness ensures that F(c) depends linearly on c (because c equals fnval(F(c),b) and the inverse of an invertible linear map is again a linear map).

The One-line Solution

Putting it all together, you arrive at the following code

```
csape(b,y(:).'/fnval(csape(b,eye(length(b)),'var'),x(:).'),...
'var')
```

for the least-squares approximation by "natural" cubic splines with break sequence b.

The Need for Proper Extrapolation

Let's try it on some data, the census data, say, which is provided in MATLAB by the command

```
load census
```

and which supplies the years, 1790:10:1990, as cdate and the values as pop. Use the break sequence 1810:40:1970.

```
b = 1810:40:1970;
s = csape(b, ...
pop(:)'/fnval(csape(b,eye(length(b)),'var'),cdate(:)'),'var');
fnplt(s, [1750,2050],2.2);
```

```
hold on
plot(cdate,pop,'or');
hold off
```

Have a look at Least-Squares Approximation by "Natural" Cubic Splines With Three Interior Breaks on page 11-6 which shows, in thick blue, the resulting approximation, along with the given data.

This looks like a good approximation, -- except that it doesn't look like a "natural" cubic spline. A "natural" cubic spline, to recall, must be linear to the left of its first break and to the right of its last break, and this approximation satisfies neither condition. This is due to the following facts.

The "natural" cubic spline interpolant to given data is provided by `csape` in ppform, with the interval spanned by the data sites its basic interval. On the other hand, evaluation of a ppform outside its basic interval is done, in MATLAB `ppval` or Curve Fitting Toolbox spline function `fnval`, by using the relevant polynomial end piece of the ppform, i.e., by full-order extrapolation. In case of a "natural" cubic spline, you want instead second-order extrapolation. This means that you want, to the left of the first break, the straight line that agrees with the cubic spline in value and slope at the first break. Such an extrapolation is provided by `fnxtr`. Because the "natural" cubic spline has zero second derivative at its first break, such an extrapolation is even third-order, i.e., it satisfies three matching conditions. In the same way, beyond the last break of the cubic spline, you want the straight line that agrees with the spline in value and slope at the last break, and this, too, is supplied by `fnxtr`.
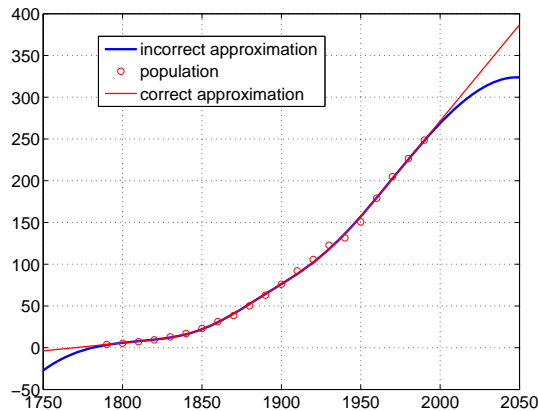
Least-Squares Approximation by "Natural" Cubic Splines With Three Interior Breaks

The Correct One-Line Solution

The following one-line code provides the correct least-squares approximation to data (x,y) by "natural" cubic splines with break sequence b:

```
fnxtr(csape(b,y(:).'/ ...
    fnval(fnxtr(csape(b,eye(length(b)),'var')),x(:).'),'var'))
```

But it is, admittedly, a rather long line.

The following code uses this correct formula and plots, in a thinner, red line, the resulting approximation on top of the earlier plots, as shown in Least-Squares Approximation by "Natural" Cubic Splines With Three Interior Breaks on page 11-6.

```
 ss = fnxtr(csape(b,pop(:)'/ ...
  fnval(fnxtr(csape(b,eye(length(b)),'var')),cdate(:)'),'var'));
hold on, fnplt(ss,[1750,2050],1.2,'r'),grid, hold off
legend('incorrect approximation','population', ...
'correct approximation')
```

Least-Squares Approximation by Cubic Splines

The one-line solution works perfectly if you want to approximate by the space S of all cubic splines with the given break sequence b. You don't even have to use the Curve Fitting Toolbox spline functions for this because you can rely on the MATLAB spline. You know that, with c a sequence containing two more entries than does b, spline(b,c) provides the unique cubic spline with break sequence b that takes the value c(i+1) at b(i), all i, and takes the slope c(1) at b(1), and the slope c(end) at b(end). Hence, spline(b,.) is a basis map for S.

More than that, you know that spline(b,c,xi) provides the value(s) at xi of this interpolating spline. Finally, you know that spline can handle vector-valued data. Therefore, the following one-line code constructs the least-squares approximation by cubic splines with break sequence b to data (x,y) :

```
spline(b,y(:)'/spline(b,eye(length(b)),x(:)'))
```

Solving A Nonlinear ODE

This section discusses these aspects of a nonlinear ODE problem:

- "Problem" on page 11-8
- "Approximation Space" on page 11-8
- "Discretization" on page 11-9
- "Numerical Problem" on page 11-9
- "Linearization" on page 11-10
- "Linear System to Be Solved" on page 11-10
- "Iteration" on page 11-11

You can run this example: "Solving a Nonlinear ODE with a Boundary Layer by Collocation".

Problem

Consider the nonlinear singularly perturbed problem:

$$\varepsilon D^2 g(x) + \left(g(x) \right)^2 = 1 \quad on \quad [0..1]$$

$$Dg(0) = g(1) = 0$$

Approximation Space

Seek an approximate solution by collocation from C^1 piecewise cubics with a suitable break sequence; for instance,

```
breaks = (0:4)/4;
```

Because cubics are of order 4, you have

```
k = 4;
```

Obtain the corresponding knot sequence as

```
knots = augknt(breaks,k,2);
```

This gives a quadruple knot at both 0 and 1, which is consistent with the fact that you have cubics, i.e., have order 4.

This implies that you have

```
n = length(knots)-k;
n = 10;
```

i.e., 10 degrees of freedom.

Discretization

You collocate at two sites per polynomial piece, i.e., at eight sites altogether. This, together with the two side conditions, gives us 10 conditions, which matches the 10 degrees of freedom.

Choose the two Gaussian sites for each interval. For the *standard* interval [–0.5,0.5] of length 1, these are the two sites

```
gauss = .5773502692*[-1/2; 1/2];
```

From this, you obtain the whole collection of collocation sites by

```
ninterv = length(breaks)-1;
temp = ((breaks(2:ninterv+1)+breaks(1:ninterv))/2);
temp = temp([1 1],:) + gauss*diff(breaks);
colsites = temp(:).';
```

Numerical Problem

With this, the numerical problem you want to solve is to find $y \in S_{4,knots}$ that satisfies the nonlinear system

$$Dy(0) = 0$$
$$(y(x))^2 + \varepsilon D^2 y(x) = 1 \text{ for } x \in \text{ colsites}$$
$$y(1) = 0$$

Linearization

If y is your current approximation to the solution, then the linear problem for the supposedly better solution z by Newton's method reads

$$Dz(0) = 0$$

$$w_0(x)z(x) + \varepsilon D^2 z(x) = b(x) \text{ for } x \in \text{ colsites}$$

$$z(1) = 0$$

with $w_0(x) = 2y(x), b(x) = (y(x))^2 + 1$. In fact, by choosing

$$w_0(1) := 1, \; w_1(0) := 1$$
$$w_1(x) := 0, \; w_2(x) := \varepsilon \text{ for } x \in \text{ colsites}$$

and choosing all other values of w_0, w_1, w_2, b not yet specified to be zero, you can give your system the uniform shape

$$w_0(x)z(x) + w_1(x)Dz(x) + w_2(x)D^2 z(x) = b(x), \quad \text{for} \quad x \in \text{sites}$$

with

```
sites = [0,colsites,1];
```

Linear System to Be Solved

Because $z \epsilon S_{4,\text{knots}}$, convert this last system into a system for the B-spline coefficients of z. This requires the values, first, and second derivatives at every $x \epsilon sites$ and for all the relevant B-splines. The command `spcol` was expressly written for this purpose.

Use `spcol` to supply the matrix

```
colmat = ...
spcol(knots,k,brk2knt(sites,3));
```

From this, you get the collocation matrix by combining the row triple of `colmat` for x using the weights $w_0(x), w_1(x), w_2(x)$ to get the row for x of the actual matrix. For this, you need a current approximation y. Initially, you get it by interpolating some reasonable initial guess from your piecewise-polynomial

space at the `sites`. Use the parabola x^2-1, which satisfies the end conditions as the initial guess, and pick the matrix from the full matrix `colmat`. Here it is, in several cautious steps:

```
intmat = colmat([2 1+(1:(n-2))*3,1+(n-1)*3],:);
coefs = intmat\[0 colsites.*colsites-1 0].';
y = spmak(knots,coefs.');
```

Plot the initial guess, and turn hold on for subsequent plotting:

```
fnplt(y,'g');
legend('Initial Guess (x^2-1)','location','NW');
axis([-0.01 1.01 -1.01 0.01]);
hold on
```

Iteration

You can now complete the construction and solution of the linear system for the improved approximate solution z from your current guess y. In fact, with the initial guess y available, you now set up an iteration, to be terminated when the change $z-y$ is *small enough*. Choose a relatively mild $\varepsilon = .1$.

```
tolerance = 6.e-9;
epsilon = .1;
while 1
   vtau = fnval(y,colsites);
   weights=[0 1 0;
            [2*vtau.' zeros(n-2,1) repmat(epsilon,n-2,1)];
            1 0 0];
   colloc = zeros(n,n);
   for j=1:n
      colloc(j,:) = weights(j,:)*colmat(3*(j-1)+(1:3),:);
   end
   coefs = colloc\[0 vtau.*vtau+1 0].';
   z = spmak(knots,coefs.');
   fnplt(z,'k');
   maxdif = max(max(abs(z.coefs-y.coefs)));
   fprintf('maxdif = %g\n',maxdif)
   if (maxdif<tolerance), break, end
% now reiterate
   y = z;
```

```
end
legend({'Initial Guess (x^2-1)' 'Iterates'},'location','NW');
```

The resulting printout of the errors is:

```
maxdif = 0.206695
maxdif = 0.01207
maxdif = 3.95151e-005
maxdif = 4.43216e-010
```

If you now decrease ε, you create more of a boundary layer near the right endpoint, and this calls for a nonuniform mesh.

Use newknt to construct an appropriate finer mesh from the current approximation:

```
knots = newknt(z, ninterv+1); breaks = knt2brk(knots);
knots = augknt(breaks,4,2);
n = length(knots)-k;
```

From the new break sequence, you generate the new collocation site sequence:

```
ninterv = length(breaks)-1;
temp = ((breaks(2:ninterv+1)+breaks(1:ninterv))/2);
temp = temp([1 1], :) + gauss*diff(breaks);
colpnts = temp(:).';
sites = [0,colpnts,1];
```

Use spcol to supply the matrix

```
colmat = spcol(knots,k,sort([sites sites sites]));
```

and use your current approximate solution z as the initial guess:

```
intmat = colmat([2 1+(1:(n-2))*3,1+(n-1)*3],:);
y = spmak(knots,[0 fnval(z,colpnts) 0]/intmat.');
```

Thus set up, divide ε by 3 and repeat the earlier calculation, starting with the statements

```
tolerance=1.e-9;
while 1
```

```
vtau=fnval(y,colpnts);
.
.
.
```

Repeated passes through this process generate a sequence of solutions, for ε = 1/10, 1/30, 1/90, 1/270, 1/810. The resulting solutions, ever flatter at 0 and ever steeper at 1, are shown in the example plot. The plot also shows the final break sequence, as a sequence of vertical bars. To view the plots, run the example "Solving a Nonlinear ODE with a Boundary Layer by Collocation".

In this example, at least, newknt has performed satisfactorily.

Construction of the Chebyshev Spline

This section discusses these aspects of the Chebyshev spline construction:

- "What Is a Chebyshev Spline?" on page 11-14
- "Choice of Spline Space" on page 11-14
- "Initial Guess" on page 11-15
- "Remez Iteration" on page 11-16

What Is a Chebyshev Spline?

The *Chebyshev spline* $C=C_t=C_{k,t}$ of order k for the knot sequence $t=(t_i:\ i=1:n+k)$ is the unique element of $S_{k,t}$ of max-norm 1 that maximally oscillates on the interval $[t_k..t_{n+1}]$ and is positive near t_{n+1}. This means that there is a unique strictly increasing n-sequence τ so that the function $C=C_t\epsilon S_{k,t}$ given by $C(\tau_i)=(-1)^{n-1}$, all i, has max-norm 1 on $[t_k..t_{n+1}]$. This implies that $\tau_1=t_k,\tau_n=t_{n+1}$, and that $t_i < \tau_i < t_{k+i}$, for all i. In fact, $t_{i+1} \le \tau_i \le t_{i+k-1}$, all i. This brings up the point that the knot sequence is assumed to make such an inequality possible, i.e., the elements of $S_{k,t}$ are assumed to be continuous.

In short, the Chebyshev spline C looks just like the Chebyshev polynomial. It performs similar functions. For example, its extreme sites τ are particularly good sites to interpolate at from $S_{k,t}$ because the norm of the resulting projector is about as small as can be; see the toolbox command `chbpnt`.

You can run the example Construction of a Chebyshev Spline to construct C for a particular knot sequence t.

Choice of Spline Space

You deal with cubic splines, i.e., with order

```
k = 4;
```

and use the break sequence

```
breaks = [0 1 1.1 3 5 5.5 7 7.1 7.2 8];
lp1 = length(breaks);
```

and use simple interior knots, i.e., use the knot sequence

```
t = breaks([ones(1,k) 2:(lp1-1) lp1(:,ones(1,k))]);
n = length(t)-k;
```

Note the quadruple knot at each end. Because `k = 4`, this makes [0..8] = [breaks(1)..breaks(lp1)] the interval $[t_k..t_{n+1}]$ of interest, with `n = length(t)-k` the dimension of the resulting spline space $S_{k,t}$. The same knot sequence would have been supplied by

```
t=augknt(breaks,k);
```

Initial Guess

As the initial guess for the τ, use the knot averages

$$t_i = (t_{i+1} + \ldots + t_{i+k-1})/(k-1)$$

recommended as good interpolation site choices. These are supplied by

```
tau=aveknt(t,k);
```

Plot the resulting first approximation to C, i.e., the spline c that satisfies $c(\tau_i)=(-1)^{n-i}$, all i:

```
b = cumprod(repmat(-1,1,n)); b = b*b(end);
c = spapi(t,tau,b);
fnplt(c,'-.')
grid
```

Here is the resulting plot.
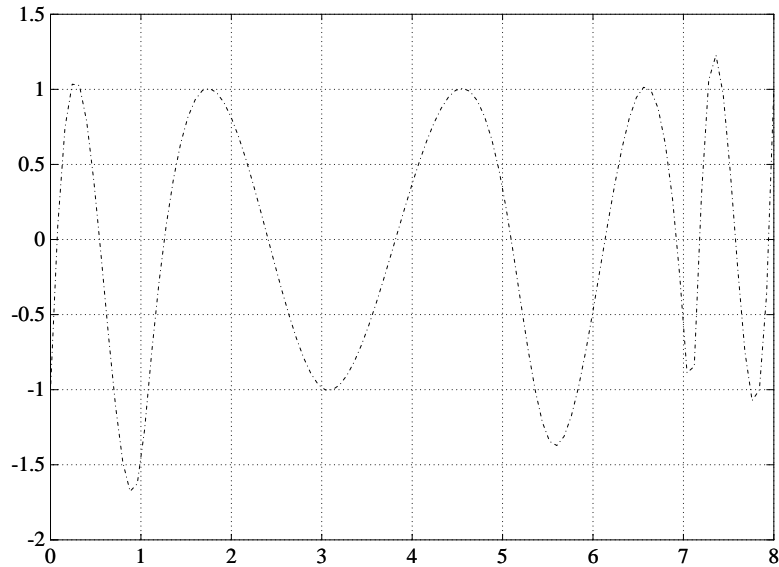
First Approximation to a Chebyshev Spline

Remez Iteration

Starting from this approximation, you use the Remez algorithm to produce a sequence of splines converging to C. This means that you construct new τ as the extrema of your current approximation c to C and try again. Here is the entire loop.

You find the new interior τ_i as the zeros of Dc, i.e., the first derivative of c, in several steps. First, differentiate:

```
Dc = fnder(c);
```

Next, take the zeros of the control polygon of Dc as your first guess for the zeros of Dc. For this, you must take apart the spline Dc.

```
[knots,coefs,np,kp] = fnbrk(Dc,'knots','coefs','n','order');
```

The control polygon has the vertices `(tstar(i),coefs(i))`, with `tstar` the knot averages for the spline, provided by `aveknt`:

```
tstar = aveknt(knots,kp);
```

Here are the zeros of the resulting control polygon of `Dc`:

```
npp = (1:np-1);
guess = tstar(npp) -coefs(npp).*(diff(tstar)./diff(coefs));
```

This provides already a very good first guess for the actual zeros.

Refine this estimate for the zeros of *Dc* by two steps of the secant method, taking `tau` and the resulting `guess` as your first approximations. First, evaluate *Dc* at both sets:

```
sites = tau(ones(4,1),2:n-1);
sites(1,:) = guess;
values = zeros(4,n-2);
values(1:2,:) = reshape(fnval(Dc,sites(1:2,:)),2,n-2);
```

Now come two steps of the secant method. You guard against division by zero by setting the function value difference to 1 in case it is zero. Because *Dc* is strictly monotone near the sites sought, this is harmless:

```
for j=2:3
  rows = [j,j-1];Dcd=diff(values(rows,:));
  Dcd(find(Dcd==0)) = 1;
  sites(j+1,:) = sites(j,:) ...
          -values(j,:).*(diff(sites(rows,:))./Dcd);
  values(j+1,:) = fnval(Dc,sites(j+1,:));
end
```

The check

```
max(abs(values.'))
  ans = 4.1176 5.7789 0.4644 0.1178
```

shows the improvement.

Now take these sites as your new `tau`,

```
tau = [tau(1) sites(4,:) tau(n)];
```

and check the extrema values of your current approximation there:

```
extremes = abs(fnval(c, tau));
```

The difference

```
max(extremes)-min(extremes)
  ans = 0.6905
```

is an estimate of how far you are from total leveling.

Construct a new spline corresponding to the new choice of `tau` and plot it on top of the old:

```
c = spapi(t,tau,b);
sites = sort([tau (0:100)*(t(n+1)-t(k))/100]);
values = fnval(c,sites);
hold on, plot(sites,values)
```

The following code turns on the grid and plots the locations of the extrema.

```
grid on
plot( tau(2:end-1), zeros( 1, np-1 ), 'o' )
hold off
legend( 'Initial Guess', 'Current Guess', 'Extreme Locations',...
 'location', 'NorthEastOutside' );
```

Following is the resulting figure (legend not shown).
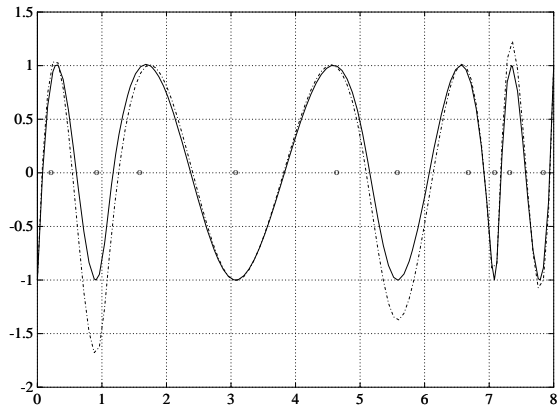
A More Nearly Level Spline

If this is not close enough, one simply reiterates the loop. For this example, the next iteration already produces C to graphic accuracy.

Approximation by Tensor Product Splines

Because the toolbox can handle splines with *vector* coefficients, it is easy to implement interpolation or approximation to gridded data by tensor product splines, as the following illustration is meant to show. You can also run the example "Bivariate Tensor Product Splines".

To be sure, most tensor product spline approximation to gridded data can be obtained directly with one of the spline construction commands, like `spapi` or `csape`, in this toolbox, without concern for the details discussed in this example. Rather, this example is meant to illustrate the theory behind the tensor product construction, and this will be of help in situations not covered by the construction commands in this toolbox.

This section discusses these aspects of the tensor product spline problem:

- "Choice of Sites and Knots" on page 11-20
- "Least Squares Approximation as Function of y" on page 11-21
- "Approximation to Coefficients as Functions of x" on page 11-22
- "The Bivariate Approximation" on page 11-27
- "Switch in Order" on page 11-25
- "Approximation to Coefficients as Functions of y" on page 11-26
- "The Bivariate Approximation" on page 11-27
- "Comparison and Extension" on page 11-28

Choice of Sites and Knots

Consider, for example, least squares approximation to given data $z(i,j)=f(x(i),y(j)),i=1:Nx,j=1:Ny$. You take the data from a function used extensively by Franke for the testing of schemes for surface fitting (see R. Franke, "A critical comparison of some methods for interpolation of scattered data," *Naval Postgraduate School Techn. Rep. NPS-53-79-003*, March 1979). Its domain is the unit square. You choose a few more data sites in the *x*-direction than the *y*-direction; also, for a better definition, you use higher data density near the boundary.

```
x = sort([(0:10)/10,.03 .07, .93 .97]);
```

```
y = sort([(0:6)/6,.03 .07, .93 .97]);
[xx,yy] = ndgrid(x,y); z = franke(xx,yy);
```

Least Squares Approximation as Function of y

Treat these data as coming from a vector-valued function, namely, the function of *y* whose value at *y*(*j*) is the vector *z*(:,*j*), all *j*. For no particular reason, choose to approximate this function by a vector-valued parabolic spline, with three uniformly spaced interior knots. This means that you choose the spline order and the knot sequence for this vector-valued spline as

```
ky = 3; knotsy = augknt([0,.25,.5,.75,1],ky);
```

and then use `spap2` to provide the least squares approximant to the data:

```
sp = spap2(knotsy,ky,y,z);
```

In effect, you are finding simultaneously the discrete least squares approximation from $S_{ky,knotsy}$ to each of the `Nx` data sets

$$\left(y(j), z(i,j) \right)_{j=1}^{Ny}, \quad i = 1 : Nx$$

In particular, the statements

```
yy = -.1:.05:1.1;
vals = fnval(sp,yy);
```

provide the array `vals`, whose entry `vals(i,j)` can be taken as an approximation to the value *f*(*x*(*i*),*yy*(*j*))of the underlying function *f* at the mesh-point *x*(*i*),*yy*(*j*) because `vals(:,j)` is the value at *yy*(*j*) of the approximating spline curve in `sp`.

This is evident in the following figure, obtained by the command:

```
mesh(x,yy,vals.'), view(150,50)
```

Note the use of `vals.'`, in the `mesh` command, needed because of the MATLAB matrix-oriented view when plotting an array. This can be a serious problem in bivariate approximation because there it is customary to think of *z*(*i*, *j*) as

the function value at the point $(x(i), y(j))$, while MATLAB thinks of $z(i, j)$ as the function value at the point $(x(j), y(i))$.
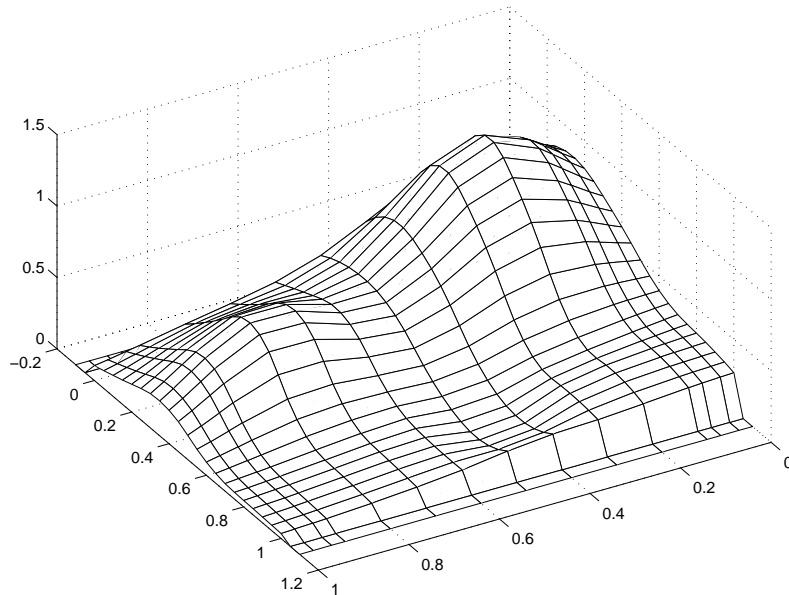
A Family of Smooth Curves Pretending to Be a Surface

Note that both the first two and the last two values on each smooth curve are actually zero because both the first two and the last two sites in yy are outside the basic interval for the spline in sp.

Note also the ridges. They confirm that you are plotting smooth curves in one direction only.

Approximation to Coefficients as Functions of x

To get an actual surface, you now have to go a step further. Look at the coefficients coefsy of the spline in sp:

```
coefsy = fnbrk(sp,'coefs');
```

Abstractly, you can think of the spline in `sp` as the function

$$y \mapsto \sum_r coefsy(:,r) B_{r,ky}(y)$$

with the *i*th entry `coefsy(i,r)` of the vector coefficient `coefsy(:,r)` corresponding to $x(i)$, for all i. This suggests approximating each coefficient vector `coefsy(q,:)` by a spline of the same order `kx` and with the same appropriate knot sequence `knotsx`. For no particular reason, this time use *cubic* splines with *four* uniformly spaced interior knots:

```
kx = 4; knotsx = augknt([0:.2:1],kx);
sp2 = spap2(knotsx,kx,x,coefsy.');
```

Note that `spap2(knots,k,x,fx)` expects `fx(:,j)` to be the datum at `x(j)`, i.e., expects each *column* of `fx` to be a function value. To fit the datum `coefsy(q, :)` at $x(q)$, for all q, present `spap2` with the *transpose* of `coefsy`.

The Bivariate Approximation

Now consider the transpose of the coefficients `cxy` of the resulting spline *curve*:

```
coefs = fnbrk(sp2,'coefs').';
```

It provides the *bivariate* spline approximation

$$(x,y) \mapsto \sum_q \sum_r coefs(q,r) B_{q,kx}(x) B_{r,ky}(y)$$

to the original data

$$\big(x(i),y(j)\big) \mapsto z\big(x(i),y(j)\big), i = 1:Nx, j = 1:Ny$$

To plot this spline surface over a grid, e.g., the grid

```
xv = 0:.025:1; yv = 0:.025:1;
```

you can do the following:

```
values = spcol(knotsx,kx,xv)*coefs*spcol(knotsy,ky,yv).';
mesh(xv,yv,values.'), view(150,50);
```

This results in the following figure.
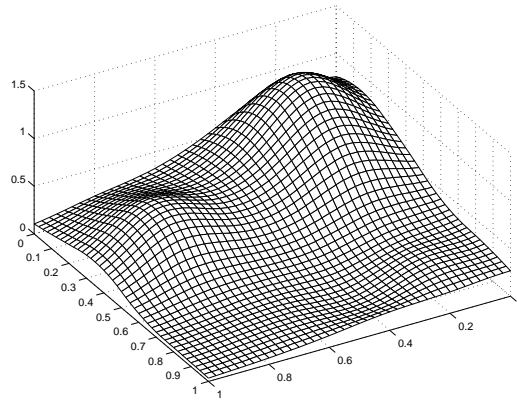
Spline Approximation to Franke's Function

This makes good sense because spcol(knotsx,kx,xv) is the matrix whose (i,q)th entry equals the value $B_{q,kx}(xv(i))$ at $xv(i)$ of the qth B-spline of order kx for the knot sequence knotsx.

Because the matrices spcol(knotsx,kx,xv) and spcol(knotsy,ky,yv) are banded, it may be more efficient, though perhaps more memory-consuming, for *large* xv and yv to make use of fnval, as follows:

```
value2 = ...
  fnval(spmak(knotsx,fnval(spmak(knotsy,coefs),yv).'),xv).';
```

This is, in fact, what happens internally when fnval is called directly with a tensor product spline, as in

```
value2 = fnval(spmak({knotsx,knotsy},coefs),{xv,yv});
```

Here is the calculation of the relative error, i.e., the difference between the given data and the value of the approximation at those data sites as compared with the magnitude of the given data:

```
errors = z - spcol(knotsx,kx,x)*coefs*spcol(knotsy,ky,y).';
disp( max(max(abs(errors)))/max(max(abs(z))) )
```

The output is `0.0539`, perhaps not too impressive. However, the coefficient array was only of size `8  6`

```
disp(size(coefs))
```

to fit a data array of size `15  11`.

```
disp(size(z))
```

Switch in Order

The approach followed here seems *biased*, in the following way. First think of the given data z as describing a vector-valued function of *y*, and then treat the matrix formed by the vector coefficients of the approximating curve as describing a vector-valued function of *x*.

What happens when you take things in the opposite order, i.e., think of z as describing a vector-valued function of *x*, and then treat the matrix made up from the vector coefficients of the approximating curve as describing a vector-valued function of *y*?

Perhaps surprisingly, the final approximation is the same, up to roundoff. Here is the numerical experiment.

Least Squares Approximation as Function of x

First, fit a spline curve to the data, but this time with *x* as the independent variable, hence it is the *rows* of z that now become the data values. Correspondingly, you must supply z.`'`, rather than z, to `spap2`,

```
spb = spap2(knotsx,kx,x,z.');
```

thus obtaining a spline approximation to all the curves $(x ; z (:, j))$. In particular, the statement

```
valsb = fnval(spb,xv).';
```

provides the matrix valsb, whose entry valsb(i, j) can be taken as an approximation to the value $f(xv(i),y(j))$ of the underlying function f at the mesh-point $(xv(i),y(j))$. This is evident when you plot valsb using mesh:

```
mesh(xv,y,valsb.'), view(150,50)
```
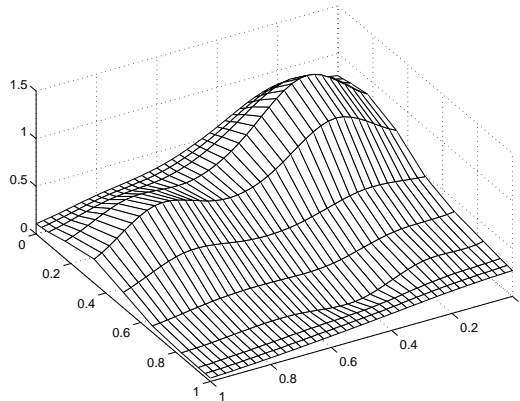

Another Family of Smooth Curves Pretending to Be a Surface

Note the ridges. They confirm that you are, once again, plotting smooth curves in one direction only. But this time the curves run in the other direction.

Approximation to Coefficients as Functions of y

Now comes the second step, to get the actual surface. First, extract the coefficients:

```
coefsx = fnbrk(spb,'coefs');
```

Then fit each coefficient vector coefsx(r,:) by a spline of the same order ky and with the same appropriate knot sequence knotsy:

```
spb2 = spap2(knotsy,ky,y,coefsx.');
```

Note that, once again, you need to transpose the coefficient array from spb, because spap2 takes the columns of its last input argument as the data values.

Correspondingly, there is now no need to transpose the coefficient array `coefsb` of the resulting *curve*:

```
coefsb = fnbrk(spb2,'coefs');
```

The Bivariate Approximation

The claim is that `coefsb` equals the earlier coefficient array `coefs`, up to round-off, and here is the test:

```
disp( max(max(abs(coefs - coefsb))) )
```

The output is `1.4433e-15`.

The explanation is simple enough: The coefficients `c` of the spline *s* contained in `sp = spap2(knots,k,x,y)` depend *linearly* on the input values *y*. This implies, given that both `c` and `y` are 1-row matrices, that there is some matrix $A=A_{\text{knots},k,x}$ so that

$$c = yA_{\text{knots},k,x}$$

for any data `y`. This statement even holds when `y` is a *matrix*, of size *d*-by-*N*, say, in which case each datum $y(:,j)$ is taken to be a point in R^d, and the resulting spline is correspondingly *d*-vector-valued, hence its coefficient array `c` is of size *d*-by-`n`, with `n = length(knots)-k`.

In particular, the statements

```
sp = spap2(knotsy,ky,y,z);
coefsy =fnbrk(sp,'coefs');
```

provide us with the matrix `coefsy` that satisfies

$$\text{coefs}y = z.A_{\text{knotsy},\text{ky},\text{y}}$$

The subsequent computations

```
 sp2 = spap2(knotsx,kx,x,coefsy.');
 coefs = fnbrk(sp2,'coefs').';
```

generate the coefficient array `coefs`, which, taking into account the two transpositions, satisfies

$$\text{coefs} = \left(\left(z A_{\text{knotsy,ky,y}} \right)' . A_{\text{knotsx,kx,x}} \right)'$$

$$= \left(A_{\text{knotsx,kx,x}} \right)' . z . A_{\text{knotsy,ky,y}}$$

In the second, alternative, calculation, you first computed

```
spb = spap2(knotsx,kx,x,z.');
coefsx = fnbrk(spb,'coefs');
```

hence `coefsx=z'.A`$_{\text{knotsx,kx,x}}$. The subsequent calculation

```
spb2 = spap2(knotsy,ky,y,coefsx.');
coefsb = fnbrk(spb,'coefs');
```

then provided

$$\text{coefsb} = \text{coefsx}.' . A_{\text{knotsy,ky,y}} = \left(A_{\text{knotsx,kx,x}} \right).' . z . A_{\text{knotsy,ky},y}$$

Consequently, `coefsb = coefs`.

Comparison and Extension

The second approach is more symmetric than the first in that transposition takes place in each call to `spap2` and nowhere else. This approach can be used for approximation to gridded data in any number of variables.

If, for example, the given data over a *three*-dimensional grid are contained in some three-dimensional array `v` of size `[Nx,Ny,Nz]`, with `v(i,j,k)` containing the value $f(x(i),y(j),z(k))$, then you would start off with

```
coefs = reshape(v,Nx,Ny*Nz);
```

Assuming that `nj = knotsj - kj`, for `j = x,y,z`, you would then proceed as follows:

```
sp = spap2(knotsx,kx,x,coefs.');
```

```
coefs = reshape(fnbrk(sp,'coefs'),Ny,Nz*nx);
sp = spap2(knotsy,ky,y,coefs.');
coefs = reshape(fnbrk(sp,'coefs'),Nz,nx*ny);
sp = spap2(knotsz,kz,z,coefs.');
coefs = reshape(fnbrk(sp,'coefs'),nx,ny*nz);
```

See Chapter 17 of *PGS* or [C. de Boor, "Efficient computer manipulation of tensor products," *ACM Trans. Math. Software 5* (1979), 173–182; Corrigenda, 525] for more details. The same references also make clear that there is nothing special here about using least squares approximation. Any approximation process, including spline interpolation, whose resulting approximation has coefficients that depend linearly on the given data, can be extended in the same way to a multivariate approximation process to gridded data.

This is exactly what is used in the spline construction commands `csapi`, `csape`, `spapi`, `spaps`, and `spap2`, when gridded data are to be fitted. It is also used in `fnval`, when a tensor product spline is to be evaluated on a grid.

Splines Glossary

The Glossary consists of these sections:

- "Introduction" on page A-2
- "List of Terms" on page A-3

Introduction

This glossary provides brief definitions of the basic mathematical terms and notation used in this guide. But, in contrast to standard glossaries, the terms do not appear here in alphabetical order. This is not much of a disadvantage because the glossary is quite short (and all the terms appear in the Index in any case). The order is carefully chosen to have the explanation of each term only use terms discussed earlier.

In this way, you may, the first time around, choose to read the entire glossary from start to finish, for a cohesive introduction to these terms.

List of Terms

Intervals

Because MATLAB uses the notation [a,b] to indicate a matrix with the two columns, a and b, this guide uses the notation [a .. b] to indicate the closed interval with endpoints a and b. This guide does the same for open and half-open intervals. For example, [a .. b) denotes the interval that includes its left endpoint, a, and excludes its right endpoint, b.

Vectors

A *d-vector* is a list of d real numbers, i.e., a point in $\Re^d$. In MATLAB, a d-vector is stored as a matrix of size [1,d], i.e., as a row-vector, or as a matrix of size [d,1], i.e., as a column-vector. In the Curve Fitting Toolbox spline functions, vectors are column vectors.

Functions

In this toolbox, the term *function* is used in its mathematical sense, and so describes any rule that associates, to each element of a certain set called its *domain*, some element in a certain set called its *target*. Common examples in this toolbox are polynomials and splines. But even a point x in $\Re^d$, i.e., a d-vector, may be thought of as a function, namely the function, with domain the set {1,...,d} and target the real numbers $\Re$, that, for $i = 1,...,d$, associates to i the real number $x(i)$.

The *range* of a function is the set of its values.

There are scalar-valued, vector-valued, matrix-valued, and *ND*-valued splines. Scalar-valued functions have the real numbers $\Re$ (or, more generally, the complex numbers) as their target, while d-vector-valued functions have $\Re^d$ as their target; if, more generally, **d** is a vector of positive integers, then **d**-valued functions have the **d**-dimensional real arrays as their target. Curve Fitting Toolbox spline functions can deal with univariate and multivariate functions. The former have some real interval, or, perhaps, all of $\Re$ as their domain, while m-variate functions have some subset, or perhaps all, of $\Re^m$ as their domain.

Placeholder notation

If f is a *bivariate* function, and y is some specific value of its second variable, then

$$f(\cdot, y)$$

is the *univariate* function whose value at x is $f(x,y)$.

Curves and surfaces vs. functions

In this toolbox, the term *function* usually refers to a scalar-valued function. A vector-valued function is called here a:

curve if its domain is some interval

surface if its domain is some rectangle

To be sure, to a mathematician, a curve is *not* a vector-valued function on some interval but, rather, the range of such a (continuous) function, with the function itself being just one of infinitely many possible parametrizations of that curve.

Tensor products

A bivariate *tensor product* is any weighted sum of products of a function in the first variable with a function in the second variable, i.e., any function of the form

$$f(x, y) = \sum_i \sum_j a(i, j) g_i(x) h_j(y).$$

More generally, an *m*-variate tensor product is any weighted sum of products $g_1(x_1)g_2(x_2)...g_m(x_m)$ of *m* univariate functions.

Polynomials

A univariate scalar-valued polynomial is specified by the list of its polynomial coefficients. The length of that list is the order of that polynomial, and, in this toolbox, the list is always stored as a row vector. Hence an *m*-list of polynomials of order *k* is always stored as a matrix of size `[m,k]`.

The coefficients in a list of polynomial coefficients are listed from "highest" to "lowest", to conform to the MATLAB convention, as in the command `polyval(a,x)`. To recall: assuming that `x` is a scalar and that `a` has `k` entries, this command returns the number

$$a(1)x^{k-1} + a(2)x^{k-2} + \cdots + a(k-1)x + a(k).$$

In other words, the command treats the list a as the coefficients in a *power form*. For reasons of numerical stability, such a coefficient list is treated in this toolbox, more generally, as the coefficients in a *shifted*, or, *local* power form, for some given *center* c. This means that the value of the polynomial at some point x is supplied by the command polyval(a,x-c).

A vector-valued polynomial is treated in exactly the same way, except that now each polynomial coefficient is a vector, say a ***d***-vector. Correspondingly, the coefficient list now becomes a matrix of size [d,k].

Multivariate polynomials appear in this toolbox mainly as *tensor products*. Assuming first, for simplicity, that the polynomial in question is scalar-valued but *m*-variate, this means that its coefficient "list" *a* is an *m*-dimensional array, of size [$k_1,...,k_m$] say, and its value at some *m*-vector *x* is, correspondingly, given by

$$\sum_{i_1=1}^{k_1} \cdots \sum_{i_m=1}^{k_m} a(i_1,...,i_m)\left(x\left(i_1\right)-c\left(i_1\right)\right)^{k_1-i_1} \cdots \left(x\left(i_m\right)-c\left(i_m\right)\right)^{k_m-i_m},$$

for some "center" *c*.

Piecewise-polynomials

A *piecewise-polynomial* function refers to a function put together from polynomial pieces. If the function is univariate, then, for some strictly increasing sequence $\xi_1 < ... < \xi_{l+1}$, and for $i = 1{:}l$, it agrees with some polynomial p_i on the interval $[\xi_i .. \xi_{i+1})$. Outside the interval $[\xi_1 .. \xi_{l+1})$, its value is given by its first, respectively its last, polynomial piece. The ξ_i are its *breaks*. All the multivariate piecewise-polynomials in this toolbox are tensor products of univariate ones.

B-splines

In this toolbox, the term *B-spline* is used in its original meaning only, as given to it by its creator, I. J. Schoenberg, and further amplified in his basic 1966 article with Curry, and used in *PGS* and many other books on splines. According to Schoenberg, the B-spline with knots $t_j, ..., t_{j+k}$ is given by the following somewhat obscure formula (see, e.g., IX(1) in *PGS*):

$$B_{j,k}(x) = B\big(x \mid t_j,...,t_{j+k}\big) = \big(t_{j+k} - t_j\big)\big[t_j,...,t_{j+k}\big](x - \cdot)_+^{k-1}.$$

To be sure, this is only one of several reasonable normalizations of the B-spline, but it is the one used in this toolbox. It is chosen so that

$$\sum_{j=1}^{n} B_{j,k}(x) = 1, \quad t_k \le x \le t_{n+1}.$$

But, instead of trying to understand the above formula for the B-spline, look at the reference pages for the GUI `bspligui` for some of the basic properties of the B-spline, and use that GUI to gain some firsthand experience with this intriguing function. Its most important property for the purposes of this toolbox is also the reason Schoenberg used the letter B in its name:

Every space of (univariate) piecewise-polynomials of a given order has a Basis consisting of B-*splines (hence the "B" in B-spline).*

Splines

Consider the set

$$S := \Pi_{\xi,k}^{\mu}$$

of all (scalar-valued) piecewise-polynomials of order k with breaks $\xi_1 < ... < \xi_{l+1}$ that, for $i = 2...l$, may have a jump across ξ_i in its μ_ith derivative but have no jump there in any lower order derivative. This set is a linear space, in the sense that any scalar multiple of a function in S is again in S, as is the sum of any two functions in S.

Accordingly, S contains a basis (in fact, infinitely many bases), that is, a sequence $f_1,...,f_n$ so that every f in S can be written *uniquely* in the form

$$f(x) = \sum_{j=1}^{n} f_j(x)a_j,$$

for suitable coefficients a_j. The number n appearing here is the *dimension* of the linear space S. The coefficients a_j are often referred to as the *coordinates* of f with respect to this basis.

In particular, according to the Curry-Schoenberg Theorem, our space S has a basis consisting of B-splines, namely the sequence of all

B-splines of the form $B\left(\cdot\,|\,t_j,...,t_{j+k}\right)$, $j = 1...n$, with the knot sequence t obtained from the break sequence ξ and the sequence μ by the following conditions:

- Have both ξ_1 and ξ_{l+1} occur in t exactly k times

- For each $i = 2:l$, have ξ_i occur in t exactly $k - \mu_i$ times

- Make sure the sequence is nondecreasing and only contains elements from ξ

Note the correspondence between the multiplicity of a knot and the smoothness of the spline across that knot. In particular, at a simple knot, that is a knot that appears exactly once in the knot sequence, only the $(k-1)$st derivative may be discontinuous.

Rational splines

A *rational spline* is any function of the form $r(x) = s(x)/w(x)$, with both s and w splines and, in particular, w a scalar-valued spline, while s often is vector-valued. In this toolbox, there is the additional requirement that both s and w be of the same form and even of the same order, and with the same knot or break sequence. This makes it possible to store the rational spline r as the ordinary spline R whose value at x is the vector $[s(x);w(x)]$. It is easy to obtain r from R. For example, if `v` is the value of R at x, then `v(1:end-1)/v(end)` is the value of r at x. As another example, consider getting derivatives of r from those of R. Because $s = wr$, Leibniz' rule tells us that

$$D^m s = \sum_{j=0}^{m}\binom{m}{j}D^j w D^{m-j} r.$$

Hence, if `v(:,j)` contains $D^{j-1}R(x)$, $j = 1...m + 1$, then

$$\left(\left(\left(v(1:\mathrm{end}-1,m+1)-\sum_{j=1}^{m}\binom{m}{j}v(\mathrm{end},j+1)v(1:\mathrm{end}-1,j+1)\right)\bigg/v(\mathrm{end},1)\right)\right)$$

provides the value of $D^m R(x)$.

Thin-plate splines

A bivariate thin-plate spline is of the form

$$f(x) = \sum_{j=1}^{n-3} \varphi\left(\left|x - c_j\right|^2\right) a_j + x(1)a_{n-2} + x(2)a_{n-1} + a_n,$$

with $\varphi(t) = t \log t$ a univariate function, and $\|y\|$ denoting the Euclidean length of the vector y. The sites c_j are called the *centers*, and the radially symmetric function $\psi(x) := \varphi(|x|^2)$ is called the *basis function*, of this particular stform.

Interpolation

Interpolation is the construction of a function f that matches given *data values*, y_i, at given *data sites*, x_i, in the sense that $f(x_i) = y_i$, all i.

The interpolant, f, is usually constructed as the unique function of the form

$$f(x) = \sum_j f_j(x) a_j$$

that matches the given data, with the functions f_j chosen "appropriately". Many considerations might enter that choice. One of these considerations is sure to be that one can match in this way arbitrary data. For example, polynomial interpolation is popular because, for arbitrary n *data points* (x_i, y_i) with distinct data sites, there is exactly one polynomial of order $n - 1$ that matches these data. Explicitly, choose the f_j in the above "model" to be

$$f_j(x) = \prod_{i \neq j} (x - x_i),$$

which is an $n - 1$ degree polynomial for each j. $f_j(x_i) = 0$ for every $i \neq j$, but $f_j(x_j) \neq 0$ as long as the x_i are all distinct. Set $a_j = y_j/f_j(x_j)$ so that

$$f(x_j) = f_j(x_j) a_j = y_j \text{ for all } j.$$

In spline interpolation, one chooses the f_j to be the n consecutive B-splines $B_j(x) = B(x \mid t_j,...,t_{j+k})$, $j = 1{:}n$, of order k for some knot sequence $t_1 \leq t_2 \leq ... \leq t_{n+k}$. For this choice, there is the following important theorem.

Schoenberg-Whitney Theorem

Let $x_1 < x_2 < ... < x_n$. For arbitrary corresponding values y_i, $i = 1...n$, there exists exactly one spline f of order k with knot sequence t_j, $j = 1...n+k$, so that $f(x_i) = y_i$, $i = 1...n$ if and only if the sites satisfy the Schoenberg-Whitney conditions of order k with respect to that knot sequence t, namely

$$t_i \leq x_i \leq t_{i+k}, \ i = 1...n,$$

with equality allowed only if the knot in question has multiplicity k, i.e., appears k times in t. In that case, the spline being constructed may have a jump discontinuity across that knot, and it is its limit from the right or left at that knot that matches the value given there.

Least-squares approximation

In least-squares approximation, the data may be matched only approximately. Specifically, the linear system

$$f(x_i) = \sum_j f_j(x_i) a_j = y_i, \ i = 1...n,$$

is solved in the least-squares sense. In this, some weighting is involved, i.e., the coefficients aj are determined so as to minimize the error measure

$$E(f) = \sum_i w_i \left| y_i - f(x_i) \right|^2$$

for certain nonnegative weights w_i at the user's disposal, with the default being to have all these weights the same.

Smoothing

In spline smoothing, one also tries to make such an error measure small, but tries, at the same time, to keep the following roughness measure small,

$$F\left(D^m f\right) = \int\limits_{x_1}^{x_n} \lambda(x) \left| D^m f(x) \right|^2 dx,$$

with λ a nonnegative weight function that is usually just the constant function 1, and $D^m f$ the mth derivative of f. The competing claims of small $E(f)$ and small $F(D^m f)$ are mediated by a smoothing parameter, for example, by minimizing

$$\rho E(f) + F\left(D^m f\right) \ \text{ or } \ \mathsf{p} E(f) + (1 - \mathsf{p}) F\left(D^m f\right),$$

for some choice of ρ or of p, and over all f for which this expression makes sense.

Remarkably, if the roughness weight λ is constant, then the unique minimizer f is a spline of order $2m$, with knots only at the data sites, and all the interior knots simple, and with its derivatives of orders $m,...,2m-2$ equal to zero at the two extreme data sites, the so-called "natural" end conditions. The larger the smoothing parameter $\rho \geq 0$ or $\mathsf{p} \ \epsilon \ [0..1]$ used, the more closely f matches the given data, and the larger is its mth derivative.

For data values y_i at sites c_i in the *plane*, one uses instead the error measure and roughness measure

$$E(f) = \sum_i \left|y_i - f\left(c_i\right)\right|^2, \ \ F\left(D^2 f\right) = \int\left(\left|D_{11}f\right|^2 + 2\left|D_{12}f\right|^2 + \left|D_{22}f\right|^2\right),$$

and, correspondingly, the minimizer of the sum $\rho E(f) + F(D^2 f)$ is not a polynomial spline, but is a thin-plate spline.

Note that the unique minimizer of $\rho E(f) + F(D^2 f)$ for given $0 < \rho < \infty$ is also the unique minimizer of $\mathsf{p} E(f) + (1 - \mathsf{p}) F(D^2 f)$ for $\mathsf{p} = \rho/(1 + \rho) \ \epsilon \ (0 .. 1)$ and *vice versa*.

2D, 3D, ND

Terms such as "a 2D problem" or "a 3D problem" are not used in this toolbox, because they are not well defined. For example a 2D problem could be any one of the following:

- Points on some curve, where you must construct a spline curve, i.e., a vector-valued spline function of one variable.

- Points on the graph of some function, where you must construct a scalar-valued spline function of one variable.

- Data sites in the plane, where you must construct a bivariate scalar-valued spline function.

A "3D problem" is similarly ambiguous. It could involve a curve, a surface, a function of three variables, ... Better to classify problems by the domain and target of the function(s) to be constructed.

Almost all the spline construction commands in this toolbox can deal with ND-valued data, meaning that the data values are ND-arrays. If `d` is the size of such an array, then the resulting spline is called `d`-valued.

12

Function Reference

Interactive Fitting

cftool   Open Curve Fitting Tool

Linear and Nonlinear Regression

argnames	Input argument names of `cfit`, `sfit`, or `fittype` object
category	Category of fit of `cfit`, `sfit`, or `fittype` object
cftool	Open Curve Fitting Tool
coeffnames	Coefficient names of `cfit`, `sfit`, or `fittype` object
dependnames	Dependent variable of `cfit`, `sfit`, or `fittype` object
feval	Evaluate `cfit`, `sfit`, or `fittype` object
fit	Fit curve or surface to data
fitoptions	Create or modify fit options structure
fittype	Fit type for curve and surface fitting
formula	Formula of `cfit`, `sfit`, or `fittype` object
get	Get fit options structure property names and values
indepnames	Independent variable of `cfit`, `sfit`, or `fittype` object
islinear	Determine if `cfit`, `sfit`, or `fittype` object is linear
numargs	Number of input arguments of `cfit`, `sfit`, or `fittype` object
numcoeffs	Number of coefficients of `cfit`, `sfit`, or `fittype` object
probnames	Problem-dependent parameter names of `cfit`, `sfit`, or `fittype` object
set	Assign values in fit options structure

setoptions	Set model fit options
type	Name of `cfit`, `sfit`, or `fittype` object

Interpolation

cftool	Open Curve Fitting Tool
feval	Evaluate `cfit`, `sfit`, or `fittype` object
fit	Fit curve or surface to data
fittype	Fit type for curve and surface fitting

Smoothing

cftool	Open Curve Fitting Tool
fit	Fit curve or surface to data
fitoptions	Create or modify fit options structure
fittype	Fit type for curve and surface fitting
get	Get fit options structure property names and values
set	Assign values in fit options structure
smooth	Smooth response data

Fit Postprocessing

cfit	Constructor for cfit object
cftool	Open Curve Fitting Tool
coeffvalues	Coefficient values of cfit or sfit, object
confint	Confidence intervals for fit coefficients of cfit or sfit object
differentiate	Differentiate cfit or sfit object
excludedata	Exclude data from fit
feval	Evaluate cfit, sfit, or fittype object
fitoptions	Create or modify fit options structure
get	Get fit options structure property names and values
integrate	Integrate cfit object
plot	Plot cfit or sfit object
predint	Prediction intervals for cfit or sfit object
prepareSurfaceData	Prepare data inputs for surface fitting
probvalues	Problem-dependent parameter values of cfit or sfit object
quad2d	Numerically integrate sfit object
set	Assign values in fit options structure
sfit	Constructor for sfit object

Splines

Spline Construction

csape	Cubic spline interpolation with end conditions
csapi	Cubic spline interpolation
csaps	Cubic smoothing spline
cscvn	"Natural" or periodic interpolating cubic spline curve
getcurve	Interactive creation of cubic spline curve
ppmak	Put together spline in ppform
rpmak	Put together rational spline
rscvn	Piecewise biarc Hermite interpolation
rsmak	Put together rational spline for standard geometric shapes
spap2	Least-squares spline approximation
spapi	Spline interpolation
spaps	Smoothing spline
spcrv	Spline curve by uniform subdivision
splinetool	Experiment with some spline approximation methods

spmak	Put together spline in B-form
stmak	Put together function in stform
tpaps	Thin-plate smoothing spline

Spline Postprocessing

fn2fm	Convert to specified form
fnbrk	Name and part(s) of form
fnchg	Change part(s) of form
fncmb	Arithmetic with function(s)
fnder	Differentiate function
fndir	Directional derivative of function
fnint	Integrate function
fnjmp	Jumps, i.e., $f(x+)$-$f(x-)$
fnmin	Minimum of function in given interval
fnplt	Plot function
fnrfn	Refine partition of form
fntlr	Taylor coefficients or polynomial
fnval	Evaluate function
fnxtr	Extrapolate function
fnzeros	Find zeros of function in given interval

Spline Breaks, Knots, and Sites

aptknt	Acceptable knot sequence
augknt	Augment knot sequence
aveknt	Provide knot averages

brk2knt	Convert breaks with multiplicities into knots
chbpnt	Good data sites, Chebyshev-Demko points
knt2brk, knt2mlt	Convert knots to breaks and their multiplicities
newknt	New break distribution
optknt	Knot distribution "optimal" for interpolation
sorted	Locate sites with respect to mesh sites

Spline Utilities

bkbrk	Part(s) of almost block-diagonal matrix
bspline	Plot B-spline and its polynomial pieces
franke	Franke's bivariate test function
slvblk	Solve almost block-diagonal linear system
spcol	B-spline collocation matrix
splpp, sprpp	Taylor coefficients from local B-coefficients
spterms	Explain spline terms
stcol	Scattered translates collocation matrix
subplus	Positive part
titanium	Titanium test data

Functions — Alphabetical List

aptknt

Purpose	Acceptable knot sequence
Syntax	knots = aptknt(tau,k) [knots,k] = aptknt(tau,k)

Description knots = aptknt(tau,k) returns a knot sequence suitable for interpolation at the data sites tau by splines of order k with that knot sequence, provided tau has at least k entries, is nondecreasing, and satisfies tau(i)<tau(i+k-1) for all i. In that case, there is exactly one spline of order k with knot sequence knots that matches given values at those sites. This is so because the sequence knots returned satisfies the Schoenberg-Whitney conditions

```
knots(i) < tau(i) < knots(i+k),          i=1:length(tau)
```

with equality only at the extreme knots, each of which occurs with exact multiplicity k.

If tau has fewer than k entries, then k is reduced to the value length(tau). An error results if tau fails to be nondecreasing and/or tau(i) equals tau(i+k-1) for some i.

[knots,k] = aptknt(tau,k) also returns the actual k used (which equals the smaller of the input k and length(tau)).

Examples If tau is equally spaced, e.g., equal to linspace(a,b,n) for some n>=4, and y is a sequence of the same size as tau, then sp = spapi(aptknt(tau,4),tau,y) gives the cubic spline interpolant with the not-a-knot end condition. This is the same cubic spline as produced by the command spline(tau,y), but in B-form rather than ppform.

Algorithms The (k-1)-point averages sum(tau(i+1:i+k-1))/(k-1) of the sequence tau, as supplied by aveknt(tau,k), are augmented by a k-fold tau(1) and a k-fold tau(end). In other words, the command gives the same result as augknt([tau(1),aveknt(tau,k),tau(end)],k), provided tau has at least k entries and k is greater than 1.

Cautionary Note If tau is very nonuniform, then use of the resulting knot sequence for interpolation to data at the sites tau may lead to unsatisfactory results.

See Also augknt | aveknt | newknt | optknt

argnames

Purpose	Input argument names of cfit, sfit, or fittype object
Syntax	args = argnames(fun)
Description	args = argnames(fun) returns the input argument (variable and coefficient) names of the cfit, sfit, or fittype object fun as an n-by-1 cell array of strings args, where n = numargs(fun).

Examples
```
f = fittype('a*x^2+b*exp(n*x)');
nargs = numargs(f)
nargs =
     4
args = argnames(f)
args =
    'a'
    'b'
    'n'
    'x'
```

See Also   fittype | formula | numargs

Purpose   Augment knot sequence

Syntax   augknt(knots,k)
                   augknt(knots,k,mults)
                   [augknot,addl] = augknt(...)

Description   augknt(knots,k) returns a nondecreasing and augmented knot
                   sequence that has the first and last knot with exact multiplicity k. (This
                   may actually shorten the knot sequence.))

                   augknt(knots,k,mults) makes sure that the augmented knot
                   sequence returned will, in addition, contain each interior knot mults
                   times. If mults has exactly as many entries as there are interior
                   knots, then the *j*th one will appear mults(j) times. Otherwise, the
                   uniform multiplicity mults(1) is used. If knots is strictly increasing,
                   this ensures that the splines of order k with knot sequence augknot
                   satisfy k-mults(j) smoothness conditions across knots(j+1),
                   j=1:length(knots)-2.

                   [augknot,addl] = augknt(...) also returns the number addl of knots
                   added on the left. (This number may be negative.)

Examples   If you want to construct a cubic spline on the interval [a..b], with two
                   continuous derivatives, and with the interior break sequence xi, then
                   augknt([a,b,xi],4) is the knot sequence you should use.

                   If you want to use Hermite cubics instead, i.e., a cubic spline with
                   only one continuous derivative, then the appropriate knot sequence is
                   augknt([a,xi,b],4,2).

                   augknt([1 2 3 3 3],2) returns the vector [1 1 2 3 3], as does
                   augknt([3 2 3 1 3],2). In either case, addl would be 1.

aveknt

Purpose	Provide knot averages

Syntax

```
tstar = aveknt(t,k)
```

Description   tstar = aveknt(t,k) returns the averages of successive k-1 knots, i.e., the sites

$$t_i^* := (t_{i+1} + \cdots + t_{i+k-1}) / (k-1), \quad i = 1:n$$

which are recommended as good interpolation site choices when

interpolating from splines of order k with knot sequence $t = (t_i)_{i=1}^{n+k}$.

Examples   aveknt([1 2 3 3 3],3) returns the vector [2.5000 3.0000], while aveknt([1 2 3],3) returns the empty vector.

With k and the strictly increasing sequence breaks given, the statements

```
t = augknt(breaks,k); x = aveknt(t);
sp = spapi(t,x,sin(x));
```

provide a spline interpolant to the sine function on the interval [breaks(1)..breaks(end)].

For sp the B-form of a scalar-valued univariate spline function, and with tstar and a computed as

```
tstar = aveknt(fnbrk(sp,'knots'),fnbrk(sp,'order'));
a = fnbrk(sp,'coefs');
```

the points (tstar(*i*), a(*i*)) constitute the *control points* of the spline, i.e., the vertices of the spline's *control polygon*.

See Also   aptknt | chbpnt | optknt

Purpose   Part(s) of almost block-diagonal matrix

Syntax   `[nb,rows,ncols,last,blocks] = bkbrk(blokmat)`
                   `bkbrk(blokmat)`

Description   `[nb,rows,ncols,last,blocks] = bkbrk(blokmat)` returns the
                   details of the almost block-diagonal matrix contained in `blokmat`,
                   with `rows` and `last` `nb`-vectors, and `blocks` a matrix of size
                   `[sum(rows),ncols]`.

                   This utility program is not likely to be of interest to the casual user. It
                   is used in `slvblk` to decode the information, provided by `spcol`, about
                   a spline collocation matrix in an almost block diagonal form especially
                   suited for splines. But `bkbrk` can also decode the almost block-diagonal
                   form used in [1].

                   `bkbrk(blokmat)` returns nothing, but the details are printed out. This
                   is of use when trying to understand what went wrong with such a
                   matrix.

References   [1] C. de Boor and R. Weiss. "SOLVEBLOK: A package for solving
                   almost block diagonal linear systems." *ACM Trans. Mathem. Software
                   6* (1980), 80–87.

See Also   `slvblk` | `spcol`

brk2knt

Purpose	Convert breaks with multiplicities into knots
Syntax	[knots,index] = brk2knt(breaks,mults)

Description [knots,index] = brk2knt(breaks,mults) returns the sequence knots that is the sequence breaks but with breaks(i) occurring mults(i) times, all i. In particular, breaks(i) will not appear unless mults(i)>O. If, as one would expect, breaks is a strictly increasing sequence, then knots contains each breaks(i) exactly mults(i) times.

If mults does not have exactly as many entries as does breaks, then all mults(i) are set equal to mults(1).

If, as one would expect, breaks is strictly increasing and all multiplicities are positive, then, for each i, index(i) is the first place in knots at which breaks(i) appears.

Examples The statements

```
t = [1 1 2 2 2 3 4 5 5];
[xi,m] = knt2brk(t);
tt = brk2knt(xi,m)
```

give [1 2 3 4 5] for xi, [2 3 1 1 2] for m, and, finally, t for tt.

See Also augknt

Purpose   Experiment with B-spline as function of its knots

Syntax   bspligui

Description   bspligui starts a graphical user interface (GUI) for exploring how a B-spline depends on its knots. As you add, move, or delete knots, you see the B-spline and its first three derivatives change accordingly.

You observe the following basic facts about the B-spline with knot sequence $t_0 \leq \cdots \leq t_k$:

- The B-spline is positive on the open interval $(t_0..t_k)$. It is zero at the end knots, t_0 and t_k, unless they are knots of multiplicity k. The B-spline is also zero outside the closed interval $[t_0..t_k]$, but that part of the B-spline is not shown in the GUI.

- Even at its maximum, the B-spline is never bigger than 1. It reaches the value 1 inside the interval $(t_0..t_k)$ only at a knot of multiplicity at least k–1. On the other hand, that maximum cannot be arbitrarily small; it seems smallest when there are no interior knots.

- The B-spline is piecewise polynomial of order k, i.e., its polynomial pieces all are of degree $<k$. For k = 1:4, you can even observe that all its nonzero polynomial pieces are of exact degree $k - 1$, by looking at the first three derivatives of the B-spline. This means that the degree goes up/down by 1 every time you add/delete a knot.

- Each knot t_j is a break for the B-spline, but it is permissible for several knots to coincide. Therefore, the number of nontrivial polynomial pieces is maximally k (when all the knots are different) and minimally 1 (when there are no "interior" knots), and any number between 1 and k is possible.

- The smoothness of the B-spline across a break depends on the multiplicity of the corresponding knot. If the break occurs in the knot sequence m times, then the $(k$–$m)$th derivative of the B-spline has a jump across that break, while all derivatives of order lower than $(k$–$m)$ are continuous across that break. Thus, by varying the

multiplicity of a knot, you can control the smoothness of the B-spline across that knot.

- As one knot approaches another, the highest derivative that is continuous across both develops a jump and the higher derivatives become unbounded. But nothing dramatic happens in any of the lower-order derivatives.

- The B-spline is *bell-shaped* in the following sense: if the first derivative is not identically zero, then it has exactly one sign change in the interval $(t_0..t_k)$, hence the B-spline itself is *unimodal*, meaning that it has exactly one maximum. Further, if the second derivative is not identically zero, then it has exactly two sign changes in that interval. Finally, if the third derivative is not identically zero, then it has exactly three sign changes in that interval. This illustrates the fact that, for $j = 0:k - 1$, if the jth derivative is not identically zero, then it has exactly j sign changes in the interval $(t_0..t_k)$; it is this property that is meant by the term "bell-shaped". For this claim to be strictly true, one has to be careful with the meaning of "sign change" in case there are knots with multiplicities. For example, the $(k-1)$st derivative is piecewise constant, hence it cannot have $k-1$ sign changes in the straightforward sense unless there are k polynomial pieces, i.e., unless all the knots are simple.

See Also bspline | chbpnt | spcol

Purpose	Plot B-spline and its polynomial pieces

Syntax

```
bspline(t)
bspline(t,window)
pp = bspline(t)
```

Description   bspline(t) plots the B-spline with knot sequence t, as well as the polynomial pieces of which it is composed.

bspline(t,window) does the plotting in the subplot window specified by window; see the MATLAB command subplot for details.

pp = bspline(t) plots nothing but returns the ppform of the B-spline.

Examples   The statement pp=fn2fm(spmak(t,1),'pp') has the same effect as the statement pp=bspline(t).

See Also   bspligui

category

Purpose	Category of fit of `cfit`, `sfit`, or `fittype` object
Syntax	`cname = category(fun)`
Description	`cname = category(fun)` returns the fit category `cname` of the `cfit`, `sfit`, or `fittype` object `fun`, where `cname` is one of `'custom'`, `'interpolant'`, `'library'`, or `'spline'`.

Examples

```
f1 = fittype('a*x^2+b*exp(n*x)');
category(f1)
ans =
custom

f2 = fittype('pchipinterp');
category(f2)
ans =
interpolant

f3 = fittype('fourier4');
category(f3)
ans =
library

f4 = fittype('smoothingspline');
category(f4)
ans =
spline
```

See Also   `fittype` | `type`

How To   • "List of Library Models for Curve and Surface Fitting" on page 4-13

Purpose   Constructor for `cfit` object

Syntax   `cfun = cfit(ffun,coeff1,coeff2,...)`

Description `cfun = cfit(ffun,coeff1,coeff2,...)` constructs the `cfit` object
`cfun` using the model type specified by the `fittype` object `ffun` and the
coefficient values `coeff1`, `coeff2`, etc.

Note `cfit` is called by the `fit` function when fitting `fittype` objects to
data. To create a `cfit` object that is the result of a regression, use `fit`.

You should only call `cfit` directly if you want to assign values to
coefficients and problem parameters of a `fittype` object *without*
performing a fit.

Examples
```
f = fittype('a*x^2+b*exp(n*x)')
f =
    General model:
      f(a,b,n,x) = a*x^2+b*exp(n*x)
c = cfit(f,1,10.3,-1e2)
c =
    General model:
      c(x) = a*x^2+b*exp(n*x)
    Coefficients:
      a =            1
      b =         10.3
      n =         -100
```

See Also   `fit` | `fittype` | `feval`

cftool

Purpose   Open Curve Fitting Tool

Syntax
```
cftool
cftool( x, y )
cftool( x, y, z )
cftool( x, y, [], w )
cftool( x, y, z, w )
cftool( filename )
cftool -v1
cftool( '-v1', x, y )
cftool( '-v1', x, y, w )
```

Description   cftool opens Curve Fitting Tool or brings focus to the tool if it is already open.

cftool(x, y) creates a curve fit to x input and y output. x and y must be numeric, have two or more elements, and have the same number of elements. cftool opens Curve Fitting Tool if necessary.

cftool(x, y, z) creates a surface fit to x and y inputs and z output. x, y, and z must be numeric, have two or more elements, and have compatible sizes. Sizes are compatible if x, y, and z all have the same number of elements or x and y are vectors, z is a 2D matrix, length(x) = n, and length(y) = m where [m,n] = size(z). cftool opens Curve Fitting Tool if necessary.

cftool(x, y, [], w) creates a curve fit with weights w. w must be numeric and have the same number of elements as x and y.

cftool(x, y, z, w) creates a surface fit with weights w. w must be numeric and have the same number of elements as z.

cftool(filename) loads the surface fitting session in filename into Curve Fitting Tool. The filename should have the extension .sfit.

cftool -v1 opens the legacy Curve Fitting Tool.

cftool('-v1', x, y) starts the legacy Curve Fitting Tool with an initial data set containing the x and y data you supply. x and y must be numeric vectors having the same length.

cftool('-v1', x, y, w) also includes the weight vector w in the initial data set.   w must have the same length as x and y.

Infs, NaNs, and imaginary parts of complex numbers are ignored in the data.

Curve Fitting Tool provides a flexible graphical user interface where you can interactively fit curves and surfaces to data and view plots. You can:

- Create, plot, and compare multiple fits

- Use linear or nonlinear regression, interpolation, local smoothing regression, or custom equations

- View goodness-of-fit statistics, display confidence intervals and residuals, remove outliers and assess fits with validation data

- Automatically generate code for fitting and plotting surfaces, or export fits to workspace for further analysis

How To • "Interactive Curve and Surface Fitting" on page 2-2

chbpnt

Purpose	Good data sites, Chebyshev-Demko points

Syntax
```
tau = chbpnt(t,k)
chbpnt(t,k,tol)
[tau,sp] = chbpnt(...)
```

Description   `tau = chbpnt(t,k)` are the extreme sites of the Chebyshev spline of order k with knot sequence t. These are particularly good sites at which to interpolate data by splines of order k with knot sequence t because the resulting interpolant is often quite close to the best uniform approximation from that spline space to the function whose values at tau are being interpolated.

`chbpnt(t,k,tol)` also specifies the tolerance tol to be used in the iterative process that constructs the Chebyshev spline. This process is terminated when the relative difference between the absolutely largest and the absolutely smallest local extremum of the spline is smaller than tol. The default value for tol is .001.

`[tau,sp] = chbpnt(...)` also returns, in sp, the Chebyshev spline.

Examples   `chbpnt([-ones(1,k),ones(1,k)],k)` provides (approximately) the extreme sites on the interval [–1 .. 1] of the Chebyshev polynomial of degree k-1.

If you have decided to approximate the square-root function on the interval [0 .. 1] by cubic splines, with knot sequence t as given by

```
k = 4; n = 10; t = augknt(((0:n)/n).^8,k);
```

then a good approximation to the square-root function from that specific spline space is given by

```
x = chbpnt(t,k); sp = spapi(t,x,sqrt(x));
```

as is evidenced by the near equi-oscillation of the error.

Algorithms   The Chebyshev spline for the given knot sequence and order is constructed iteratively, using the Remez algorithm, using as initial guess the spline that takes alternately the values 1 and −1 at the sequence `aveknt(t,k)`. The example "Constructing the Chebyshev Spline" gives a detailed discussion of one version of the process as applied to a particular example.

See Also   `aveknt`

coeffnames

Purpose	Coefficient names of `cfit`, `sfit`, or `fittype` object
Syntax	`coeffs = coeffnames(fun)`
Description	`coeffs = coeffnames(fun)` returns the coefficient (parameter) names of the `cfit`, `sfit`, or `fittype` object `fun` as an n-by-1 cell array of strings `coeffs`, where `n = numcoeffs(fun)`.

Examples
```
f = fittype('a*x^2+b*exp(n*x)');
ncoeffs = numcoeffs(f)
ncoeffs =
     3
coeffs = coeffnames(f)
coeffs =
    'a'
    'b'
    'n'
```

See Also   `fittype` | `formula` | `numcoeffs` | `probnames` | `coeffvalues`

coeffvalues

Purpose   Coefficient values of `cfit` or `sfit`, object

Syntax   `coeffvals = coeffvalues(fun)`

Description `coeffvals = coeffvalues(fun)` returns the values of the coefficients
(parameters) of the `cfit` object `fun` as a 1-by-n vector `coeffvals`, where
`n = numcoeffs(fun)`.

Examples
```
load census

f = fittype('poly2');
coeffnames(f)
ans =
    'p1'
    'p2'
    'p3'
formula(f)
ans =
p1*x^2 + p2*x + p3

c = fit(cdate,pop,f);
coeffvalues(c)
ans =
  1.0e+004 *
    0.0000   -0.0024    2.1130
```

See Also   coeffnames | confint | predint | probvalues

confint

Purpose	Confidence intervals for fit coefficients of `cfit` or `sfit` object

Syntax

```
ci = confint(fitresult)
ci = confint(fitresult,level)
```

Description   `ci = confint(fitresult)` returns 95% confidence bounds `ci` on the coefficients associated with the `cfit` or `sfit` object `fitresult`. `fitresult` must be an output from the `fit` function to contain the necessary information for `ci`. `ci` is a 2-by-n array where `n = numcoeffs(fitresult)`. The top row of `ci` contains the lower bound for each coefficient; the bottom row contains the upper bound.

`ci = confint(fitresult,level)` returns confidence bounds at the confidence level specified by `level`. `level` must be between `0` and `1`. The default value of `level` is `0.95`.

Tips   To calculate confidence bounds, `confint` uses R^{-1} (the inverse R factor from QR decomposition of the Jacobian), the degrees of freedom for error, and the root mean squared error. This information is automatically returned by the `fit` function and contained within `fitresult`.

If coefficients are bounded and one or more of the estimates are at their bounds, those estimates are regarded as fixed and do not have confidence bounds.

Note that you cannot calculate confidence bounds if `category(fitresult)` is `'spline'` or `'interpolant'`.

Examples
```
load census

fitresult = fit(cdate,pop,'poly2')
fitresult =
     Linear model Poly2:
       fitresult(x) = p1*x^2 + p2*x + p3
     Coefficients (with 95% confidence bounds):
       p1 =    0.006541  (0.006124, 0.006958)
       p2 =       -23.51  (-25.09, -21.93)
```

```
      p3 =  2.113e+004  (1.964e+004, 2.262e+004)

ci = confint(fitresult,0.95)
ci =

   0.0061242        -25.086         19641
   0.0069581        -21.934         22618
```

Note that `fit` and `confint` display the confidence bounds in slightly different formats.

See Also   fit | predint

csape

Purpose	Cubic spline interpolation with end conditions

Syntax

```
pp = csape(x,y)
pp = csape(x,y,conds)
```

Description pp = csape(x,y) is the ppform of a cubic spline *s* with knot sequence x that satisfies *s*(x(j)) = y(:,j) for all j, as well as an additional *end condition* at the ends (meaning the leftmost and at the rightmost data site), namely the default condition listed below. The data values y(:,j) may be scalars, vectors, matrices, even ND-arrays. Data values at the same data site are averaged.

pp = csape(x,y,conds) lets you choose the end conditions to be used, from a rather large and varied catalog, by proper choice of conds. If needed, you supply the corresponding end condition values as additional data values, with the first (last) data value taken as the end condition value at the left (right) end. In other words, in that case, *s*(x(j)) matches y(:,j+1) for all j, and the variable endcondvals used in the detailed description below is set to y(:,[1 end]). For some choices of conds, these end condition values need not be present and/or are ignored when present.

conds may be a *string* whose first character matches one of the following: 'complete' or 'clamped', 'not-a-knot', 'periodic', 'second', 'variational', with the following meanings.

'complete' or 'clamped'	Match endslopes (as given, with default as under "default").
'not-a-knot'	Make second and second-last sites inactive knots (ignoring end condition values if given).
'periodic'	Match first and second derivatives at left end with those at right end.
'second'	Match end second derivatives (as given, with default [0 0], i.e., as in 'variational').

| 'variational' | Set end second derivatives equal to zero (ignoring end condition values if given). |
| default | Match endslopes to the slope of the cubic that matches the first four data at the respective end (i.e., Lagrange). |

By giving conds as a 1-by-2 matrix instead, it is possible to specify *different* conditions at the two ends. Explicitly, the *i*th derivative, $D^i s$, is given the value endcondvals(:,*j*) at the left (*j* is 1) respectively right (*j* is 2) end in case conds(*j*) is *i*,*i* = 1:2. There are default values for conds and/or endcondvals.

Available conditions are:

clamped	$Ds(e) = $ endcondvals(:,j)	if conds(j) == 1
curved	$D^2 s(e) = $ endcondvals(:,j)	if conds(j) == 2
Lagrange	$Ds(e) = Dp(e)$	default
periodic	$D^r s(a) = D^r s(b)$, $r = 1,2$	if conds == [0 0]
variational	$D^2 s(e) = 0$	if conds(j) == 2 & endcondvals(:,j) == 0

Here, *e* is *a* (*e* is *b*), i.e., the left (right) end, in case j is 1 (j is 2), and (in the Lagrange condition) *P* is the cubic polynomial that interpolates to the given data at *e* and the three sites nearest *e*.

If conds(j) is not specified or is different from 0, 1, or 2, then it is taken to be 1 and the corresponding endcondvals(:,j) is taken to be the corresponding default value.

The default value for endcondvals(:,j) is the derivative of the cubic interpolant at the nearest four sites in case conds(j) is 1, and is 0 otherwise.

It is also possible to handle gridded data, by having x be a cell array containing *m* univariate meshes and, correspondingly, having y be an *m*-dimensional array (or an *m*+*r*-dimensional array if the function is to be *r*-valued). Correspondingly, conds is a cell array with *m* entries,

and end condition values may be correspondingly supplied in each of the *m* variables. This, as the last example below, of bicubic spline interpolation, makes clear, may require you to supply end conditions for end conditions.

This command calls on a much expanded version of the Fortran routine CUBSPL in *PGS*.

Examples   csape(x,y) provides the cubic spline interpolant with the Lagrange end conditions, while csape(x,y,[2 2]) provides the variational, or *natural* cubic spline interpolant, as does csape(x,y,'v'). csape([-1 1],[3 -1 1 6],[1 2]) provides the cubic polynomial *p* for which $Dp(-1) = 3$, $p(-1) = -1$, $p(1) = 1$, $D^2p(1) = 6$, i.e., $p(x) = x^3$. Finally, csape([-1 1],[-1 1]) provides the straight line *p* for which $p(\pm 1) = \pm 1$, i.e., $p(x) = x$.

End conditions other than the ones listed earlier can be handled along the following lines. Suppose that you want to enforce the condition

$$\lambda(s) := aDs(e) + bD^2s(e) = c$$

for given scalars *a*, *b*, and *c*, and with *e* equal to x(1). Then one could compute the cubic spline interpolant s_1 to the given data using the default end condition as well as the cubic spline interpolant s_0 to zero data and some (nontrivial) end condition at *e*, and then obtain the desired interpolant in the form

$$s = s_1 + ((c - \lambda)(s_1)) / \lambda(s_0)s_0$$

Here are the (not inconsiderable) details (in which the first polynomial piece of s_1 and s_0 is pulled out to avoid differentiating all of s_1 and s_0):

```
% Data: x and y
[x, y] = titanium();

% Scalars a, b, and c
a = -2;
b = -1;
```

```
c = 0;

% End condition at left
e = x(1);

% The cubic spline interpolant s1 to the
% given data using the default end
% condition
s1 = csape(x,y);

% The cubic spline interpolant s0 to
% zero data and some (nontrivial) end
% condition at e
s0 = csape(x,[1,zeros(1,length(y)),0],[1,0]);

% Compute the derivatives of the first
%   polynomial piece of s1 and s0
ds1 = fnder(fnbrk(s1,1));
ds0 = fnder(fnbrk(s0,1));

% Compute interpolant with desired end conditions
lam1 = a*fnval(ds1,e) + b*fnval(fnder(ds1),e);
lam0 = a*fnval(ds0,e) + b*fnval(fnder(ds0),e);
pp = fncmb(s0,(c-lam1)/lam0,s1);
```

Plot to see the results:

```
fnplt( pp, [594, 632] )
hold on
fnplt( s1, 'b--', [594, 632] )
plot( x, y, 'ro', 'MarkerFaceColor', 'r' )
hold off
axis( [594, 632, 0.62, 0.655] )
legend 'Desired end-conditions' ...
'Default end-conditions' 'Data' ...
    Location SouthEast
```

As a multivariate vector-valued example, here is a sphere, done as a parametric bicubic spline, 3D-valued, using prescribed slopes in one direction and periodic end conditions in the other:

```
x = 0:4; y=-2:2; s2 = 1/sqrt(2);

v = zeros( 3, 7, 5 );
v(1,:,:) = [1 0 s2 1 s2 0 -1].'*[1 0 -1 0 1];
v(2,:,:) = [1 0 s2 1 s2 0 -1].'*[0 1 0 -1 0];
v(3,:,:) = [0 1 s2 0 -s2 -1 0].'*[1 1 1 1 1];

sph = csape({x,y},v,{'clamped','periodic'});
values = fnval(sph,{0:.1:4,-2:.1:2});

surf( squeeze(values(1,:,:)), ...
squeeze(values(2,:,:)), squeeze(values(3,:,:)) );

axis equal
axis off
```

The lines involving fnval and surf could have been replaced by the simple command: fnplt(sph). Note that v is a 3-dimensional array, with v(:,i+1,j) the 3-vector to be matched at (x(i),y(j)), i=1:5, j=1:5. Note further that, in accordance with conds{1} being 'clamped', size(v,2) is 7 (and not 5), with the first and last entry of v(r,:,j) specifying the end slopes to be matched.

Here is a bivariate example that shows the need for supplying end conditions of end conditions when supplying end conditions in both variables. You reproduce the bicubic polynomial $g(x,y) = x^3y^3$ by complete bicubic interpolation. You then derive the needed data, including end condition values, directly from g in order to make it easier for you to see just how the end condition values must be placed. Finally, you check the result.

```
sites = {[0 1],[0 2]}; coefs = zeros(4, 4); coefs(1,1) = 1;
g = ppmak(sites,coefs);
Dxg = fnval(fnder(g,[1 0]),sites);
```

```
Dyg = fnval(fnder(g,[0 1]),sites);
Dxyg = fnval(fnder(g,[1 1]),sites);

f = csape(sites,[Dxyg(1,1),   Dxg(1,:),   Dxyg(1,2); ...
                 Dyg(:,1), fnval(g,sites), Dyg(:,2) ; ...
                 Dxyg(2,1),   Dxg(2,:),   Dxyg(2,2)], ...
                                       {'complete','complete'});
if any(squeeze(fnbrk(f,'c'))-coefs)
    disp( 'this is wrong' )
end
```

Algorithms   The relevant tridiagonal linear system is constructed and solved using the sparse matrix capabilities of MATLAB.

Cautionary Note   csape recognizes that you supplied explicit end condition values by the fact that you supplied exactly two more data values than data sites. In particular, even when using different end conditions at the two ends, if you wish to supply an end condition value at one end, you must also supply one for the other end.

See Also   csapi | spapi | spline

csapi

Purpose	Cubic spline interpolation

Syntax

```
pp=csapi(x,y)
values = csapi(x,y,xx)
```

Description   pp=csapi(x,y) returns the ppform of a cubic spline *s* with knot
sequence x that takes the value y(:,j) at x(j) for j=1:length(x).
The values y(:,j) can be scalars, vectors, matrices, even ND-arrays.
Data points with the same data site are averaged and then sorted by
their sites. With x the resulting sorted data sites, the spline s satisfies
the not-a-knot end conditions, namely $\mathrm{jump}_{x(2)}D^3s = 0 = \mathrm{jump}_{x(end-1)}D^3s$
(with D^3s the third derivative of *s*).

If x is a cell array, containing sequences x1, ..., xm, of lengths n1, ..., nm
respectively, then y is expected to be an array, of size [n1,...,nm]
(or of size [d,n1,...,nm] if the interpolant is to be d-valued). In that
case, pp is the ppform of an m-cubic spline interpolant *s* to such data. In
particular, now $s(xl(i_1), ..., xm(i_m))$ equals $y(:,i_1, ..., i_m)$ for i_1 = 1:nl, ...,
i_m = 1:nm.

You can use the structure pp, in fnval, fnder, fnplt, etc, to evaluate,
differentiate, plot, etc, this interpolating cubic spline.

values = csapi(x,y,xx) is the same as fnval(csapi(x,y),xx), i.e.,
the values of the interpolating cubic spline at the sites specified by
xx are returned.

This command is essentially the MATLAB function spline, which, in
turn, is a stripped-down version of the Fortran routine CUBSPL in PGS,
except that csapi (and now also spline) accepts vector-valued data and
can handle gridded data.

Examples   See the example "Spline Interpolation" for various examples.

Up to rounding errors, and assuming that x is a vector with at least four
entries, the statement pp = csapi(x,y) should put the same spline
into pp as does the statement

```
pp = fn2fm(spapi(augknt(x([1 3:(end-2) end]),4),x,y),'pp');
```

except that the description of the spline obtained this second way will use no break at `x(2)` and `x(n-1)`.

Here is a simple bivariate example, a bicubic spline interpolant to the Mexican Hat function being plotted:

```
x =.0001+[-4:.2:4]; y = -3:.2:3;
[yy,xx] = meshgrid(y,x); r = pi*sqrt(xx.^2+yy.^2); z = sin(r)./r;
bcs = csapi( {x,y}, z ); fnplt( bcs ), axis([-5 5 -5 5 -.5 1])
```

Note the reversal of `x` and `y` in the call to `meshgrid`, needed because MATLAB likes to think of the entry `z(i,j)` as the value at `(x(j),y(i))` while this toolbox follows the Approximation Theory standard of thinking of `z(i,j)` as the value at `(x(i),y(j))`. Similar caution has to be exerted when values of such a bivariate spline are to be plotted with the aid of the MATLAB `mesh` function, as is shown here (note the use of the transpose of the matrix of values obtained from `fnval`).

```
xf = linspace(x(1),x(end),41); yf = linspace(y(1),y(end),41);
mesh(xf, yf, fnval( bcs, {xf, yf}).')
```

Algorithms   The relevant tridiagonal linear system is constructed and solved, using the MATLAB sparse matrix capability.

The not-a-knot end condition is used, thus forcing the first and second polynomial piece of the interpolant to coincide, as well as the second-to-last and the last polynomial piece.

See Also   csape | spapi | spline

csaps

Purpose   Cubic smoothing spline

Syntax
```
pp = csaps(x,y)
csaps(x,y,p)
[...,p] = csaps(...)
csaps(x,y,p,[],w)
values = csaps(x,y,p,xx)
csaps(x,y,p,xx,w)
[...] = csaps({x1,...,xm},y,...)
```

Description   pp = csaps(x,y) returns the ppform of a cubic smoothing spline
f to the given data x,y, with the value of *f* at the data site x(j)
approximating the data value y(:,j), for j=1:length(x). The values
may be scalars, vectors, matrices, even ND-arrays. Data points with the
same site are replaced by their (weighted) average, with its weight the
sum of the corresponding weights.

This smoothing spline *f* minimizes

$$P \sum_{j=1}^{n} w(j) \left| y(:,j) - f(x(j)) \right|^2 + (1-p) \int \lambda(t) \left| D^2 f(t) \right|^2 \, dt$$

Here, $|z|^2$ stands for the sum of the squares of all the entries of *z*, *n* is
the number of entries of x, and the integral is over the smallest interval
containing all the entries of x. The default value for the weight vector
w in the *error measure* is ones(size(x)). The default value for the
piecewise constant weight function λ in the *roughness measure* is the
constant function 1. Further, D^2f denotes the second derivative of the
function *f*. The default value for the *smoothing parameter*, p, is chosen
in dependence on the given data sites x.

If the smoothing spline is to be evaluated outside its basic interval, it
must first be properly extrapolated, by the command pp = fnxtr(pp),
to ensure that its second derivative is zero outside the interval spanned
by the data sites.

csaps(x,y,p) lets you supply the smoothing parameter. The smoothing parameter determines the relative weight you would like to place on the contradictory demands of having *f* be smooth *vs* having *f* be close to the data. For p = 0, *f* is the least-squares straight line fit to the data, while, at the other extreme, i.e., for p = 1, *f* is the variational, or 'natural' cubic spline interpolant. As p moves from 0 to 1, the smoothing spline changes from one extreme to the other. The interesting range for p is often near $1/(1 + h^3/6)$, with *h* the average spacing of the data sites, and it is in this range that the default value for p is chosen. For uniformly spaced data, one would expect a close following of the data for p = $1(1 + h^3/60)$ and some satisfactory smoothing for p = $1/(1 + h^3/0.6)$. You can input a p > 1, but this leads to a smoothing spline even rougher than the variational cubic spline interpolant.

If the input p is negative or empty, then the default value for p is used.

[...,p] = csaps(...) also returns the value of p actually used whether or not you specified p. This is important for experimentation which you might start with [pp,p]=csaps(x,y) in order to obtain a 'reasonable' first guess for p.

If you have difficulty choosing p but have some feeling for the size of the noise in y, consider using instead spaps(x,y,tol) which, in effect, chooses p in such a way that the roughness measure

$$\int \lambda(t)\left|D^2 s(t)\right|^2 dt$$

is as small as possible subject to the condition that the error measure

$$\sum w(j)\left|y(:,j) - s(x(j))\right|^2$$

does not exceed the specified tol. This usually means that the error measure equals the specified tol.

The weight function λ in the roughness measure can, optionally, be specified as a (nonnegative) piecewise constant function, with breaks at the data sites x , by inputing for p a *vector* whose ith entry provides the value of λ on the interval (x(i-1) .. x(i)) for i=2:length(x). The first

entry of the input vector p continues to be used as the desired value of the smoothness parameter p. In this way, it is possible to insist that the resulting smoothing spline be smoother (by making the weight function larger) or closer to the data (by making the weight functions smaller) in some parts of the interval than in others.

`csaps(x,y,p,[],w)` lets you specify the weights w in the error measure, as a vector of nonnegative entries of the same size as x.

`values = csaps(x,y,p,xx)` is the same as `fnval(csaps(x,y,p),xx)`.

`csaps(x,y,p,xx,w)` is the same as `fnval(csaps(x,y,p,[],w),xx)`.

`[...] = csaps({x1,...,xm},y,...)` provides the ppform of an m-variate tensor-product smoothing spline to data on a rectangular grid. Here, the first argument is a cell-array, containing the vectors x1, ..., xm, of lengths n1, ..., nm, respectively. Correspondingly, y is an array of size [n1,...,nm] (or of size [d,n1,...,nm] in case the data are d-valued), with $y(:,i_1, ...,i_m)$ the given (perhaps noisy) value at the grid site $xl(i_1), ...,xm(i_m)$.

In this case, p if input must be a cell-array with m entries or else an m-vector, except that it may also be a scalar or empty, in which case it is taken to be the cell-array whose m entries all equal the p input. The optional second output argument will always be a cell-array with m entries.

Further, w if input must be a cell-array with m entries, with w{i} either empty, to indicate the default choice, or else a nonnegative vector of the same size as xi.

Examples

Example 1.

```
x = linspace(0,2*pi,21); y = sin(x)+(rand(1,21)-.5)*.1;
pp = csaps(x,y, .4, [], [ones(1,10), repmat(5,1,10), 0] );
```

returns a smooth fit to the (noisy) data that is much closer to the data in the right half, because of the much larger error weight there, except for the last data point, for which the weight is zero.

```
pp1 = csaps(x,y, [.4,ones(1,10),repmat(.2,1,10)], [], ...
             [ones(1,10), repmat(5,1,10), 0]);
```

uses the same data, smoothing parameter, and error weight but chooses
the roughness weight to be only .2 in the right half of the interval and
gives, correspondingly, a rougher but better fit there, except for the
last data point, which is ignored.

A plot showing both examples for comparison can now be obtained by

```
fnplt(pp); hold on, fnplt(pp1,'r--'), plot(x,y,'ok'), hold off
title(['cubic smoothing spline, with right half treated ',...
       'differently:'])
xlabel(['blue: larger error weights; ', ...
        'red dashed: also smaller roughness weights'])
```

The resulting plot is shown below.
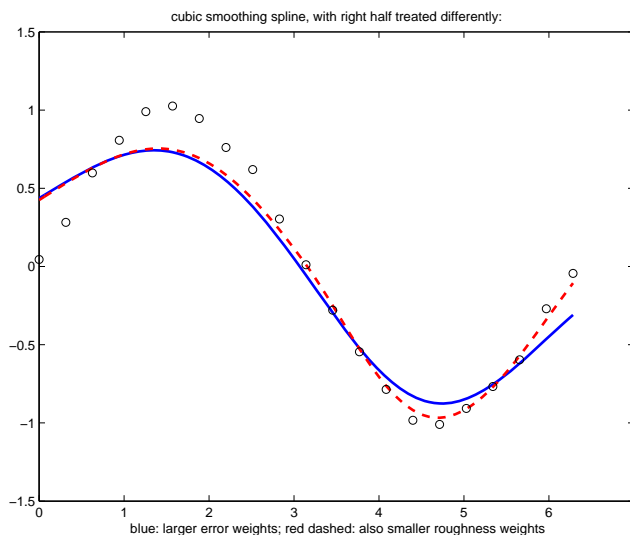
Example 2. This bivariate example adds some uniform noise, from the interval [-1/2 .. 1/2], to values of the MATLAB `peaks` function on a 51-by-61 uniform grid, obtain smoothed values for these data from `csaps`, along with the smoothing parameters chosen by `csaps`, and then plot these smoothed values.

```
x = {linspace(-2,3,51),linspace(-3,3,61)};
[xx,yy] = ndgrid(x{1},x{2}); y = peaks(xx,yy);
rand('state',0), noisy = y+(rand(size(y))-.5);
[smooth,p] = csaps(x,noisy,[],x);
surf(x{1},x{2},smooth.'), axis off
```

Note the need to transpose the array `smooth`. For a somewhat smoother approximation, use a slightly smaller value of `p` than the one, `.9998889`, used above by `csaps`. The final plot is obtained by the following:

```
smoother = csaps(x,noisy,.996,x);
figure, surf(x{1},x{2},smoother.'), axis off
```
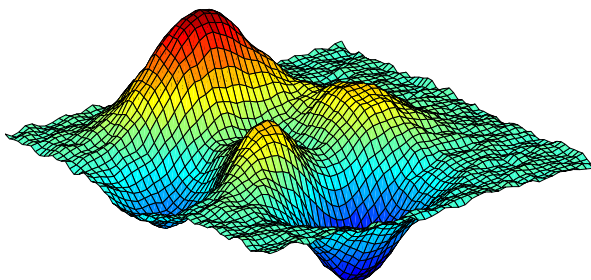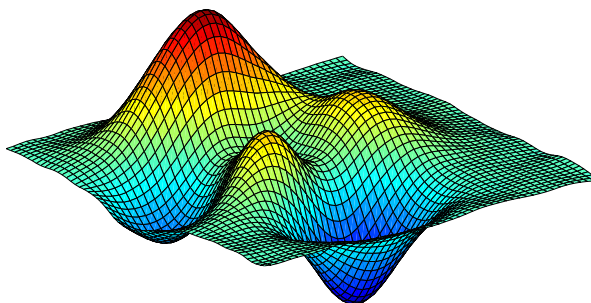
Algorithms   csaps is an implementation of the Fortran routine SMOOTH from PGS.

The default value for p is determined as follows. The calculation of the smoothing spline requires the solution of a linear system whose coefficient matrix has the form p*A + (1-p)*B, with the matrices A and B depending on the data sites x. The default value of p makes p*trace(A) equal (1-p)*trace(B).

See Also   csape | spap2 | spaps | tpaps

cscvn

Purpose   "Natural" or periodic interpolating cubic spline curve

Syntax   curve = cscvn(points)

Description curve = cscvn(points) returns a parametric variational, or *natural,* cubic spline curve (in ppform) passing through the given sequence points(:,*j*), *j* = 1:end. The parameter value *t*(*j*) for the *j*th point is chosen by Eugene Lee's [1] centripetal scheme, i.e., as accumulated square root of chord length:

$$\sum_{i<j} \sqrt{\left\| \text{points}(:,i+1) - \text{points}\,(:,i) \right\|_2}$$

If the first and last point coincide (and there are no other repeated points), then a periodic cubic spline curve is constructed. However, double points result in corners.

Examples   The following provides the plot of a questionable curve through some points (marked as circles):

```
points=[0 1 1 0 -1 -1 0 0; 0 0 1 2 1 0 -1 -2];
fnplt(cscvn(points)); hold on,
plot(points(1,:),points(2,:),'o'), hold off
```

Here is a closed curve, good for 14 February, with one double point:

```
c=fnplt(cscvn([0 .82 .92 0 0 -.92 -.82 0; .66 .9 0 ...
-.83 -.83 0 .9 .66])); fill(c(1,:),c(2,:),'r'), axis equal
```

Algorithms   The break sequence t is determined as

```
t = cumsum([0;((diff(points.').^2)*ones(d,1)).^(1/4)]).';
```

and csape (with either periodic or variational end conditions) is used to construct the smooth pieces between double points (if any).

References   [1] E. T. Y. Lee. "Choosing nodes in parametric curve interpolation." *Computer-Aided Design* 21 (1989), 363–370.

See Also   csape | fnplt | getcurve

datastats

Purpose   Data statistics

Syntax
```
xds = datastats(x)
[xds,yds] = datastats(x,y)
```

Description   `xds = datastats(x)` returns statistics for the column vector `x` to the structure `xds`. Fields in `xds` are listed in the table below.

Field	Description
num	The number of data values
max	The maximum data value
min	The minimum data value
mean	The mean value of the data
median	The median value of the data
range	The range of the data
std	The standard deviation of the data

`[xds,yds] = datastats(x,y)` returns statistics for the column vectors `x` and `y` to the structures `xds` and `yds`, respectively. `xds` and `yds` contain the fields listed in the table above. `x` and `y` must be of the same size.

Tips   If `x` or `y` contains complex values, only the real parts are used in computing the statistics. Data containing `Inf` or `NaN` are processed using the usual MATLAB rules.

Examples   Compute statistics for the census data in `census.mat`:

```
load census
[xds,yds] = datastats(cdate,pop)
xds =
        num: 21
        max: 1990
        min: 1790
```

```
      mean: 1890
    median: 1890
     range: 200
       std: 62.048
yds =
       num: 21
       max: 248.7
       min: 3.9
      mean: 85.729
    median: 62.9
     range: 244.8
       std: 78.601
```

See Also

excludedata, smooth

dependnames

Purpose	Dependent variable of cfit, sfit, or fittype object
Syntax	dep = dependnames(fun)
Description	dep = dependnames(fun) returns the (single) dependent variable name of the cfit, sfit, or fittype object fun as a 1-by-1 cell array of strings dep.

Examples

```
f1 = fittype('a*x^2+b*exp(n*x)');
dep1 = dependnames(f1)
dep1 =
    'y'

f2 = fittype('a*x^2+b*exp(n*x)','dependent','power');
dep2 = dependnames(f2)
dep2 =
    'power'
```

See Also   indepnames | fittype | formula

Purpose   Differentiate cfit or sfit object

Syntax
```
fx = differentiate(FO, X)
[fx, fxx] = differentiate(...)
[fx, fy] = differentiate(FO, X, Y)
[fx, fy] = differentiate(FO, [x, y])
[fx, fy, fxx, fxy, fyy] = differentiate(FO, ...)
```

Description   **For Curves**

fx = differentiate(FO, X) differentiates the cfit object FO at the points specified by the vector X and returns the result in fx.

[fx, fxx] = differentiate(...) also returns the second derivative in fxx.

All return arguments are the same size and shape as X.

For Surfaces

[fx, fy] = differentiate(FO, X, Y) differentiates the surface FO at the points specified by X and Y and returns the result in fx and fy.

FO is a surface fit (sfit) object generated by the fit function.

X and Y must be double-precision arrays and the same size and shape as each other.

All return arguments are the same size and shape as X and Y.

If FO represents the surface $z = f(x, y)$, then FX contains the derivatives

with respect to *x*, that is, $\dfrac{df}{dx}$, and FY contains the derivatives with

respect to *y*, that is, $\dfrac{df}{dy}$.

[fx, fy] = differentiate(FO, [x, y]), where X and Y are column vectors, allows you to specify the evaluation points as a single argument.

[fx, fy, fxx, fxy, fyy] = differentiate(FO, ...) computes the first and second derivatives of the surface fit object FO.

differentiate

fxx contains the second derivatives with respect to x, that is, $\dfrac{\partial^2 f}{\partial x^2}$.

fxy contains the mixed second derivatives, that is, $\dfrac{\partial^2 f}{\partial x \partial y}$.

fyy contains the second derivatives with respect to y, that is, $\dfrac{\partial^2 f}{\partial y^2}$.

Tips

For library models with closed forms, the toolbox calculates derivatives analytically. For all other models, the toolbox calculates the first derivative using the centered difference quotient

$$\frac{df}{dx} = \frac{f(x + \Delta x) - f(x - \Delta x)}{2\Delta x}$$

where x is the value at which the toolbox calculates the derivative, Δx is a small number (on the order of the cube root of eps), $f(x + \Delta x)$ is fun evaluated at $x + \Delta x$, and $f(x - x\Delta)$ is fun evaluated at $x - \Delta x$.

The toolbox calculates the second derivative using the expression

$$\frac{d^2 f}{dx^2} = \frac{f(x + \Delta x) + f(x - \Delta x) - 2f(x)}{(\Delta x)^2}$$

The toolbox calculates the mixed derivative for surfaces using the expression

$$\frac{\partial^2 f}{\partial x \partial y}(x, y) = \frac{f(x + \Delta x, y + \Delta y) - f(x - \Delta x, y + \Delta y) - f(x + \Delta x, y - \Delta y) + f(x - \Delta x, y - \Delta y)}{4\Delta x \Delta y}$$

Examples

For Curves

Create a baseline sinusoidal signal:

```
xdata = (0:.1:2*pi)';
y0 = sin(xdata);
```

Add noise to the signal:

```
noise = 2*y0.*randn(size(y0));  % Response-dependent
                 % Gaussian noise
ydata = y0 + noise;
```

Fit the noisy data with a custom sinusoidal model:

```
f = fittype('a*sin(b*x)');
fit1 = fit(xdata,ydata,f,'StartPoint',[1 1]);
```

Find the derivatives of the fit at the predictors:

```
[d1,d2] = differentiate(fit1,xdata);
```

Plot the data, the fit, and the derivatives:

```
subplot(3,1,1)
plot(fit1,xdata,ydata) % cfit plot method
subplot(3,1,2)
plot(xdata,d1,'m') % double plot method
grid on
legend('1st derivative')
subplot(3,1,3)
plot(xdata,d2,'c') % double plot method
grid on
legend('2nd derivative')
```

differentiate
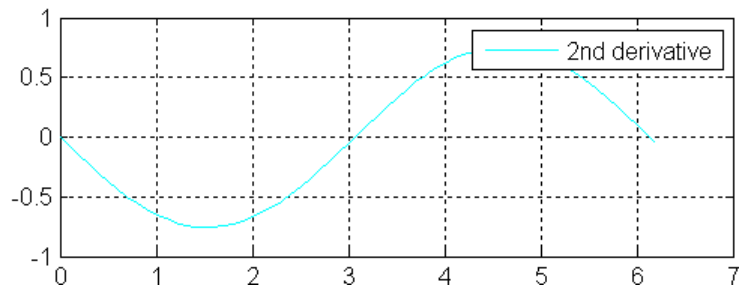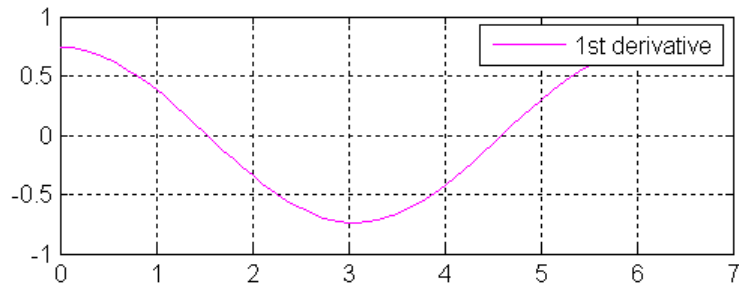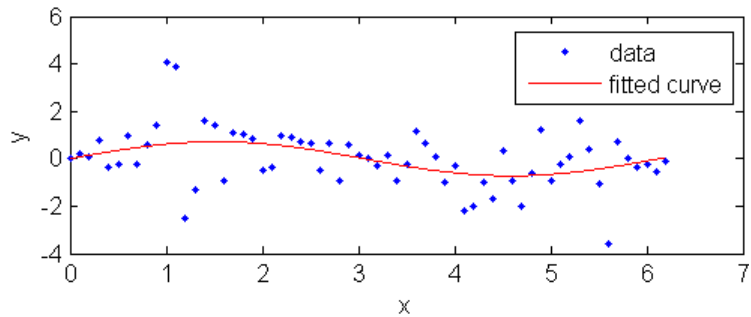
You can also compute and plot derivatives directly with the `cfit` `plot` method, as follows:

```
plot(fit1,xdata,ydata,{'fit','deriv1','deriv2'})
```

The `plot` method, however, does not return data on the derivatives, unlike the `differentiate` method.

For Surfaces

You can use the `differentiate` method to compute the gradients of a fit and then use the `quiver` function to plot these gradients as arrows. The following example plots the gradients over the top of a contour plot.

```
x = [0.64;0.95;0.21;0.71;0.24;0.12;0.61;0.45;0.46;...
0.66;0.77;0.35;0.66];
y = [0.42;0.84;0.83;0.26;0.61;0.58;0.54;0.87;0.26;...
0.32;0.12;0.94;0.65];
z = [0.49;0.051;0.27;0.59;0.35;0.41;0.3;0.084;0.6;...
0.58;0.37;0.19;0.19];
fo = fit( [x, y], z, 'poly32', 'normalize', 'on' );
[xx, yy] = meshgrid( 0:0.04:1, 0:0.05:1 );

[fx, fy] = differentiate( fo, xx, yy );

plot( fo, 'Style', 'Contour' );
hold on
h = quiver( xx, yy, fx, fy, 'r', 'LineWidth', 2 );
hold off
colormap( copper )
```
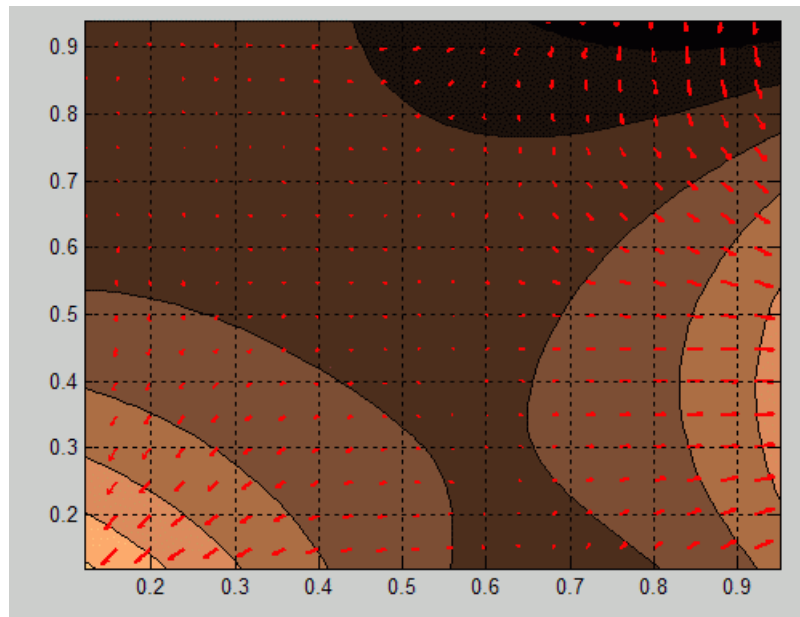
If you want to use derivatives in an optimization, you can, for example, implement an objective function for fmincon as follows.

```
function [z, g, H] = objectiveWithHessian( xy )
        % The input xy represents a single evaluation point
        z = f( xy );
        if nargout > 1
            [fx, fy, fxx, fxy, fyy] = differentiate( f, xy );
            g = [fx, fy];
            H = [fxx, fxy; fxy, fyy];
        end
    end
```

See Also   fit | plot | integrate

Purpose   Exclude data from fit

Syntax   outliers = excludedata(xdata,ydata,*MethodName*,MethodValue)

Description   outliers = excludedata(xdata,ydata,*MethodName*,MethodValue)
                  identifies data to be excluded from a fit using the specified *MethodName*
                  and MethodValue. outliers is a logical vector, with 1 marking
                  predictors (xdata) to exclude and 0 marking predictors to include.
                  Supported *MethodName* and MethodValue pairs are given in the table
                  below.

MethodName	MethodValue
'box'	A four-element vector specifying the edges of a closed box in the *xy*-plane, outside of which data is to be excluded from a fit. The vector has the form [xmin xmax ymin ymax].
'domain'	A two-element vector specifying the endpoints of a closed interval on the *x*-axis, outside of which data is to be excluded from a fit. The vector has the form [xmin xmax].
'indices'	A vector of indices specifying the data points to be excluded.
'range'	A two-element vector specifying the endpoints of a closed interval on the *y*-axis, outside of which data is to be excluded from a fit. The vector has the form [ymin ymax].

Tips   You can combine data exclusion rules using logical operators. For
                  example, to exclude data *inside* the box [-1 1 -1 1] or *outside* the
                  domain [-2 2], use:

```
outliers1 = excludedata(xdata,ydata,'box',[-1 1 -1 1]);
outliers2 = excludedata(xdata,ydata,'domain',[-2 2]);
outliers = ~outliers1|outliers2;
```

You can visualize the combined exclusion rule using random data:

```
xdata = -3 + 6*rand(1,1e4);
ydata = -3 + 6*rand(1,1e4);
plot(xdata(~outliers),ydata(~outliers),'.')
axis ([-3 3 -3 3])
axis square
```
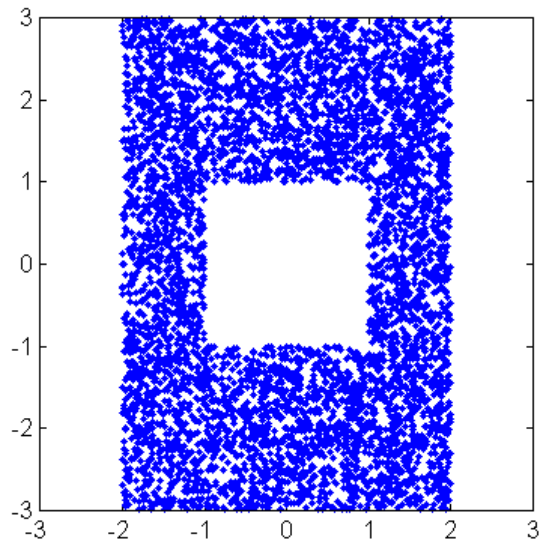

Examples   Load the vote counts and county names for the state of Florida from the 2000 U.S. presidential election:

```
load flvote2k
```

Use the vote counts for the two major party candidates, Bush and Gore, as predictors for the vote counts for third-party candidate Buchanan, and plot the scatters:

```
plot(bush,buchanan,'rs')
hold on
plot(gore,buchanan,'bo')
```

```
legend('Bush data','Gore data')
```
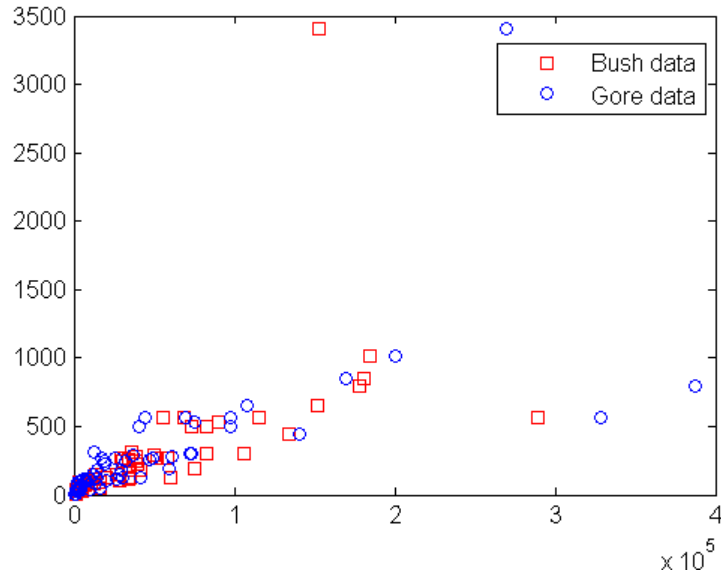

Assume a model where a fixed proportion of Bush or Gore voters choose to vote for Buchanan:

```
f = fittype({'x'})
f =
     Linear model:
       f(a,x) = a*x
```

Exclude the data from absentee voters, who did not use the controversial "butterfly" ballot:

```
absentee = find(strcmp(counties,'Absentee Ballots'));
nobutterfly = excludedata(bush,buchanan,...
                          'indices',absentee);
```

Perform a bisquare weights robust fit of the model to the two data sets, excluding absentee voters:

```
bushfit = fit(bush,buchanan,f,...
              'Exclude',nobutterfly,'Robust','on');
gorefit = fit(gore,buchanan,f,...
              'Exclude',nobutterfly,'Robust','on');
```

Robust fits give outliers a low weight, so large residuals from a robust fit can be used to identify the outliers:

```
figure
plot(bushfit,bush,buchanan,'rs','residuals')
hold on
plot(gorefit,gore,buchanan,'bo','residuals')
```
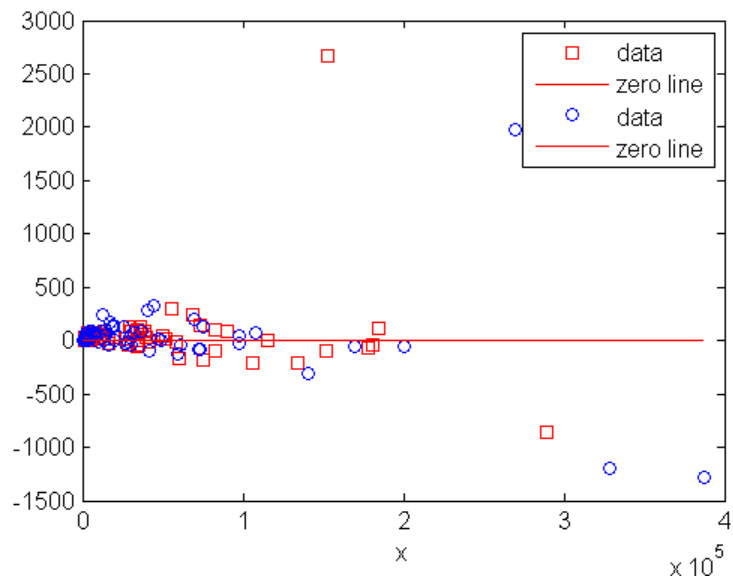

The residuals in the plot above can be computed as follows:

```
bushres = buchanan - feval(bushfit,bush);
goreres = buchanan - feval(gorefit,gore);
```

Large residuals can be identified as those outside the range [-500 500]:

```
bushoutliers = excludedata(bush,bushres,...
                           'range',[-500 500]);
goreoutliers = excludedata(gore,goreres,...
                           'range',[-500 500]);
```

The outliers for the two data sets correspond to the following counties:

```
counties(bushoutliers)
ans =
    'Miami-Dade'
    'Palm Beach'

counties(goreoutliers)
ans =
    'Broward'
    'Miami-Dade'
    'Palm Beach'
```

Miami-Dade and Broward counties correspond to the largest predictor values. Palm Beach county, the only county in the state to use the "butterfly" ballot, corresponds to the largest residual values.

See Also   fit | fitoptions

feval

Purpose   Evaluate `cfit`, `sfit`, or `fittype` object

Syntax
```
y = feval(cfun,x)
z = feval(sfun,[x,y])
z = feval(sfun,x,y)
y = feval(ffun,coeff1,coeff2,...,x)
z = feval(ffun,coeff1,coeff2,...,x,y)
```

Tips   You can also treat fit objects as functions and call `feval` indirectly
                  using the following syntax:

```
y = cfun(x)          % cfit objects;
z = sfun(x,y)        % sfit objects
z = sfun([x, y])     % sfit objects
y = ffun(coef1,coef2,...,x)    % curve fittype objects;
z = ffun(coef1,coef2,...,x,y) % surface fittype objects;
```

This simpler syntax is recommended to evaluate these objects, instead
of calling `feval` directly.

Description   Use the `feval` method to evaluate the estimated function, either at your
                  original data points, or at new locations. The latter is often referred to
                  as interpolation or prediction, depending on the type of model. You can
                  also use `feval` to extrapolate the estimated function's value at new
                  locations that are not within the range of the original data.

`y = feval(cfun,x)` evaluates the `cfit` object `cfun` at the predictor
values in the column vector `x` and returns the response values in the
column vector `y`.

`z = feval(sfun,[x,y])` evaluates the `sfit` object `sfun` at the predictor
values in the two column matrix `[x,y]` and returns the response values
in the column vector `z`.

`z = feval(sfun,x,y)` evaluates the `sfit` object `sfun` at the predictor
values in the matrices `x` and `y` that must be the same size. It
returns the response values in the matrix `z` that will be the same size
as `x` and `y`.

y = feval(ffun,coeff1,coeff2,...,x) assigns the coefficients coeff1, coeff2, etc. to the fittype object ffun, evaluates it at the predictor values in the column vector x, and returns the response values in the column vector y. ffun cannot be a cfit object in this syntax. To evaluate cfit objects, use the first syntax.

z = feval(ffun,coeff1,coeff2,...,x,y) achieves a similar result for a fittype object for a surface.

Examples

```
f = fittype('a*x^2+b*exp(n*x)');
c = cfit(f,1,10.3,-1e2);
X = rand(2)
X =
    0.0579    0.8132
    0.3529    0.0099

y1 = feval(f,1,10.3,-1e2,X)
y1 =
    0.0349    0.6612
    0.1245    3.8422
y1 = f(1,10.3,-1e2,X)
y1 =
    0.0349    0.6612
    0.1245    3.8422

y2 = feval(c,X)
y2 =
    0.0349
    0.1245
    0.6612
    3.8422
y2 = c(X)
y2 =
    0.0349
    0.1245
    0.6612
    3.8422
```

feval

See Also   fit | fittype | cfit

Purpose   Fit curve or surface to data

Syntax
```
fitobject = fit(x,y,fitType)
fitobject = fit([x,y],z, fitType)
fitobject = fit(..., Name, Value,...)
fitobject = fit(x,y,libname,options)
fitobject = fit(...,'problem',vals)
fitobject = fit(x,y,fitType,...,'Weight', Weights)
[fitobject,gof] = fit(...)
[cfun,gof,output] = fit(...)
```

Description   `fitobject = fit(x,y,fitType)` fits the data in `x` and `y` with the library model, anonymous function or `fittype` object specified by `fitType`.

- `x` must be a matrix with either one (curve fitting) or two (surface fitting) columns. For surface fitting, if your data is in separate vectors, then you can use the syntax: `fitobject = fit([x,y],z, fitType)`.

- `y` must be a column vector with the same number of rows as `x`.

- `x` and `y` cannot contain `Inf` or `NaN`. Only the real parts of complex data are used in the fit.

- `fitType` can be a string, anonymous function, or a `fittype` object specifying the model to fit. If a string, you can specify library model names. String choices include:

```
LIBNAME             DESCRIPTION
'poly1'             Linear polynomial curve
'poly11'            Linear polynomial surface
'poly2'             Quadratic polynomial curve
'linearinterp'      Piecewise linear interpolation
'cubicinterp'       Piecewise cubic interpolation
'smoothingspline'   Smoothing spline (curve)
'lowess'            Local linear regression (surface)
```

or any of the names of library models described in "List of Library Models for Curve and Surface Fitting" on page 4-13.

To fit custom models, use an anonymous function or create a `fittype` with the `fittype` function and use this as the `fitType` argument.

- `fitobject` is the fit result, a `cfit` (for curves) or `sfit` (for surfaces) object. See "Fit Postprocessing" for functions for plotting, evaluating, calculating confidence intervals, integrating, differentiating, or modifying your fit object.

`fitobject = fit(..., Name, Value,...)` fits the data using the problem and algorithm options specified in the name-value pair arguments. You can display the supported property names and default values for specific library models with the `fitoptions` function. For example:

```
fitoptions( 'cubicinterp' )
fitoptions( 'poly1' )
```

`fitobject = fit(x,y,`*`libname`*`,options)` fits the data using the algorithm options specified by the `fitoptions` object `options`. This is an alternative syntax to specifying the property-value pairs. For help on constructing `options`, see the `fitoptions` function.

`fitobject = fit(...,'problem',vals)` assigns `vals` to the problem-dependent constants. `vals` is a cell array with one element per problem dependent constant. See `fittype` for more information on problem dependent constants.

`fitobject = fit(x,y,fitType,...,'Weight', Weights)` creates a weighted fit using the given `Weights`. `Weights` must be a vector the same size as `y`.

`[fitobject,gof] = fit(...)` returns goodness-of-fit statistics to the structure `gof`. The `gof` structure includes the fields shown in the table below.

Field	Value
sse	Sum of squares due to error
R2	Coefficient of determination
adjustedR2	Degree-of-freedom adjusted coefficient of determination
stdError	Root mean squared error (standard error)

`[cfun,gof,output] = fit(...)` returns the structure `output`, which contains information associated with the fitting algorithm. Fields depend on the algorithm. For example, the `output` structure for nonlinear least-squares algorithms has the fields shown in the table below.

Field	Value
numobs	Number of observations (response values)
numparam	Number of unknown parameters (coefficients) to fit
residuals	Vector of residuals
Jacobian	Jacobian matrix
exitflag	Describes the exit condition of the algorithm. Positive flags indicate convergence, within tolerances. Zero flags indicate that the maximum number of function evaluations or iterations was exceeded. Negative flags indicate that the algorithm did not converge to a solution.
iterations	Number of iterations
funcCount	Number of function evaluations
firstorderopt	Measure of first-order optimality (absolute maximum of gradient components)
algorithm	Fitting algorithm employed

fit

Remarks on Starting Points	For rational and Weibull models, and all custom nonlinear models, the toolbox selects default initial values for coefficients uniformly at random from the interval (0,1).
	As a result, multiple fits using the same data and model may lead to different fitted coefficients. To avoid this, specify initial values for coefficients with `fitoptions` structure or a vector value for the `StartPoint` property.
Examples	Fit a cubic interpolating spline through x and y:

```
[curve, goodness] = fit( x, y, 'pchipinterp' );
```

Fit a polynomial surface of degree 2 in x and degree 3 in y using the least absolute residual robust (LAR) method:

```
sf = fit( [x, y], z, 'poly23', 'Robust', 'LAR' );
```

Fit the 1st equation in the curve fitting library of exponential models (a single-term exponential), overriding the starting point to be p0:

```
curve = fit( x, y, 'exp1', 'StartPoint', p0 );
```

Load data and fit using an anonymous function:

1 Load data and set Emax to 1 before defining your anonymous function:

```
data = importdata( 'OpioidHypnoticSynergy.txt' );
Propofol     = data.data(:,1);
Remifentanil = data.data(:,2);
Algometry    = data.data(:,3);
Emax = 1;
```

2 Define the model equation as an anonymous function:

```
Effect = @(IC50A, IC50B, alpha, n, x, y) ...
    Emax*( x/IC50A + y/IC50B + alpha*( x/IC50A )...
    .* ( y/IC50B ) ).^n ./(( x/IC50A + y/IC50B + ...
    alpha*( x/IC50A ) .* ( y/IC50B ) ).^n  + 1);
```

3 Use the anonymous function `Effect` as an input to the `fit` function, and plot the results:
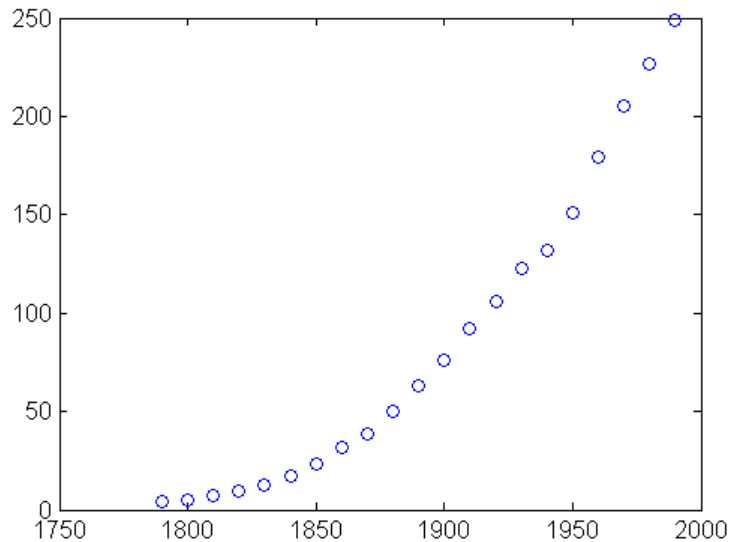
```
AlgometryEffect = fit( [Propofol, Remifentanil], Algometry, Effect,
    'StartPoint', [2, 10, 1, 0.8], ...
    'Lower', [-Inf, -Inf, -5, -Inf], ...
    'Robust', 'LAR' )
plot( AlgometryEffect, [Propofol, Remifentanil], Algometry )
```

See "Biopharmaceutical Drug Interaction Surface Fitting" on page 5-24 for more information on this example.

Load and plot data, create fit options and fit type, then create and plot fit:

1 Load and plot the data in `census.mat`:

```
load census
plot(cdate,pop,'o')
hold on
```

2 Create a fit options structure and a `fittype` object for the custom nonlinear model $y = a(x–b)^n$, where a and b are coefficients and n is a problem-dependent parameter:

```
s = fitoptions('Method','NonlinearLeastSquares',...
               'Lower',[0,0],...
               'Upper',[Inf,max(cdate)],...
               'Startpoint',[1 1]);
f = fittype('a*(x-b)^n','problem','n','options',s);
```

3 Fit the data using the fit options and a value of $n = 2$:

```
[c2,gof2] = fit(cdate,pop,f,'problem',2)
c2 =
     General model:
       c2(x) = a*(x-b)^n
     Coefficients (with 95% confidence bounds):
       a =    0.006092  (0.005743, 0.006441)
       b =        1789  (1784, 1793)
     Problem parameters:
```

```
        n =                 2
 gof2 =
            sse: 246.1543
        rsquare: 0.9980
            dfe: 19
     adjrsquare: 0.9979
           rmse: 3.5994
```

4 Fit the data using the fit options and a value of *n* = 3:

```
 [c3,gof3] = fit(cdate,pop,f,'problem',3)
 c3 =
     General model:
       c3(x) = a*(x-b)^n
     Coefficients (with 95% confidence bounds):
       a =  1.359e-005  (1.245e-005, 1.474e-005)
       b =         1725  (1718, 1731)
     Problem parameters:
       n =             3
 gof3 =
            sse: 232.0058
        rsquare: 0.9981
            dfe: 19
     adjrsquare: 0.9980
           rmse: 3.4944
```

5 Plot the fit results with the data:

```
 plot(c2,'m')
 plot(c3,'c')
```
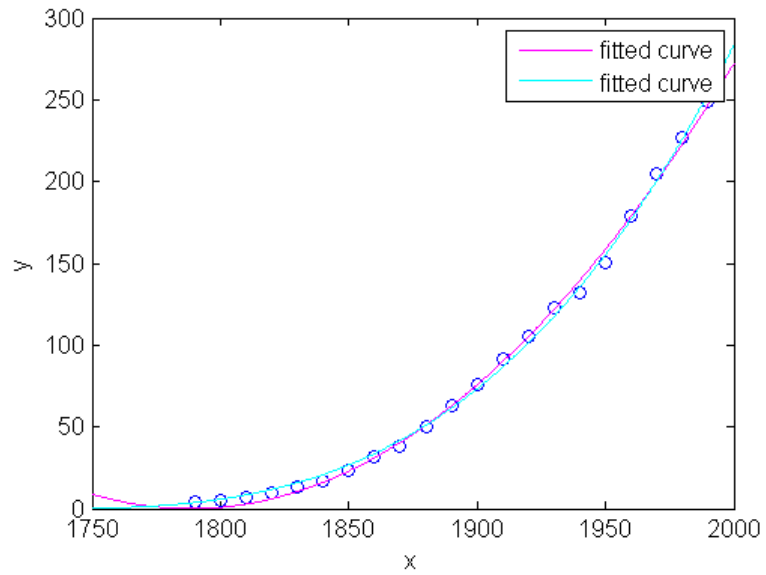
See Also   `fittype` | `fitoptions` | `feval` | `plot` | `confint`

How To   • "Fit Postprocessing"

              • "List of Library Models for Curve and Surface Fitting" on page 4-13

Purpose   Create or modify fit options structure

Syntax
```
options = fitoptions
options = fitoptions(model)
options = fitoptions(model,fld1,val1,fld2,val2,...)
options = fitoptions('Method',method)
options =
fitoptions('Method',method,fld1,val1,fld2,val2,...)
newoptions = fitoptions(options,fld1,val1,fld2,val2,...)
newoptions = fitoptions(options1,options2)
```

Description   `options = fitoptions` creates the default fit options structure
`options`. Properties in the options structure, listed in the table below
with their default values, are supported by all fitting methods.

Property Name	Values
Normalize	Specifies whether the data is centered and scaled. Values are `'off'` or `'on'`. The default is `'off'`.
Exclude	A logical vector indicating data points to exclude from the fit. The `excludedata` function can be used to create this vector. The default is empty.
Weights	A vector of weights the same size as the response data. The default is empty.

fitoptions

Property Name	Values	
Method	The fitting method. A complete list of supported fitting methods is given below. The default is `'None'`.	
	`'NearestInterpolant'`	Nearest neighbor interpolation
	`'LinearInterpolant'`	Linear interpolation
	`'PchipInterpolant'`	Piecewise cubic Hermite interpolation (curves only)
	`'CubicSplineInterpolant'`	Cubic spline interpolation
	`'BiharmonicInterpolant'`	Biharmonic surface interpolation
	`'SmoothingSpline'`	Smoothing spline
	`'LowessFit'`	Lowess smoothing (surfaces only)
	`'LinearLeastSquares'`	Linear least squares
	`'NonlinearLeastSquares'`	Nonlinear least squares

`options = fitoptions(model)` creates the default fit options structure for the library or custom model specified by the string `model`. To find library model names, see "List of Library Models for Curve and Surface Fitting" on page 4-13.

`options = fitoptions(model,`*fld1*`,val1,`*fld2*`,val2,...)` creates a fit options structure for the specified `model` with the properties specified by the strings *fld1*, *fld2*, ... set to the values val1, val2, ..., respectively.

`options = fitoptions('Method',`*method*`)` creates the default fit options structure for the fitting method specified by the string *method*. Supported fitting methods are listed in the table above.

`options = fitoptions('Method',`*method*`,`*fld1*`,val1,`*fld2*`,val2,...)` creates

the default fit options structure for the fitting method specified by the string *method* with the properties specified by the strings *fld1*, *fld2*, ... set to the values val1, val2, ..., respectively.

newoptions = fitoptions(options,*fld1*,val1,*fld2*,val2,...) modifies the existing fit options structure options by setting the properties specified by the strings *fld1*, *fld2*, ... set to the values val1, val2, ..., respectively. The new options structure is returned in newoptions.

newoptions = fitoptions(options1,options2) combines the input fit options structures options1 and options2 to create the output fit options structure newoptions. If the input structures have Method properties set to the same value, the nonempty values for the properties in options2 override the corresponding values in options1 in the output structure. If the input structures have Method properties set to different values, the output structure will have the same Method as options1, and only the values of the Normalize, Exclude, and Weights properties of options2 will override the corresponding values in options1.

Tips   Property values in a fit options structure can be referenced with the get method and assigned with the set method. For example:

```
options = fitoptions('fourier1');
get(options,'Method')
ans =
NonlinearLeastSquares
get(options,'MaxIter')
ans =
   400
set(options,'Maxiter',1e3);
get(options,'MaxIter')
ans =
        1000
```

Property values can also be referenced and assigned using the dot notation. For example:

```
options.MaxIter
ans =
        1000
options.MaxIter = 500;
options.MaxIter
ans =
   500
```

Additional Fit Options

Some fitting methods have additional properties in the fit options structure, beyond the default properties Normalize, Exclude, Weights, and Method.

There are *no* additional parameters if Method is:

- 'NearestInterpolant'
- 'LinearInterpolant'
- 'PchipInterpolant'
- 'CubicSplineInterpolant'
- 'BiharmonicInterpolant'

If the Method is SmoothingSpline, the SmoothingParam property is available to configure the smoothing parameter. Its value must be between 0 and 1. The default value depends on the data set.

If the Method is LowessFit, then the Span property is available to configure the proportion of data points to be used in local regressions. It must be a scalar in [0,1], and the default is 0.25.

If the Method property has the value LinearLeastSquares, the additional properties available in the fit options structure are listed in the table below.

Additional Properties with LinearLeastSquares	Description
Robust	Specifies the robust linear least-squares fitting method to be used. Values are `'on'`, `'off'`, `'LAR'`, or `'Bisquare'`. The default is `'off'`. `'LAR'` specifies the least absolute residual method and `'Bisquare'` specifies the bisquare weights method. `'on'` is equivalent to `'Bisquare'`, the default method.
Lower	A vector of lower bounds on the coefficients to be fitted. The default value is an empty vector, indicating that the fit is unconstrained by lower bounds. If bounds are specified, the vector length must equal the number of coefficients. Individual unconstrained lower bounds can be specified by `-Inf`.
Upper	A vector of upper bounds on the coefficients to be fitted. The default value is an empty vector, indicating that the fit is unconstrained by upper bounds. If bounds are specified, the vector length must equal the number of coefficients. Individual unconstrained upper bounds can be specified by `Inf`.

If the `Method` property has the value `NonlinearLeastSquares`, the additional properties available in the fit options structure are listed in the table below.

Additional Properties with NonlinearLeastSquares	Description
Robust	Specifies the robust linear least-squares fitting method to be used. Values are `'on'`, `'off'`, `'LAR'`, or `'Bisquare'`. The default is `'off'`. `'LAR'` specifies the least absolute residual method and `'Bisquare'` specifies the bisquare weights method. `'on'` is equivalent to `'Bisquare'`, the default method.
Lower	A vector of lower bounds on the coefficients to be fitted. The default value is an empty vector, indicating that the fit is unconstrained by lower bounds. If bounds are specified, the vector length must equal the number of coefficients. Individual unconstrained lower bounds can be specified by `-Inf`.
Upper	A vector of upper bounds on the coefficients to be fitted. The default value is an empty vector, indicating that the fit is unconstrained by upper bounds. If bounds are specified, the vector length must equal the number of coefficients. Individual unconstrained upper bounds can be specified by `Inf`.
StartPoint	A vector of initial values for the coefficients. The default value of `StartPoint` is an empty vector. If the default value is passed to the `fit` function, starting points for some library models are determined heuristically. For other models, the values are selected uniformly at random on the interval (0,1).

Additional Properties with NonlinearLeastSquares	Description
Algorithm	The algorithm used for the fitting procedure. Values are 'Levenberg-Marquardt', 'Gauss-Newton', or 'Trust-Region'. The default is 'Trust-Region'.
DiffMaxChange	The maximum change in coefficients for finite difference gradients. The default is 0.1.
DiffMinChange	The minimum change in coefficients for finite difference gradients. The default is 10^{-8}.
Display	Controls the display in the command window. 'notify', the default, displays output only if the fit does not converge. 'final' displays only the final output. 'iter' displays output at each iteration. 'off' displays no output.
MaxFunEvals	The maximum number of evaluations of the model allowed. The default is 600.
MaxIter	The maximum number of iterations allowed for the fit. The default is 400.
TolFun	The termination tolerance on the model value. The default is 10^{-6}.
TolX	The termination tolerance on the coefficient values. The default is 10^{-6}.

fitoptions

Note For the properties `Upper`, `Lower`, and `StartPoint`, the order of the entries in the vector value is the order of the coefficients returned by the `coeffnames` method. For example, if

```
f = fittype('b*x^2+c*x+a');
coeffnames(f)
ans =
    'a'
    'b'
    'c'
```

then setting

```
options.StartPoint = [1 3 5];
```

assigns initial values to the coefficients as follows: `a = 1`, `b = 3`, `c = 5`. Note that this is not the order of the coefficients in the expression used to create `f` with `fittype`.

Examples   Create the default fit options structure and set the option to center and scale the data before fitting:

```
options = fitoptions;
options.Normal = 'on';
options
options =
    Normalize: 'on'
    Exclude: [1x0 double]
    Weights: [1x0 double]
    Method: 'None'
```

Modifying the default fit options structure is useful when you want to set the `Normalize`, `Exclude`, or `Weights` properties, and then fit your data using the same options with different fitting methods. For example:

```
load census
f1 = fit(cdate,pop,'poly3',options);
f2 = fit(cdate,pop,'exp1',options);
f3 = fit(cdate,pop,'cubicsp',options);
```

Data-dependent fit options are returned in the third output argument of the fit function. For example:

```
[f,gof,out] = fit(cdate,pop,'smooth');
smoothparam = out.p
smoothparam =
    0.0089
```

The default smoothing parameter can be modified for a new fit:

```
options = fitoptions('Method','Smooth',...
                     'SmoothingParam',0.0098);
[f,gof,out] = fit(cdate,pop,'smooth',options);
```

Create a noisy sum of two Gaussian peaks—one with a small width, and one with a large width:

```
a1 = 1; b1 = -1; c1 = 0.05;
a2 = 1; b2 = 1; c2 = 50;
x = (-10:0.02:10)';
gdata = a1*exp(-((x-b1)/c1).^2) + ...
        a2*exp(-((x-b2)/c2).^2) + ...
        0.1*(rand(size(x))-.5);
plot(x,gdata)
```
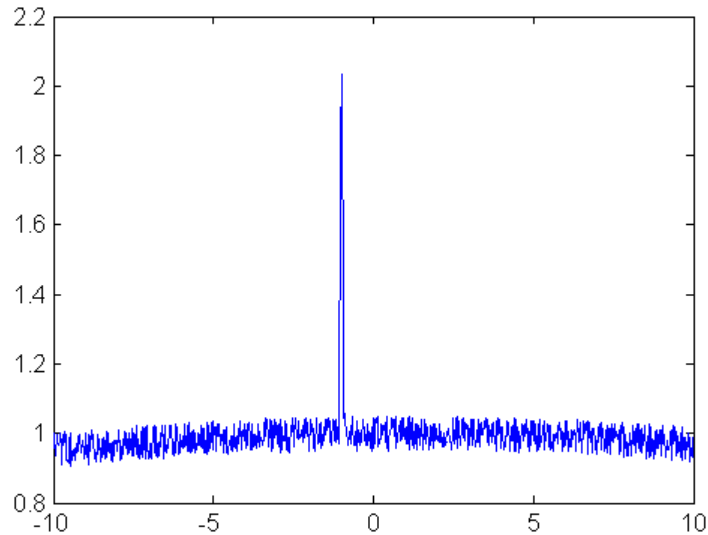
Fit the data using the two-term Gaussian library model:

```
gfit = fit(x,gdata,'gauss2')
gfit =
     General model Gauss2:
       gfit(x) =  a1*exp(-((x-b1)/c1)^2) +
                  a2*exp(-((x-b2)/c2)^2)
     Coefficients (with 95% confidence bounds):
       a1 =    -0.05388  (-0.136, 0.02826)
       b1 =      -2.651  (-2.718, -2.584)
       c1 =     0.05373  (-0.04106, 0.1485)
       a2 =       1.012  (1.006, 1.018)
       b2 =      0.6703  (0.06681, 1.274)
       c2 =        41.2  (36.54, 45.85)
```

The algorithm is having difficulty, as indicated by the wide confidence intervals for some of the coefficients. To help the algorithm, we could specify lower bounds for the nonnegative amplitudes a1, a2 and widths c1, c2:

```
options = fitoptions('gauss2');
options.Lower = [0 -Inf 0 0 -Inf 0];
```

Recompute the fit with the bound constraints on the coefficients:

```
gfit = fit(x,gdata,'gauss2',options)
gfit =
     General model Gauss2:
       gfit(x) =  a1*exp(-((x-b1)/c1)^2) +
                  a2*exp(-((x-b2)/c2)^2)
     Coefficients (with 95% confidence bounds):
       a1 =        1.003  (0.9641, 1.042)
       b1 =           -1  (-1.002, -0.9987)
       c1 =      0.04972  (0.04748, 0.05197)
       a2 =        1.002  (0.999, 1.004)
       b2 =        1.136  (0.725, 1.547)
       c2 =        48.89  (45.32, 52.47)
```

This is a much better fit. The fit can be further improved by assigning reasonable values to other properties in the fit options structure.

See Also   fit | get | set | setoptions

How To   • "Specifying Fit Options and Optimized Starting Points" on page 4-7

• "Fit Postprocessing"

• "List of Library Models for Curve and Surface Fitting" on page 4-13

fittype

Purpose   Fit type for curve and surface fitting

Syntax
```
ffun = fittype(libname)
ffun = fittype(expr)
ffun = fittype({expr1,...,exprn})
ffun = fittype(expr, Name, Value,...)
ffun = fittype({expr1,...,exprn}, Name, Value,...)
```

Description   `ffun = fittype(libname)` constructs the `fittype` object `ffun` for the library model specified by `libname`. See "Using Library Models" on page 13-74.

`ffun = fittype(expr)` constructs `ffun` for the custom model specified by the expression `expr`. The expression can be a string, cell array, or anonymous function. See "Using Custom Models" on page 13-75.

`ffun = fittype({expr1,...,exprn})` constructs a custom linear model with terms specified by the cell array of expressions in the strings `expr1`, `expr2`, ... , `exprn`. See "Using Linear Models" on page 13-76.

`ffun = fittype(expr, Name, Value,...)` or `ffun = fittype({expr1,...,exprn}, Name, Value,...)` constructs the `fittype` using specified name-value pair arguments specifying values other than the default values. For supported names-value pair arguments, see "Input Arguments" on page 13-78.

Using Library Models

`ffun = fittype(libname)` constructs the `fittype` for the library model `libname`. Choices for `libname` include any of the names of library models described in "List of Library Models for Curve and Surface Fitting" on page 4-13. The following table shows some common examples.

libname	Description
'poly1'	Linear polynomial curve
'poly11'	Linear polynomial surface

libname	Description
`'poly2'`	Quadratic polynomial curve
`'linearinterp'`	Piecewise linear interpolation
`'cubicinterp'`	Piecewise cubic interpolation
`'smoothingspline`	Smoothing spline (curve)
`'lowess'`	Local linear regression (surface)

Using Custom Models

`ffun = fittype(expr)` constructs a custom model `fittype` for the MATLAB expression contained in the string, cell array, or anonymous function `expr`.

The `fittype` automatically determines input arguments by searching `expr` for variable names. The `fittype` assumes `x` is the independent variable, `y` is the dependent variable, and all other variables are coefficients of the model. `x` is used if no variable exists. See "Dependent and Independent Variables" on page 13-77.

All coefficients must be scalars. You cannot use the following coefficient names in the expression string `expr`: `i`, `j`, `pi`, `inf`, `nan`, and `eps`.

If `expr` is a string or anonymous function, then the toolbox uses a nonlinear fitting algorithm to fit the model to data (see Using Anonymous Functions on page 75). To use a linear fitting algorithm, use a cell array of terms (see "Using Linear Models" on page 13-76).

Using Anonymous Functions

If `expr` is an anonymous function, then the order of inputs must be correct. The input order enables the `fittype` class to determine which inputs are coefficients to estimate, problem-dependent parameters and independent variables. The order of the input arguments to the anonymous function must be:

```
EXPR = @(coefficients, problemparameters, x, y) expression
```

fittype

You need at least one coefficient. The problem parameters and y are optional. The last arguments, x and y, represent the independent variables: just x for curves, but x and y for surfaces. If you don't want to use x and/or y as the names of the independent variables, then specify different names using the 'independent' argument name-value pair. However, whatever name or names you choose, these arguments must be the last arguments to the anonymous function.

Anonymous functions make it easier to pass other data into the fittype and fit functions. For example, to create a fittype using an anonymous function and a variable value (c) from the workspace:

```
c = 1;
g = fittype( @(a, b, x) a*x.^2+b*x+c )
```

The fittype can use the variable values in your workspace at the time you create the fittype. To pass in new data from the workspace, create the fittype again, e.g.,

```
c=5 % Change value of c
g = fittype( @(a, b, x) a*x.^2+b*x+c )
```

In this case the value of c is fixed when you create the fittype. To specify the value of c at the time you call fit, you can use problem parameters. For example, to make a fit with c = 2 and then a new fit with c = 3:

```
g = fittype( @(a,b,x) a*x.^2+b*x+c, 'problem', 'c' )
f1 = fit( xdata, ydata, g, 'problem', 2 )
f2 = fit( xdata, ydata, g, 'problem', 3 )
```

See "Examples" on page 13-79.

Using Linear Models

To use a linear fitting algorithm, specify expr as a cell array of terms, as follows: ffun = fittype({expr1,...,exprn}). Specify the model terms by the expressions in the strings expr1, expr2, ... , exprn. Do

not include coefficients in the expressions for the terms. If there is a constant term, use `'1'` as the corresponding expression in the cell array.

To specify a linear model of the following form:

```
coeff1 * term1 + coeff2 * term2 + coeff3 * term3 + ...
```

(where no coefficient appears within any of `term1`, `term2`, etc.), use a cell array where each term, without coefficients, is specified in a cell of `expr`, as follows:

```
EXPR = {'term1', 'term2', 'term3', ... }
```

For example, the model

```
a*x + b*sin(x) + c
```

is linear in `a`, `b` and `c`. It has three terms `x`, `sin(x)` and `1` (because c=c*1) and so `expr` is

```
EXPR = {'x','sin(x)','1'}
```

Dependent and Independent Variables

To determine what are dependent and independent variables and coefficients, consider this equation:

$$y = f(x) = a + (b * x) + (c * x^2)$$

- y is the dependent variable.
- x is the independent variable.
- a, b, and c are the coefficients.

The `'independent'` variable is what you control. The `'dependent'` variable is what you measure, i.e., it depends on the independent variable. The `'coefficients'` are the parameters that the fitting algorithm estimates.

fittype

For example, if you have census data, then the year is the independent variable because it does not depend on anything. Population is the dependent variable, because its value depends on the year in which the census is taken. If a parameter like growth rate is part of the model, if the fitting algorithm estimates it, then it is one of the 'coefficients'.

See "Examples" on page 13-79 for how to specify an independent variable and coefficient names.

Input Arguments

libname

Library model name. See "List of Library Models for Curve and Surface Fitting" on page 4-13.

expr

Custom model expression. The expression can be a string, cell array, or anonymous function.

You can specify any MATLAB command and therefore any `.m` file.

Name-Value Pair Arguments

Specify optional comma-separated pairs of Name,Value arguments, where Name is the argument name and Value is the corresponding value. Name must appear inside single quotes (' '). You can specify several name and value pair arguments in any order as Name1,Value1,...,NameN,ValueN.

coefficients

The coefficient names. Use a cell array for multiple names. You can use multicharacter symbol names. You cannot use the following names: `i`, `j`, `pi`, `inf`, `nan`, `eps`.

dependent

The dependent (response) variable name.

Default: y

independent

The independent (predictor) variable name.

Default: x

options

The default fit options for the object.

problem

The problem-dependent (fixed) parameter names. Use a cell array for multiple names.

Default: None

Output Arguments

ffun

fittype object. You can use a fittype object as an input to the fit function.

Examples

Construct a fittype object for the rat33 library model:

```
f = fittype('rat33')
f =
   General model Rat33:
   f(p1,p2,p3,p4,q1,q2,q3,x) =
         (p1*x^3 + p2*x^2 + p3*x + p4)/
                  (x^3 + q1*x^2 + q2*x + q3)
```

Construct a fittype object for a custom nonlinear model, designating n as a problem-dependent parameter and u as the independent variable:

```
g = fittype('a*u+b*exp(n*u)',...
            'problem','n',...
            'independent','u')
g =
```

```
General model:
   g(a,b,n,u) = a*u+b*exp(n*u)
```

Construct a `fittype` object for a custom linear model, specifying the coefficient names:

```
h = fittype({'cos(x)','1'},'coefficients',{'a1','a2'})
h =
    Linear model:
      h(a1,a2,x) = a1*cos(x) + a2
```

Fit a curve defined by a file:

1 Define a function in a MATLAB file:

```
function y = piecewiseLine( x, a, b, c, d, k )
% PIECEWISELINE   A line made of two pieces
% that is not continuous.

y = zeros( size( x ) );

% This example includes a for-loop and if statement
% purely for example purposes.
for i = 1:length( x )
    if x(i) < k,
        y(i) = a + b.* x(i);
    else
        y(i) = c + d.* x(i);
    end
end
end
```

2 Define some data, create a `fittype` specifying the function `piecewiseLine`, create a fit using the `fittype`, and plot the results:

```
x = [0.81;0.91;0.13;0.91;0.63;0.098;0.28;0.55;...
0.96;0.96;0.16;0.97;0.96];
y = [0.17;0.12;0.16;0.0035;0.37;0.082;0.34;0.56;...
0.15;-0.046;0.17;-0.091;-0.071];
ft = fittype( 'piecewiseLine( x, a, b, c, d, k )' )
f = fit( x, y, ft, 'StartPoint', [1, 0, 1, 0, 0.5] )
plot( f, x, y )
```

Create a `fittype` using an anonymous function:

```
g = fittype( @(a, b, c, x) a*x.^2+b*x+c )
```

Create a `fittype` for a surface using an anonymous function and specifying independent and dependent parameters:

```
g = fittype( @(a, b, c, d, x, y) a*x.^2+b*x+c*exp...
    ( -(y-d).^2 ), 'independent', {'x', 'y'},...
     'dependent', 'z' );
```

Create a `fittype` for a surface using an anonymous function and specifying independent and dependent parameters, and problem parameters that you will specify later when you call `fit`:

```
g = fittype( @(a,b,c,d,x,y) a*x.^2+b*x+c*exp( -(y-d).^2 ), ...
        'problem', {'c','d'}, 'independent', {'x', 'y'}, ...
        'dependent', 'z' );
```

Use an anonymous function to pass workspace data into the `fittype` and `fit` functions.

1 Create and plot an S-shaped curve. In later steps you will stretch and move this curve to fit to some data.

```
% Breakpoints
xs = (0:0.1:1).';
% Height of curve at breakpoints
ys = [0; 0; 0.04; 0.1; 0.2; 0.5; 0.8; 0.9; 0.96; 1; 1];
% Plot S-shaped curve
xi = linspace( 0, 1, 241 );
plot( xi, interp1( xs, ys, xi, 'pchip' ), 'LineWidth', 2 )
hold on
plot( xs, ys, 'o', 'MarkerFaceColor', 'r' )
hold off
title S-curve
```

2 Create a `fittype` using an anonymous function, taking the values
from the workspace for the curve breakpoints (`xs`) and the height
of the curve at the breakpoints (`ys`). Coefficients are `b` (base) and
`h` (height).

```
ft = fittype( @(b, h, x) interp1( xs, b+h*ys, x, 'pchip' ) )
```

3 Plot the `fittype` specifying example coefficients of base `b=1.1` and
height `h=-0.8`.

```
plot( xi, ft( 1.1, -0.8, xi ), 'LineWidth', 2 )
title 'Fittype with b=1.1 and h=-0.8'
```

4 Load and fit some data, using the `fittype` `ft` created using
workspace values:

```
% Load some data
xdata = [0.012;0.054;0.13;0.16;0.31;0.34;0.47;0.53;0.53;...
   0.57;0.78;0.79;0.93];
ydata = [0.78;0.87;1;1.1;0.96;0.88;0.56;0.5;0.5;0.5;0.63;...
   0.62;0.39];
% Fit the curve to the data
f = fit( xdata, ydata, ft, 'Start', [0, 1] )
% Plot fit
plot( f, xdata, ydata )
title 'Fitted S-curve'
```

The following example shows the differences between using anonymous functions with problem parameters and workspace variable values.

1 Load data, create a `fittype` for a curve using an anonymous function with problem parameters, and call `fit` specifying the problem parameters:

```
% Load some random data:
xdata = [0.098;0.13;0.16;0.28;0.55;0.63;0.81;0.91;0.91;...
    0.96;0.96;0.96;0.97];
ydata = [0.52;0.53;0.53;0.48;0.33;0.36;0.39;0.28;0.28;...
    0.21;0.21;0.21;0.2];

% Create a fittype that has a problem parameter:
g = fittype( @(a,b,c,x) a*x.^2+b*x+c, 'problem', 'c' )

% Examine coefficients. Observe c is not a coefficient.
coeffnames( g )

% Examine arguments. Observe that c is an argument.
argnames( g )

% Call fit and specify the value of c:
f1 = fit( xdata, ydata, g, 'problem', 0, 'start', [1, 2] )

% Note: specify start points in the calls to fit to
% avoid warning messages about random start points
% and to ensure repeatability of results.

% Call fit again and specify a different value of c,
% to get a new fit:
f2 = fit( xdata, ydata, g, 'problem', 1, 'start', [1, 2] )

% Plot results. Note example specified c constants
```

```
% do not make a good fit.
plot( f1, xdata, ydata )
hold on
plot( f2, 'b' )
hold off
```

2 Modify the example above to create the same fits using workspace
values for variables, instead of using problem parameters. Using the
same data, create a fittype for a curve using an anonymous function
with a workspace value for variable c:

```
% Remove c from the argument list:
try
    g = fittype( @(a,b,x) a*x.^2+b*x+c )
catch e
    disp( e.message )
end
% Observe error because now c is undefined.
% Define c and create fittype:
c = 0;
g1 = fittype( @(a,b,x) a*x.^2+b*x+c )

% Call fit (now no need to specify problem parameter)
f1 = fit( xdata, ydata, g1, 'start', [1, 2] )
% Note that this f1 is the same as the f1 above.
% To change the value of c, create the fittype again:
c = 1;
g2 = fittype( @(a,b,x) a*x.^2+b*x+c ) % uses c = 1
f2 = fit( xdata, ydata, g2, 'start', [1, 2] )
% Note that this f2 is the same as the f2 above.
% Plot results
plot( f1, xdata, ydata )
hold on
plot( f2, 'b' )
hold off
```

Here are other examples using `fittype`:

```
g = fittype('a*x^2+b*x+c')
g = fittype('a*x^2+b*x+c','coeff',{'a','b','c'})
g = fittype('a*time^2+b*time+c','indep','time')
g = fittype('a*time^2+b*time+c','indep','time','depen','height')
g = fittype('a*x+n*b','problem','n')
g = fittype({'cos(x)','1'})                          % linear
g = fittype({'cos(x)','1'}, 'coefficients', {'a','b'}) % linear
```

See Also   `fit` | `fitoptions` | `cfit` | `sfit`

How To   • "List of Library Models for Curve and Surface Fitting" on page 4-13

• "Parametric Fitting" on page 4-2

fn2fm

Purpose   Convert to specified form

Syntax   ```
g = fn2fm(f,form)
sp = fn2fm(f,'B-',sconds)
fn2fm(f)
```

**Description**  `g = fn2fm(f,form)` describes the same function as is described by `f`,
                 but in the form specified by the string `form`. Choices for `form` are `'B-'`,
                 `'pp'`, `'BB'`, `'rB'`, `'rp'`, for the B-form, the ppform, the BBform, and the
                 two rational spline forms, respectively.

                 The B-form describes a function as a weighted sum of the B-splines
                 of a given order k for a given knot sequence, and the BBform (or,
                 Bernstein-Bézier form) is the special case when each knot in that
                 sequence appears with maximal multiplicity, k. The ppform describes
                 a function in terms of its local polynomial coefficients. The B-form is
                 good for constructing and/or shaping a function, while the ppform is
                 cheaper to evaluate.

                 Conversion from a polynomial form to the corresponding rational form
                 is possible only if the function in the polynomial form is vector-valued,
                 in which case its last component is designated as the denominator.
                 Converting from a rational form to the corresponding polynomial form
                 simply reverses this process by reinterpreting the denominator of
                 the function in the rational form as an additional component of the
                 piecewise polynomial function.

                 Conversion to or from the stform is not possible at present.

                 If `form` is `'B-'` (and `f` is in ppform), then the actual smoothness of
                 the function in `f` across each of its interior breaks has to be guessed.
                 This is done by looking, for each interior break, for the first derivative
                 whose jump across that break is not *small* compared to the size of that
                 derivative nearby. The default tolerance used in this is `1.e-12`.

                 `sp = fn2fm(f,'B-',sconds)` permits you to supply, as the input
                 argument `sconds`, a tolerance (strictly between 0 and 1) to be used in
                 the conversion from ppform to B-form.

Alternatively, you can input sconds as a vector with integer entries, with at least as many entries as the ppform in f has *interior* breaks. In that case, sconds(i) specifies the number of smoothness conditions to be used across the ith *interior* break. If the function in f is a tensor product, then sconds, if given, must be a cell array.

fn2fm(f) converts a possibly old version of a form into its present version.

**Examples**     sp = fn2fm(spline(x,y),'B-') gives the interpolating cubic spline provided by the MATLAB command spline, but in B-form rather than in ppform.

```
p0 = ppmak([0 1],[3 0 0]);
p1 = fn2fm(fn2fm(fnrfn(p0,[.4 .6]),'B-'),'pp');
```

gives p1 identical to p0 (up to round-off in the coefficients) since the spline has no discontinuity in any derivative across the additional breaks introduced by fnrfn, hence conversion to B-form ignores these additional breaks, and conversion to ppform does not retain any knot multiplicities (like the knot multiplicities introduced, by conversion to B-form, at the endpoints of the spline's basic interval).

Algorithms   For a multivariate (tensor-product) function, univariate algorithms are applied in each variable.

For the conversion from B-form (or BBform) to ppform, the utility command sprpp is used to convert the B-form of all polynomial pieces to their local power form, using repeated knot insertion at the left endpoint.

The conversion from B-form to BBform is accomplished by inserting each knot enough times to increase its multiplicity to the order of the spline.

The conversion from ppform to B-form makes use of the dual functionals discussed in . Without further information, such a conversion has to ascertain the actual smoothness across each interior break of the function in f.

fn2fm

Cautionary Note

When going from B-form to ppform, any jump discontinuity at the first and last knot, `t(1)` or `t(end)`, will be lost since the ppform considers *f* to be defined outside its basic interval by extension of the first, respectively, the last polynomial piece. For example, while `sp=spmak([0 1],1)` gives the characteristic function of the interval [0..1], `pp=fn2fm(spmak([0 1],1),'pp')` is the constant polynomial, $x| \rightarrow 1$.

See Also

ppmak | spmak | rsmak | stmak

Purpose Name and part(s) of form

Syntax

```
[out1,...,outn] = fnbrk(f,part1,...,partm)
fnbrk(f,interval)
fnbrk(pp,j)
fnbrk(f)
```

Description `[out1,...,outn] = fnbrk(f,part1,...,partm)` returns the part(s) of the form in f specified by part1,...,partn (assuming that n<=m). These are the parts used when the form was put together, in spmak or ppmak or rpmak or rsmak or stmak, but also other parts derived from these.

You only need to specify the beginning character(s) of the relevant string.

Regardless of what particular form f is in, parti can be one of the following.

`'form'`	The particular form used
`'variables'`	The dimension of the function's domain
`'dimension'`	The dimension of the function's target
`'coefficients'`	The coefficients in that particular form
`'interval'`	The basic interval of that form

Depending on the form in f, additional parts may be asked for.

If f is in B-form (or BBform or rBform), then additional choices for parti are

`'knots'`	The knot sequence
`'coefficients'`	The B-spline coefficients
`'number'`	The number of coefficients
`'order'`	The polynomial order of the spline

If `f` is in ppform (or rpform), then additional choices for `parti` are

`'breaks'`	The break sequence
`'coefficients'`	The local polynomial coefficients
`'pieces'`	The number of polynomial pieces
`'order'`	The polynomial order of the spline
`'guide'`	The local polynomial coefficients, but in the form needed for PPVALU in PGS

If the function in `f` is multivariate, then the corresponding multivariate parts are returned. This means, e.g., that knots, breaks, and the basic interval, are cell arrays, the coefficient array is, in general, higher than two-dimensional, and order, number and pieces are vectors.

If `f` is in stform, then additional choices for `parti` are

`'centers'`	The centers
`'coefficients'`	The coefficients
`'number'`	Number of coefficients or terms
`'type'`	The particular type

`fnbrk(f,interval)` with `interval` a 1-by-2 matrix `[a b]` with `a<b` does not return a particular part. Rather, it returns a description of the univariate function described by `f` and in the same form but with the basic interval changed, to the interval given. If, instead, `interval` is `[ ]`, `f` is returned unchanged. This is of particular help when the function in `f` is *m*-variate, in which case `interval` must be a cell array with *m* entries, with the *i*th entry specifying the desired interval in the *i*th dimension. If that *i*th entry is `[ ]`, the basic interval in the *i*th dimension is unchanged.

`fnbrk(pp,j)`, with `pp` the ppform of a univariate function and `j` a positive integer, does not return a particular part, but returns the

ppform of the jth polynomial piece of the function in pp. If pp is the ppform of an *m*-variate function, then j must be a cell array of length *m*. In that case, each entry of j must be a positive integer or else an interval, to single out a particular polynomial piece or else to specify the basic interval in that dimension.

fnbrk(f) returns nothing, but a description of the various parts of the form is printed at the command line instead.

Examples   If p1 and p2 contain the B-form of two splines of the same order, with the same knot sequence, and the same target dimension, then

```
p1plusp2 = spmak(fnbrk(p1,'k'),fnbrk(p1,'c')+fnbrk(p2,'c'));
```

provides the (pointwise) sum of those two functions.

If pp contains the ppform of a bivariate spline with at least four polynomial pieces in the first variable, then ppp=fnbrk(pp,{4,[-1 1]}) gives the spline that agrees with the spline in pp on the rectangle [b4 .. b5] x [-1 .. 1] , where b4, b5 are the fourth and fifth entry in the break sequence for the first variable.

See Also   ppmak | rpmak | rsmak | spmak | stmak

fnchg

Purpose	Change part(s) of form

Syntax

`f = fnchg(f,part,value)`

Description

`f = fnchg(f,part,value)` returns the given function description `f` but with the specified `part` changed to the specified `value`.

The string `part` can be (the beginning character(s) of) :

`'dimension'`	The dimension of the function's target
`'interval'`	The basic interval of that form

The specified `value` for `part` is not checked for consistency with the rest of the description in `f` in case the string `part` terminates with the letter `z`.

Examples

`fndir(f,directions)` returns a vector-valued function even when the function described by `f` is ND-valued. You can correct this by using `fnchg` as follows:

```
fdir = fnchg( fndir(f,directions),...
              'dim',[fnbrk(f,'dim'),size(directions,2)] );
```

See Also

`fnbrk`

Purpose   Arithmetic with function(s)

Syntax   ```
fn = fncmb(function,operation)
f = fncmb(function,function)
fncmb(function,matrix,function)
fncmb(function,matrix,function,matrix)
f = fncmb(function,op,function)
```

Description The intent is to make it easy to carry out the standard linear operations of scaling and adding within a spline space without having to deal explicitly with the relevant parts of the function(s) involved.

fn = fncmb(function,operation) returns (a description of) the function obtained by applying to the values of the function in function the operation specified by operation. The nature of the operation depends on whether operation is a *scalar*, a *vector*, a *matrix*, or a *string*, as follows.

Scalar	Multiply the function by that scalar.
Vector	Add that vector to the function's values; this requires the function to be vector-valued.
Matrix	Apply that matrix to the function's coefficients.
String	Apply the function specified by that string to the function's coefficients.

The remaining options only work for *univariate* functions. See Limitations for more information.

f = fncmb(function,function) returns (a description of) the pointwise sum of the two functions. The two functions must be of the same form. This particular case of just two input arguments is not included in the above table since it only works for univariate functions.

fncmb(function,matrix,function) is the same as fncmb(fncmb(function,matrix),function).

fncmb

fncmb(function,matrix,function,matrix) is the same as
fncmb((fncmb(function,matrix),fncmb(function,matrix)).

f = fncmb(function,op,function) returns the ppform of the spline
obtained by the pointwise combining of the two functions, as specified
by the string op. op can be one of the strings '+', '-', '*'. If the
second function is to be a constant, it is sufficient simply to supply
here that constant.

Examples   fncmb(fn,3.5) multiplies (the coefficients of) the function in fn by 3.5.

fncmb(f,3,g,-4) returns the linear combination, with weights 3 and
−4, of the function in f and the function in g.

fncmb(f,3,g) adds 3 times the function in f to the function in g.

If the function f in f happens to be scalar-valued, then
f3=fncmb(f,[1;2;3]) contains the description of the function whose
value at x is the 3-vector ($f(x)$, $2f(x)$, $3f(x)$). Note that, by the convention
throughout this toolbox, the subsequent statement fnval(*f*3, *x*) returns
a 1-*column*-matrix.

If f describes a surface in R^3, i.e., the function in f is 3-vector-valued
bivariate, then f2 = fncmb(f,[1 0 0;0 0 1]) describes the projection
of that surface to the (*x, z*)-plane.

The following commands produce the picture of a ... spirochete?

```
c = rsmak('circle');
fnplt(fncmb(c,diag([1.5,1]))); axis equal, hold on
sc = fncmb(c,.4);
fnplt(fncmb(sc,-[.2;-.5]))
fnplt(fncmb(sc,-[.2,-.5]))
hold off, axis off
```

If t is a knot sequence of length n+k and a is a matrix with n columns,
then fncmb(spmak(t,eye(n)),a) is the same as spmak(t,a).

fncmb(spmak([0:4],1),'+',ppmak([-1 5],[1 -1])) is the
piecewise-polynomial with breaks -1:5 that, on the interval [0 .. 4],

agrees with the function $x \mapsto B(x \mid 0,1,2,3,4) + x$ (but has no active break at 0 or 1, hence differs from this function outside the interval $[0 .. 4]$).

`fncmb(spmak([0:4],1),'-',0)` has the same effect as `fn2fm(spmak([0:4],1),'pp')`.

Assuming that `sp` describes the B-form of a spline of order <k, the output of

```
fn2fm(fncmb(sp,'+',ppmak(fnbrk(sp,'interv'),zeros(1,k))),'B-')
```

describes the B-form of the same spline, but with its order raised to `k`.

Algorithms   The coefficients are extracted (via `fnbrk`) and operated on by the specified matrix or operation (and, possibly, added), then recombined with the rest of the function description (via `ppmak`, `spmak,rpmak,rsmak,stmak`). To be sure, when the function is rational, the matrix is only applied to the coefficients of the numerator. Again, if we are to translate the function values by a given vector and the function is in ppform, then only the coefficients corresponding to constant terms are so translated.

If there are two functions input, then they must be of the same type (see Limitations, below) *except* for the following.

`fncmb(f1,op,f2)` returns the ppform of the function

$$x \mapsto f1(x) \; op \; f2(x)$$

with `op` one of `'+'`, `'-'`, `'*'`, and `f1`, `f2` of arbitrary polynomial form. If, in addition, `f2` is a scalar or vector, it is taken to be the function that is constantly equal to that scalar or vector.

Limitations   `fncmb` only works for *univariate* functions, except for the case `fncmb(function,operation)`, i.e., when there is just one function in the input.

Further, if two functions are involved, then they must be of the same type. This means that they must either both be in B-form or both be in

ppform, and, moreover, have the same knots or breaks, the same order, and the same target. The only exception to this is the command of the form `fncmb(function,op,function)`.

Purpose   Differentiate function

Syntax   fprime = fnder(f,dorder)
                fnder(f)

Description fprime = fnder(f,dorder) is the description of the dorderth
                derivative of the function whose description is contained in f. The
                default value of dorder is 1. For negative dorder, the particular
                |dorder|th indefinite integral is returned that vanishes |dorder|-fold
                at the left endpoint of the basic interval.

                The output is of the same form as the input, i.e., they are both ppforms
                or both B-forms or both stforms. fnder does not work for rational
                splines; for them, use fntlr instead. fnder works for stforms only in a
                limited way: if the type is tp00, then dorder can be [1,0] or [0,1].

                fnder(f) is the same as fnder(f,1).

                If the function in f is multivariate, say *m*-variate, then dorder must be
                given, and must be of length *m*.

Examples   If f is in ppform, or in B-form with its last knot of sufficiently high
                multiplicity, then, up to rounding errors, f and fnder(fnint(f)) are
                the same.

                If f is in ppform and fa is the value of the function in f at the
                left end of its basic interval, then, up to rounding errors, f and
                fnint(fnder(f),fa) are the same, unless the function described by f
                has jump discontinuities.

                If f contains the B-form of *f*, and t_1 is its leftmost knot, then, up to
                rounding errors, fnint(fnder(f)) contains the B-form of $f - f(t_1)$.
                However, its leftmost knot will have lost one multiplicity (if it had
                multiplicity > 1 to begin with). Also, its rightmost knot will have full
                multiplicity even if the rightmost knot for the B-form of *f* in f doesn't.

                Here is an illustration of this last fact. The spline in sp = spmak([0 0
                1], 1) is, on its basic interval [0..1], the straight line that is 1 at 0 and
                0 at 1. Now integrate its derivative: spdi = fnint(fnder(sp)). As

you can check, the spline in `spdi` has the same basic interval, but, on that interval, it agrees with the straight line that is 0 at 0 and −1 at 1.

See the examples "Intro to B-form" and "Intro to ppform" for examples.

Algorithms   For differentiation of either polynomial form, the derivatives are found in the piecewise-polynomial sense. This means that, in effect, each polynomial piece is differentiated separately, and jump discontinuities between polynomial pieces are ignored during differentiation.

For the B-form, the formulas [PGS; (X.10)] for differentiation are used.

For the stform, differentiation relies on knowing a formula for the relevant derivative of the basis function of the particular type.

See Also   fndir | fnint | fnplt | fnval

Purpose　　Directional derivative of function

Syntax　　`df = fndir(f,y)`

Description　　`df = fndir(f,y)` is the ppform of the directional derivative, of the function *f* in f, in the direction of the (column-)vector y. This means

that df describes the function $D_y f(x) := \lim_{t\to 0}(f(x+ty)-f(x))/t$.

If y is a matrix, with n columns, and *f* is d-valued, then the function in df is `prod(d)*n`-valued. Its value at *x*, reshaped to be of size [d,n], has in its *j*th "column" the directional derivative of *f* at *x* in the direction of the *j*th column of y. If you prefer df to reflect explicitly the actual size of *f*, use instead

```
df = fnchg( fndir(f,y), 'dim',[fnbrk(f,'dim'),size(y,2)] );
```

Since fndir relies on the ppform of the function in f, it does not work for rational functions nor for functions in stform.

Examples　　For example, if f describes an m-variate d-vector-valued function and x is some point in its domain, then, e.g., with this particular ppform f that describes a scalar-valued bilinear polynomial,

```
f = ppmak({0:1,0:1},[1 0;0 1]); x = [0;0];
[d,m] = fnbrk(f,'dim','var');
jacobian = reshape(fnval(fndir(f,eye(m)),x),d,m)
```

is the Jacobian of that function at that point (which, for this particular *scalar*-valued function, is its gradient, and it is zero at the origin).

As a related example, the next statements plot the gradients of (a good approximation to) the Franke function at a regular mesh:

```
xx = linspace(-.1,1.1,13); yy = linspace(0,1,11);
[x,y] = ndgrid(xx,yy); z = franke(x,y);
pp2dir = fndir(csapi({xx,yy},z),eye(2));
grads = reshape(fnval(pp2dir,[x(:) y(:)].'),...
```

```
        [2,length(xx),length(yy)]);
  quiver(x,y,squeeze(grads(1,:,:)),squeeze(grads(2,:,:)))
```

Here is the resulting plot.
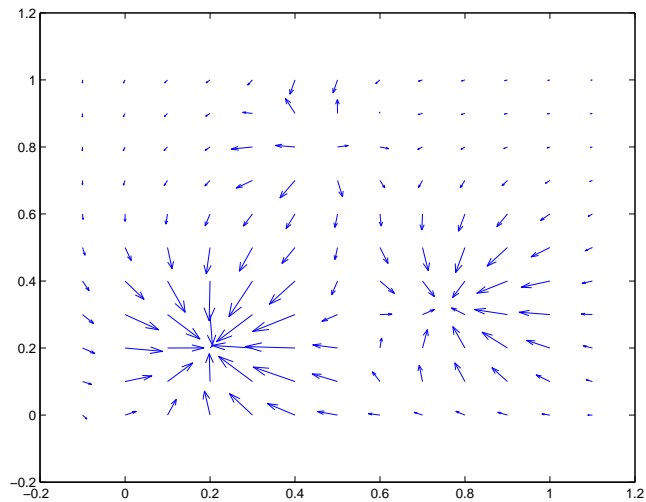
Algorithms   The function in `f` is converted to ppform, and the directional derivative of its polynomial pieces is computed formally and in one vector operation, and put together again to form the ppform of the directional derivative of the function in `f`.

See Also   fnchg | fnder | fnint | franke

Purpose   Integrate function

Syntax   intgrf = fnint(f,value)
                     fnint(f)

Description   intgrf = fnint(f,value) is the description of an indefinite integral
                     of the *univariate* function whose description is contained in f. The
                     integral is normalized to have the specified value at the left endpoint of
                     the function's basic interval, with the default value being zero.

                     The output is of the same type as the input, i.e., they are both ppforms
                     or both B-forms. fnint does not work for rational splines nor for
                     functions in stform.

                     fnint(f) is the same as fnint(f,0).

                     Indefinite integration of a *multivariate* function, in coordinate directions
                     only, is available via fnder(f,dorder) with dorder having nonpositive
                     entries.

Examples   The statement diff(fnval(fnint(f),[a b])) provides the definite
                     integral over the interval [a .. b] of the function described by f.

                     If f is in ppform, or in B-form with its last knot of sufficiently high
                     multiplicity, then, up to rounding errors, f and fnder(fnint(f)) are
                     the same.

                     If f is in ppform and fa is the value of the function in f at the
                     left end of its basic interval, then, up to rounding errors, f and
                     fnint(fnder(f),fa) are the same, unless the function described by f
                     has jump discontinuities.

                     If f contains the B-form of *f*, and t_1 is its leftmost knot, then, up to
                     rounding errors, fnint(fnder(f)) contains the B-form of $f - f(t_1)$.
                     However, its leftmost knot will have lost one multiplicity (if it had
                     multiplicity > 1 to begin with). Also, its rightmost knot will have full
                     multiplicity even if the rightmost knot for the B-form of *f* in f doesn't.

                     Here is an illustration of this last fact. The spline in sp = spmak([0 0
                     1], 1) is, on its basic interval [0..1], the straight line that is 1 at 0 and

0 at 1. Now integrate its derivative: `spdi = fnint(fnder(sp))`. As you can check, the spline in `spdi` has the same basic interval, but, on that interval, it agrees with the straight line that is 0 at 0 and -1 at 1.

See the examples "Intro to B-form" and "Intro to ppform" for examples.

Algorithms   For the B-form, the formula [PGS; (X.22)] for integration is used.

See Also   fnder | fnplt | fnval

Purpose   Jumps, i.e., f(x+)-f(x-)

Syntax   `jumps = fnjmp(f,x)`

Description   `jumps = fnjmp(f,x)` is like `fnval(f,x)` except that it returns the
jump $f(x+) - f(x-)$ across `x` (rather than the value at `x`) of the function f
described by `f` and that it only works for univariate functions.

This is a function for spline specialists.

Examples   `fnjmp(ppmak(1:4,1:3),1:4)` returns the vector `[0,1,1,0]` since the
pp function here is 1 on [1 .. 2], 2 on [2 .. 3], and 3 on [3 .. 4], hence has
zero jump at 1 and 4 and a jump of 1 across both 2 and 3.

If `x` is `cos([4:-1:0]*pi/4)`, then `fnjmp(fnder(spmak(x,1),3),x)`
returns the vector `[12 -24 24 -24 12]` (up to round-off). This is
consistent with the fact that the spline in question is a so called perfect
cubic B-spline, i.e., has an absolutely constant third derivative (on its
basic interval). The modified command

`fnjmp(fnder(fn2fm(spmak(x,1),'pp'),3),x)`

returns instead the vector `[0 -24 24 -24 0]`, consistent with the
fact that, in contrast to the B-form, a spline in ppform does not have
a discontinuity in any of its derivatives at the endpoints of its basic
interval. Note that `fnjmp(fnder(spmak(x,1),3),-x)` returns the
vector `[12,0,0,0,12]` since `-x`, though theoretically equal to `x`, differs
from `x` by roundoff, hence the third derivative of the B-spline provided
by `spmak(x,1)` does not have a jump across `-x(2)`,`-x(3)`, and `-x(4)`.

fnmin

Purpose	Minimum of function in given interval

Syntax

```
fnmin(f)
fnmin(f,interv)
[minval,minsite] = fnmin(f,...)
```

Description   fnmin(f) returns the minimum value of the *scalar-valued univariate* spline in f on its basic interval.

fnmin(f,interv) returns the minimum value on the interval [a..b] specified by interv.

[minval,minsite] = fnmin(f,...) also returns a location, minsite, at which the function in f takes that minimum value, minval.

Examples   **Example 1.** We construct and plot a spline *f* with many local extrema, then compute its *maximum* as the negative of the minimum of –*f*. We indicate this maximum value by adding a horizontal line to the plot at the height of the computed maximum.

```
rand('seed',21);
f = spmak(1:21,rand(1,15)-.5);
fnplt(f)
maxval = -fnmin(fncmb(f,-1));
hold on, plot(fnbrk(f,'interv'),maxval([1 1])), hold off
```
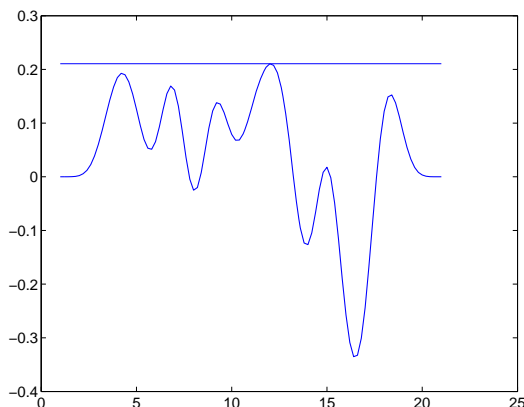
Example 2. Since `spmak(1:5,-1)` provides the negative of the cubic B-spline with knot sequence `1:5`, we expect the command

```
[y,x] = fnmin(spmak(1:5,-1))
```

to return `-2/3` for `y` and `3` for `x`.

Algorithm   `fnmin` first changes the basic interval of the function to the given interval, if any. On the interval, `fnmin` then finds all local extrema of the function as left and right limits at a jump and as zeros of the function's first derivative. It then evaluates the function at these extrema and at the endpoints of the interval, and determines the minimum over all these values.

fnplt

Purpose   Plot function

Syntax   fnplt(f)
                 fnplt(f,arg1,arg2,arg3,arg4)
                 points = fnplt(f,...)
                 [points, t] = fnplt(f,...)

Description fnplt(f) plots the function, described by f, on its basic interval.

If *f* is univariate, the following is plotted:

- If *f* is scalar-valued, the graph of *f* is plotted.

- If *f* is 2-vector-valued, the planar curve is plotted.

- If *f* is *d*-vector-valued with *d* > 2, the space curve given by the first three components of *f* is plotted.

If *f* is bivariate, the following is plotted:

- If *f* is scalar-valued, the graph of *f* is plotted (via surf).

- If *f* is 2-vector-valued, the image in the plane of a regular grid in its domain is plotted.

- If *f* is *d*-vector-valued with *d* > 2, then the parametric surface given by the first three components of its values is plotted (via surf).

If *f* is a function of more than two variables, then the bivariate function, obtained by choosing the midpoint of the basic interval in each of the variables other than the first two, is plotted.

fnplt(f,arg1,arg2,arg3,arg4) permits you to modify the plotting by the specification of additional input arguments. You can place these arguments in whatever order you like, chosen from the following list:

- A *string* that specifies a plotting symbol, such as '-.' or '*'; the default is '-'.

- A *scalar* to specify the linewidth; the default value is 1.

- A *string* that starts with the letter `'j'` to indicate that any jump in the *univariate* function being plotted should actually appear as a jump. The default is to fill in any jump by a (near-)vertical line.

- A *vector* of the form `[a,b]`, to indicate the interval over which to plot the univariate function in `f`. If the function in `f` is *m*-variate, then this optional argument must be a cell array whose ith entry specifies the interval over which the *i*th argument is to vary. In effect, for this arg, the command `fnplt(f,arg,...)` has the same effect as the command `fnplt(fnbrk(f,arg),...)`. The default is the basic interval of `f`.

- An empty matrix or string, to indicate use of default(s). You will find this option handy when your particular choice depends on some other variables.

`points = fnplt(f,...)` plots nothing, but the two-dimensional points or three-dimensional points it would have plotted are returned instead.

`[points, t] = fnplt(f,...)` also returns, for a vector-valued `f`, the corresponding vector `t` of parameter values.

Algorithms   A vector `x` of evaluation points is generated by the union of:

1 101 equally spaced sites filling out the plotting interval

2 Any breakpoints in the plotting interval

The univariate function *f* described by `f` is evaluated at these `x` evaluation points. If *f* is real-valued, the points $(x, f(x))$ are plotted. If *f* is vector-valued, then the first two or three components of $f(x)$ are plotted.

The bivariate function *f* described by `f` is evaluated on a 51-by-51 uniform grid if *f* is scalar-valued or *d*-vector-valued with $d > 2$ and the result plotted by `surf`. In the contrary case, *f* is evaluated along the meshlines of a 11-by-11 grid, and the resulting planar curves are plotted.

fnplt

Cautionary Note   The basic interval for *f* in B-form is the interval containing *all* the knots. This means that, e.g., *f* is sure to vanish at the endpoints of the basic interval unless the first and the last knot are both of full multiplicity *k*, with *k* the order of the spline *f*. Failure to have such full multiplicity is particularly annoying when *f* is a spline curve, since the plot of that curve as produced by fnplt is then bound to start and finish at the origin, regardless of what the curve might otherwise do.

Further, since B-splines are zero outside their support, any function in B-form is zero outside the basic interval of its form. This is very much in contrast to a function in ppform whose values outside the basic interval of the form are given by the extension of its leftmost, respectively rightmost, polynomial piece.

See Also   fnder | fnint | fnval

Purpose   Refine partition of form

Syntax   g = fnrfn(f,addpts)

Description   g = fnrfn(f,addpts) describes the same function as does f, but uses more terms to do it. This is of use when the sum of two or more functions of different forms is wanted or when the number of degrees of freedom in the form is to be increased to make fine local changes possible. The precise action depends on the form in f.

If the form in f is a B-form or BBform, then the entries of addpts are inserted into the existing knot sequence, subject to the following restriction: The multiplicity of no knot exceed the order of the spline. The equivalent B-form with this refined knot sequence for the function given by f is returned.

If the form in f is a ppform, then the entries of addpts are inserted into the existing break sequence, subject to the following restriction: The break sequence be strictly increasing. The equivalent ppform with this refined break sequence for the function in f is returned.

fnrfn does not work for functions in stform.

If the function in f is m-variate, then addpts must be a cell array, {addpts1,..., addptsm}, and the refinement is carried out in each of the variables. If the *i*th entry in this cell array is empty, then the knot or break sequence in the *i*th variable is unchanged.

Examples   Construct a spline in B-form, plot it, then apply two midpoint refinements, and also plot the control polygon of the resulting refined spline, expecting it to be quite close to the spline itself:

```
k = 4; sp = spapi( k, [1,1:10,10], [cos(1),sin(1:10),cos(10)] );
      fnplt(sp), hold on
      sp3 = fnrfn(fnrfn(sp));
      plot( aveknt( fnbrk(sp3,'knots'),k), fnbrk(sp3,'coefs'), 'r')
      hold off
```

A third refinement would have made the two curves indistiguishable.

Use fnrfn to add two B-splines of the same order:

```
B1 = spmak([0:4],1); B2 = spmak([2:6],1);
      B1r = fnrfn(B1,fnbrk(B2,'knots'));
      B2r = fnrfn(B2,fnbrk(B1,'knots'));
      B1pB2 = spmak(fnbrk(B1r,'knots'),fnbrk(B1r,'c')+fnbrk(B2r,'c'));
      fnplt(B1,'r'),hold on, fnplt(B2,'b'), fnplt(B1pB2,'y',2)
      hold off
```

Algorithms   The standard *knot insertion* algorithm is used for the calculation of the B-form coefficients for the refined knot sequence, while Horner's method is used for the calculation of the local polynomial coefficients at the additional breaks in the refined break sequence.

See Also   fncmb | ppmak | spmak

Purpose   Taylor coefficients or polynomial

Syntax   taylor = fntlr(f,dorder,x)
                    p = fntlr(f,dorder,x,interv)

Description   taylor = fntlr(f,dorder,x) returns the unnormalized Taylor
                    coefficients, up to the given order dorder and at the given x, of the
                    function described in f .

For a univariate function and a scalar x, this is the vector

$$T(f, \text{dorder}, x) := [f(x); Df(x); ...; D^{\text{dorder}-1}f(x)]$$

If, more generally, the function in f is d-valued with d>1 or
even prod(d)>1 and/or is m-variate for some m>1, then dorder is
expected to be an m-vector of positive integers, x is expected to
be a matrix with m rows, and, in that case, the output is of size
[prod(d)*prod(dorder),size(x,2)], with its j-th column containing

$$T(f, \text{dorder}, x(:,j))(i1,...,im) = D_1^{i1-1}...D_m^{im-1}f(x(:,j))$$

for i1=1:dorder(1), ..., im=1:dorder(m). Here, $D_i f$ is the partial
derivative of f with respect to its ith argument.

p = fntlr(f,dorder,x,interv) returns instead a ppform of the
Taylor polynomial at x of order dorder for the function described by f.
The basic interval for this ppform is as specified by interv. In this case
and assuming that the function described by f is m-variate, x is expected
to be of size [m,1], and interv is either of size [m,2] or else a cell array
of length m containing m vectors of size [1,2].

Examples   If f contains a univariate function and x is a scalar or a 1-row matrix,
                    then fntlr(f,3,x) produces the same output as the statements

```
df = fnder(f); [fnval(f,x); fnval(df,x); fnval(fnder(df),x)];
```

As a more complicated example, look at the Taylor vectors of order 3 at 21 equally spaced points for the rational spline whose graph is the unit circle:

```
ci = rsmak('circle'); in = fnbrk(ci,'interv');
t = linspace(in(1),in(2),21); t(end)=[];
v = fntlr(ci,3,t);
```

We plot `ci` along with the points `v(1:2,:)`, to verify that these are, indeed, points on the unit circle.

```
fnplt(ci), hold on, plot(v(1,:),v(2,:),'o')
```

Next, to verify that `v(3:4,j)` is a vector tangent to the circle at the point `v(1:2,j)`, we use the MATLAB `quiver` command to add the corresponding arrows to our plot:

```
quiver(v(1,:),v(2,:),v(3,:),v(4,:))
```

Finally, what about `v(5:6,:)`? These are second derivatives, and we add the corresponding arrows by the following `quiver` command, thus finishing First and Second Derivative of a Rational Spline Giving a Circle on page 13-113.

```
quiver(v(1,:),v(2,:),v(5,:),v(6,:)), axis equal, hold off
```
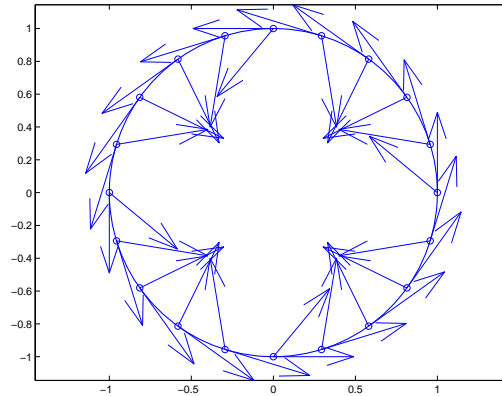
First and Second Derivative of a Rational Spline Giving a Circle

Now, our curve being a circle, you might have expected the 2nd derivative arrows to point straight to the center of that circle, and that would have been indeed the case if the function in ci had been using arclength as its independent variable. Since the parameter used is not arclength, we use the formula, given in "Example: B-form Spline Approximation to a Circle" on page 10-24, to compute the curvature of the curve given by ci at these selected points. For ease of comparison, we switch over to the variables used there and then simply use the commands from there.

```
dspt = v(3:4,:); ddspt = v(5:6,:);
kappa = abs(dspt(1,:).*ddspt(2,:)-dspt(2,:).*ddspt(1,:))./...
   (sum(dspt.^2)).^(3/2);
max(abs(kappa-1))
ans = 2.2204e-016
```

The numerical answer is reassuring: at all the points tested, the curvature is 1 to within roundoff.
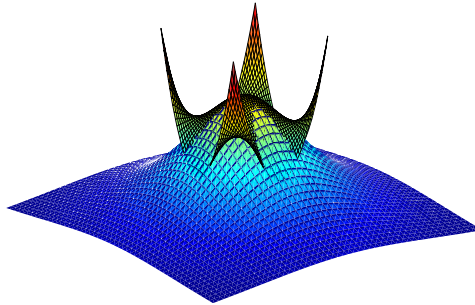
The Function 1/(1+x^2+y^2) and Its Taylor Polynomial of Order [3,3] at the Origin

As a final example, we start with a bivariate version of the Runge function, obtaining, for variety, a ppform for its denominator, $1 + x^2 + y^2$, by bicubic spline interpolation:

```
w = csapi({-1:1,-1:1},[3 2 3;2 1 2;3 2 3]);
```

Next, we make up the coefficient array for the numerator, 1, using exactly the same size, and put the two together into a rational spline:

```
wcoefs = fnbrk(w,'coef');
scoefs = zeros(size(wcoefs)); scoefs(end)=1;
runge2 = rpmak(fnbrk(w,'breaks'),[scoefs;wcoefs]);
```

Then we enlarge the basic interval for this rational spline, plot it and plot, on top of it, its Taylor polynomial at (0,0) of order [3,3].

```
fnplt(fnbrk(runge2,{[-2 2],[-2 2]}));  shading interp, hold on
fnplt(fntlr(runge2,[3 3],[0;0],[-.7 .7; -.7 .7]))
axis off, hold off
```

Since we shaded the function but not the Taylor polynomial, we can easily distinguish the two in the previous figure. We can also see that, in contrast to the function, the Taylor polynomial fails to be rotationally symmetric. This is due to the fact that it is a polynomial of order [3,3] rather than a polynomial of total order 3.

To obtain the Taylor polynomial of order 3, we get the Taylor polynomial of order [3,3], but with (0,0) the left point of its basic interval, set all its coefficients of total order bigger than 3 equal to zero, and then reconstruct the polynomial, and plot it, choosing a different view in order to show off the Taylor polynomial better. Here are the commands and the resulting figure.

```
taylor = fntlr(runge2,[3 3],[0;0],[0 1;0 1]);
tcoef = fnbrk(taylor,'coe'); tcoef([1 2 4]) = 0;
taylor2 = fnbrk(ppmak(fnbrk(taylor,'br'),tcoef),{[-1 1],[-1
1]});
fnplt(fnbrk(runge2,{[-2 2],[-2 2]}));  shading interp, hold on
fnplt(taylor2), view(-28,-26), axis off, hold off
```
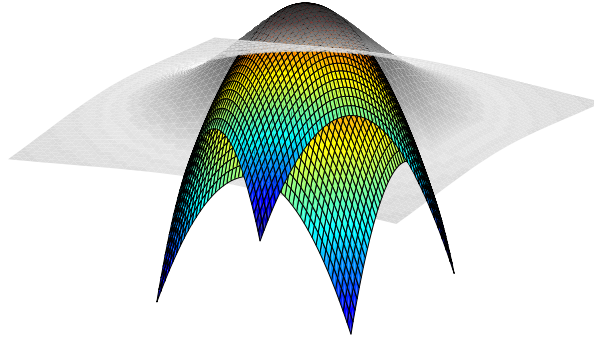
The Function 1/(1+x^2+y^2) and Its Taylor Polynomial of Order 3 at the Origin

See Also   `fnder` | `fndir`

Purpose   Evaluate function

Syntax   v = fnval(f,x)
                 fnval(x,f)
                 fnval(...,'l')

Description v = fnval(f,x) and v = fnval(x,f) both provide the value *f(x)* at
                 the points in x of the function *f* whose description is contained in f.

                 Roughly speaking, the output v is obtained by replacing each entry of x
                 by the value of *f* at that entry. This is literally true in case the function
                 in f is scalar-valued and univariate, and is the intent in all other cases,
                 except that, for a d-valued m-variate function, this means replacing
                 m-vectors by d-vectors. The full details are as follows.

                 For a univariate *f* :

                 • If *f* is scalar-valued, then v is of the same size as x.

                 • If *f* is [d1,...,dr]-valued, and x has size [n1,...,ns], then v has
                   size [d1,...,dr, n1,...,ns], with v(:,...,:, j1,...,js) the
                   value of *f* at x(j1,...,js), – except that

                   (1) n1 is ignored if it is 1 and s is 2, i.e., if x is a row vector; and

                   (2) MATLAB ignores any trailing singleton dimensions of x.

                 For an m-variate *f* with m>1, with *f* [d1,...,dr]-valued, x may be either
                 an array, or else a cell array {x1,...,xm}.

                 • If x is an array, of size [n1,...,ns] say, then n1 must equal m, and v
                   has size [d1,...,dr, n2,...,ns], with v(:,...,:, j2,...,js)
                   the value of *f* at x(:,j2,...,js), – except that

                   (1) d1, ..., dr is ignored in case *f* is scalar-valued, i.e., both r and n1
                   are 1;

                   (2) MATLAB ignores any trailing singleton dimensions of x.

                 • If x is a cell array, then it must be of the form {x1,...,xm}, with
                   xj a vector, of length nj, and, in that case, v has size [d1,...,dr,
                   n1,...,nm], with v(:,...,:, j1,...,jm) the value of *f* at (x1(j1),

..., xm(jm)), – except that d1, ..., dr is ignored in case *f* is scalar-valued, i.e., both r and n1 are 1.

If *f* has a jump discontinuity at x, then the value $f(x+)$, i.e., the limit from the right, is returned, except when x equals the right end of the basic interval of the form; for such x, the value $f(x-)$, i.e., the limit from the left, is returned.

fnval(x,f) is the same as fnval(f,x).

fnval(...,'l') treats *f* as continuous from the left. This means that if *f* has a jump discontinuity at x, then the value $f(x-)$, i.e., the limit from the left, is returned, except when x equals the left end of the basic interval; for such x, the value $f(x+)$ is returned.

If the function is *multivariate*, then the above statements concerning continuity from the left and right apply coordinatewise.

Examples   The statement fnval(csapi(x,y),xx) has the same effect as the statement csapi(x,y,xx).

Algorithms   For each entry of x, the relevant break- or knot-interval is determined and the relevant information assembled. Depending on whether f is in ppform or in B-form, nested multiplication or the B-spline recurrence (see, e.g., [*PGS*; X.(3)]) is then used vector-fashion for the simultaneous evaluation at all entries of x. Evaluation of a multivariate polynomial spline function takes full advantage of the tensor product structure.

Evaluation of a rational spline follows up evaluation of the corresponding vector-valued spline by division of all but its last component by its last component.

Evaluation of a function in stform makes essential use of stcol, and tries to keep the matrices involved to reasonable size.

See Also   fnbrk | ppmak | rsmak | spmak | stmak

Purpose   Extrapolate function

Syntax
```
g = fnxtr(f,order)
fnxtr(f)
```

Description   `g = fnxtr(f,order)` returns the spline (in ppform) that agrees with the spline in `f` on the latter's basic interval but is a polynomial of the given `order` outside it, with 2 the default for `order`, in such a way that the spline in `g` satisfies at least `order` smoothness conditions at the ends of `f`'s basic interval, i.e., at the new breaks.

f must be in B-form, BBform, or ppform.

While `order` can be any nonnegative integer, `fnxtr` is useful mainly when `order` is positive but less than the order of `f`.

If `order` is zero, then `g` describes the same spline as `fn2fm(f,'B-)` but is in ppform and has a larger basic interval.

If `order` is at least as big as `f`'s order, then `g` describes the same pp as `fn2fm(f,'pp')` but uses two more pieces and has a larger basic interval.

If `f` is m-variate, then `order` may be an m-vector, in which case `order(i)` specifies the matching order to be used in the `i`-th variable, `i = 1:m`.

If `order<0`, then `g` is exactly the same as `fn2fm(f,'pp')`. This unusual option is useful when, in the multivariate case, extrapolation is to take place in only some but not all variables.

`fnxtr(f)` is the same as `fnxtr(f,2)`.

Examples   **Example 1.** The cubic smoothing spline for given data x,y is, like any other 'natural' cubic spline, required to have zero second derivative outside the interval spanned by the data sites. Hence, if such a spline is to be evaluated outside that interval, it should be constructed as `s = fnxtr(csaps(x,y))`. A Cubic Smoothing Spline Properly Extrapolated on page 13-120, generated by the following code, shows the difference.

```
rand('seed',6); x = rand(1,21); s = csaps(x,x.^3); sn = fnxtr(s);
```

```
fnplt(s,[-.5 1.4],3), hold on, fnplt(sn,[-.5 1.4],.5,'r',2)
legend('cubic smoothing spline','... properly extrapolated')
set(gca,'Fontsize',16), axis off, hold off
```
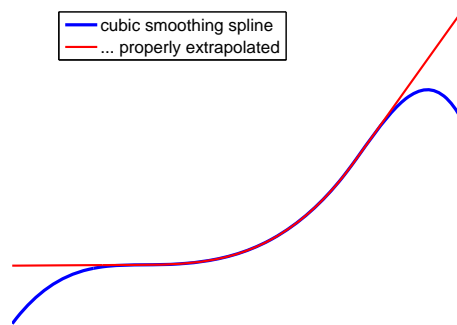
A Cubic Smoothing Spline Properly Extrapolated

Example 2. Here is the plot of a bivariate B-spline, quadratically extrapolated in the first variable and not at all extrapolated in the second, as generated by

```
fnplt(fnxtr(spmak({0:3,0:4},1),[3,-1]))
```
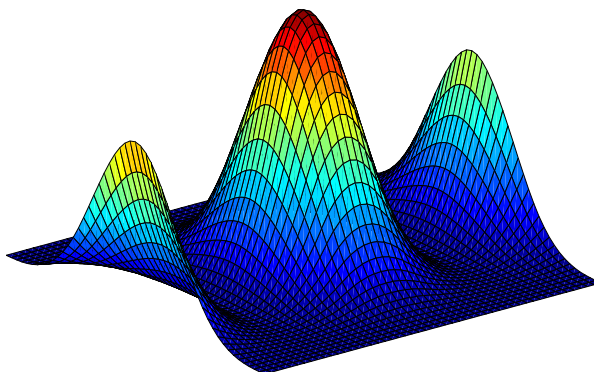
A Bivariate B-spline Quadratically Extrapolated In One Direction

See Also   ppmak | spmak | fn2fm

fnzeros

Purpose   Find zeros of function in given interval

Syntax   ```
                     z = fnzeros(f,[a b])
                     z = fnzeros(f)
                     ```

**Description**      `z = fnzeros(f,[a b])` is an ordered list of the zeros of the univariate
                     spline `f` in the interval [a .. b] .

                     `z = fnzeros(f)` is a list of the zeros in the basic interval of the spline `f`.

                     A spline zero is either a maximal closed interval over which the spline
                     is zero, or a zero crossing (a point across which the spline changes sign).

                     The list of zeros, `z`, is a matrix with two rows. The first row is the left
                     endpoint of the intervals and the second row is the right endpoint. Each
                     column `z(:,j)` contains the left and right endpoint of a single interval.

                     These intervals are of three kinds:

                     - If the endpoints are different, then the function is zero on the entire
                       interval. In this case the maximal interval is given, regardless of
                       knots that may be in the interior of the interval.

                     - If the endpoints are the same and coincident with a knot, then the
                       function in `f` has a zero at that point. The spline could cross zero,
                       touch zero or be discontinuous at this point.

                     - If the endpoints are the same and not coincident with a knot, then
                       the spline has a zero crossing at this point.

                     If the spline, `f`, touches zero at a point that is not a knot, but does not
                     *cross* zero, then this zero may not be found. If it is found, then it may
                     be found twice.

**Examples**         **Example 1.** The following code constructs and plots a piecewise linear
                     spline that has each of the three kinds of zeros: touch zero, cross zero,
                     and zero for an interval. `fnzeros` computes all the zeros, and then the
                     code plots the results on the graph.

                     ```
                     sp = spmak(augknt(1:7,2),[1,0,1,-1,0,0,1]);
                     fnplt(sp)
                     ```

```
z = fnzeros(sp)
nz = size(z,2);
hold on
plot(z(1,:),zeros(1,nz),'>',z(2,:),zeros(1,nz),'<'), hold off
```

This gives the following list of zeros:

```
z =
    2.0000    3.5000    5.0000
    2.0000    3.5000    6.0000
```

In this simple example, even for the second kind of zero, the two endpoints agree to all places.

Example 2. The following code generates and plots a spline function with many extrema and locates all extrema by computing the zeros of the spline function's first derivative there.

```
f = spmak( 1:21, rand( 1, 15 )-0.5 );
interval = fnbrk( f, 'interval' );
z = fnzeros( fnder( f ) );
z = z(1,:);
```

```
values = fnval( f, [interval, z] );
min( values )
fnplt(f)
hold on
plot(z,fnval(f,z),'ro')
hold off
```

Your plot will be different to the example following because of the use of rand to generate random coefficients.

Example 3. We construct a spline with a zero at a jump discontinuity and in B-form and find all the spline's zeros in an interval that goes beyond its basic interval.

```
sp = spmak([0 0 1 1 2],[1 0 -.2]);
fnplt(sp)
z = fnzeros(sp,[.5, 2.7])
zy = zeros(1,size(z,2));
hold on, plot(z(1,:),zy,'>',z(2,:),zy,'<'), hold off
```

This gives the following list of zeros:

```
z =
    1.0000    2.0000
    1.0000    2.7000
```

Notice the resulting zero interval [2..2.7], due to the fact that, by definition, a spline in B-form is identically zero outside its basic interval, [0..2].

Example 4. The following example shows the use of fnzeros with a discontinuous function. The following code creates and plots a discontinuous piecewise linear function, and finds the zeros.

```
 sp = spmak([0 0 1 1 2 2],[-1 1 -1 1]);
 fnplt(sp);
 fnzeros(sp)
```

This gives the following list of zeros, in (1..2) and (0..1) and the jump through zero at 1:

```
ans =
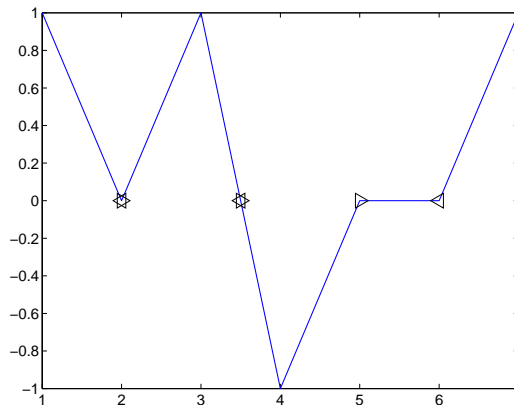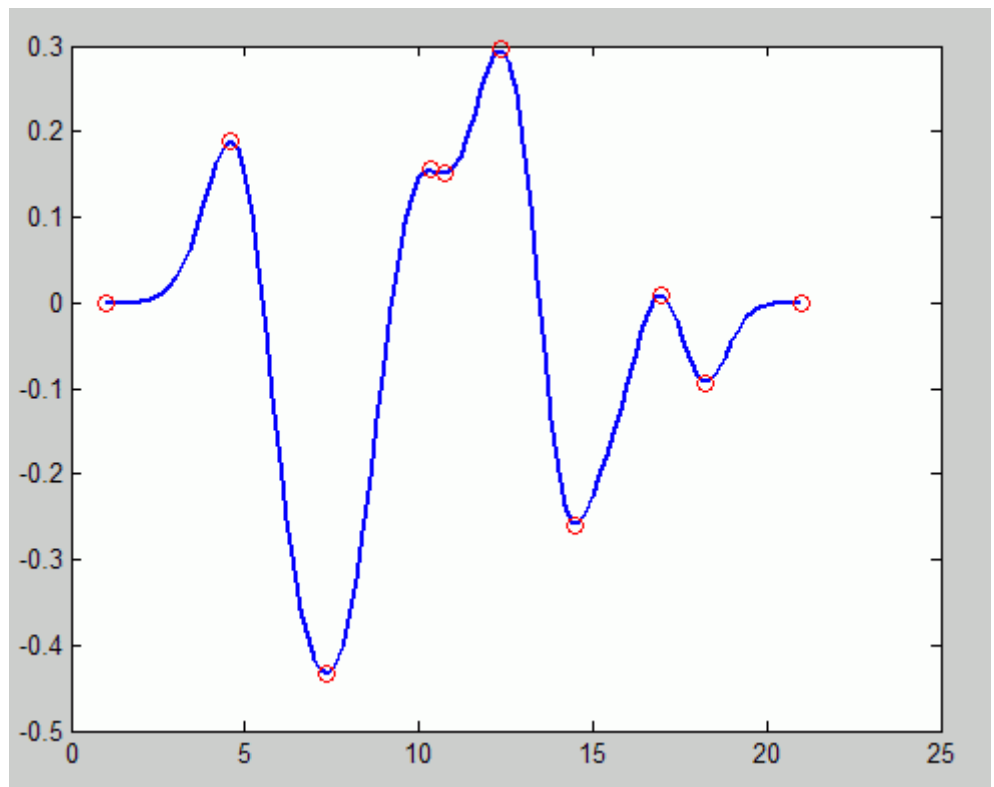
    0.5000    1.0000    1.5000
    0.5000    1.0000    1.5000
```

fnzeros

Algorithms   `fnzeros` first converts the function to B-form. The function performs some preprocessing to handle discontinuities, and then uses the algorithm of Mørken and Reimers.

Reference: Knut Mørken and Martin Reimers, An unconditionally convergent method for computing zeros of splines and polynomials, Math. Comp. 76:845--865, 2007.

See Also   `fnmin` | `fnval`

Purpose   Formula of `cfit`, `sfit`, or `fittype` object

Syntax   `formula(fun)`

Description   `formula(fun)` returns the formula of the `cfit`, `sfit`, or `fittype` object
                   `fun` as a character array.

Examples
```
f = fittype('weibull');
formula(f)
ans =
a*b*x^(b-1)*exp(-a*x^b)

g = fittype('cubicspline');
formula(g)
ans =
piecewise polynomial
```

See Also   `fittype` | `coeffnames` | `numcoeffs` | `probnames` | `coeffvalues`

franke

Purpose	Franke's bivariate test function
Syntax	`z = franke(x,y)`

Description `z = franke(x,y)` returns the value `z(i)` of Franke's function at the site `(x(i),y(i))`, `i=1:numel(x)`, with `z` of the same size as `x` and `y` (which must be of the same size).

Franke's function is the following weighted sum of four exponentials:

$$\frac{3}{4}e^{-\left((9x-2)^2+(9y-2)^2\right)/4} + \frac{3}{4}e^{-\left((9x+1)^2/49-(9y+1)/10\right)}$$

$$+ \frac{1}{2}e^{-\left((9x-7)^2+(9y-3)^2\right)/4} - \frac{1}{5}e^{-\left((9x-4)^2+(9y-7)^2\right)}$$

Examples The following commands provide a plot of Franke's function:

```
pts = (0:50)/50; [x,y] = ndgrid(pts,pts); z = franke(x,y);
surf(x,y,z), view(145,-2), set(gca,'Fontsize',16)
```


References [1] Richard Franke. "A critical comparison of some methods for interpolation of scattered data." *Naval Postgraduate School Tech.Rep.* NPS-53-79-003, March 1979.

Purpose	Get fit options structure property names and values

Syntax

```
get(options)
s = get(options)
value = get(options,fld)
```

Description   get(options) displays all property names and values of the fit options structure options.

s = get(options) returns a copy of the fit options structure options as the structure s.

value = get(options,*fld*) returns the value of the property *fld* of the fit options structure options. *fld* can be a cell array of strings, in which case value is also a cell array.

Examples

```
options = fitoptions('fourier1');
get(options,'Method')
ans =
NonlinearLeastSquares
get(options,'MaxIter')
ans =
   400
set(options,'Maxiter',1e3);
get(options,'MaxIter')
ans =
        1000
```

Property values can also be referenced and assigned using the dot notation. For example:

```
options.MaxIter
ans =
        1000
options.MaxIter = 500;
options.MaxIter
ans =
   500
```

See Also   fitoptions | set

Purpose   Interactive creation of cubic spline curve

Syntax   `[xy,spcv] = getcurve`

Description   `[xy,spcv] = getcurve` displays a gridded window and asks you for
input. As you click on points in the gridded window, the broken line
connecting these points is displayed. To indicate that you are done, click
outside the gridded window. Then a cubic spline curve, `spcv`, through
the point sequence, `xy`, is computed (via `cscvn`) and drawn. The point
sequence and, optionally, the spline curve are output.

If you want a closed curve, place the last point *close* to the initial point.

If you would like the curve to have a corner at some point, click on that
point twice (or more times) in succession.

See Also   `cscvn`

indepnames

Purpose	Independent variable of `cfit`, `sfit`, or `fittype` object
Syntax	`indep = indepnames(fun)`
Description	`indep = indepnames(fun)` returns the independent variable name or names (`indep`) of the `cfit`, `sfit`, or `fittype` object `fun`. For curves `indep` is a 1-by-1 cell array of strings, and for surfaces `indep` is a 2-by-1 cell array of strings.

Examples

```
f1 = fittype('a*x^2+b*exp(n*x)');
indep1 = indepnames(f1)
indep1 =
    'x'

f2 = fittype('a*x^2+b*exp(n*x)','independent','n');
indep2 = indepnames(f2)
indep2 =
    'n'
```

See Also   dependnames | fittype | formula

Purpose   Integrate `cfit` object

Syntax   `int = integrate(fun,x,x0)`

Description   `int = integrate(fun,x,x0)` integrates the `cfit` object `fun` at the points specified by the vector `x`, starting from `x0`, and returns the result in `int`. `int` is a vector the same size as `x`. `x0` is a scalar.

Examples   Create a baseline sinusoidal signal:

```
xdata = (0:.1:2*pi)';
y0 = sin(xdata);
```

Add noise to the signal:

```
noise = 2*y0.*randn(size(y0)); % Response-dependent
                               % Gaussian noise
ydata = y0 + noise;
```

Fit the noisy data with a custom sinusoidal model:

```
f = fittype('a*sin(b*x)');
fit1 = fit(xdata,ydata,f,'StartPoint',[1 1]);
```

Find the integral of the fit at the predictors:

```
int = integrate(fit1,xdata,0);
```

Plot the data, the fit, and the integral:

```
subplot(2,1,1)
plot(fit1,xdata,ydata) % cfit plot method
subplot(2,1,2)
plot(xdata,int,'m') % double plot method
grid on
legend('integral')
```

integrate

Note that integrals can also be computed and plotted directly with the `cfit` `plot` method, as follows:

```
plot(fit1,xdata,ydata,{'fit','integral'})
```

The `plot` method, however, does not return data on the integral.

See Also   fit | plot | differentiate

Purpose   Determine if `cfit`, `sfit`, or `fittype` object is linear

Syntax   `flag = islinear(fun)`

Description   `flag = islinear(fun)` returns a `flag` of 1 if the `cfit`, `sfit`, or `fittype` object `fun` represents a linear model, and a `flag` of 0 if it does not.

Note `islinear` assumes that all custom models specified by the `fittype` function using the syntax `ftype = fittype('expr')` are nonlinear models. To create a linear model with `fittype` that will be recognized as linear by `islinear` (and, importantly, by the algorithms of `fit`), use the syntax `ftype = fittype({'expr1','expr2',...,'exprn'})`.

Examples
```
f = fittype('a*x+b')
f =
     General model:
       f(a,b,x) = a*x+b
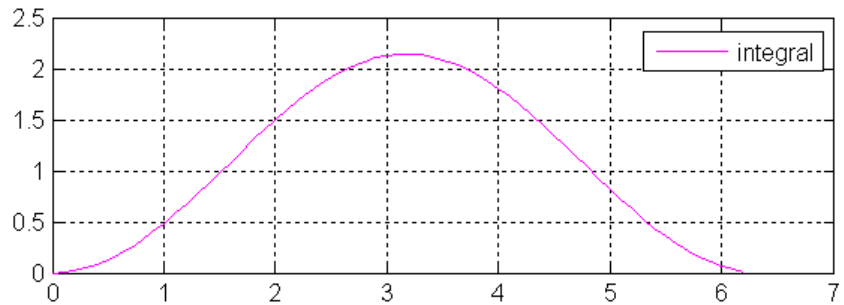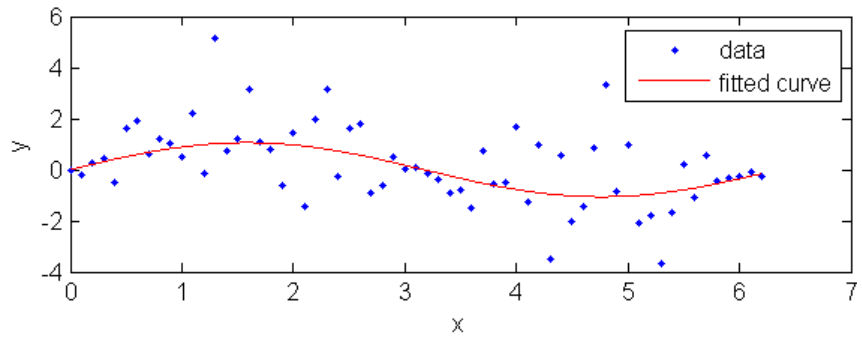
g = fittype({'x','1'})
g =
     Linear model:
       g(a,b,x) = a*x + b

h = fittype('poly1')
h =
     Linear model Poly1:
       h(p1,p2,x) = p1*x + p2

islinear(f)
ans =
     0
islinear(g)
ans =
```

islinear

```
             1
islinear(h)
ans =
             1
```

Purpose   Convert knots to breaks and their multiplicities

Syntax
```
knt2brk(knots)
[breaks,mults] = knt2brk(knots)
m = knt2mlt(t)
[m,sortedt] = knt2mlt(t)
```

Description   The commands extract the distinct elements from a sequence, as well as their multiplicities in that sequence, with *multiplicity* taken in two slightly different senses.

knt2brk(knots) returns the distinct elements in knots, and in increasing order, hence is the same as unique(knots).

[breaks,mults] = knt2brk(knots) additionally provides, in mults, the multiplicity with which each distinct element occurs in knots. Explicitly, breaks and mults are of the same length, and knt2brk is complementary to brk2knt in that, for any knot sequence knots, the two commands [xi,mlts] = knt2brk(knots); knots1 = brk2knt(xi,mlts); give knots1 equal to knots.

m = knt2mlt(t) returns a vector of the same length as t, with m(i) counting, in the vector sort(t), the number of entries before its *i*th entry that are equal to that entry. This kind of multiplicity vector is needed in spapi or spcol where such multiplicity is taken to specify which particular derivatives are to be matched at the sites in t. Precisely, if t is nondecreasing and z is a vector of the same length, then sp = spapi(knots, t, z) attempts to construct a spline *s* (with knot sequence knots) for which $D^{m(i)}s(t(i))$ equals $z(i)$, all *i*.

[m,sortedt] = knt2mlt(t) also returns the output from sort(t).

Neither knt2brk nor knt2mlt is likely to be used by the casual user of this toolbox.

Examples   [xi,mlts]=knt2brk([1 2 3 3 1 3]) returns [1 2 3] for xi and [2 1 3] for mlts.

[m,t]=knt2mlt([1 2 3 3 1 3]) returns [0 1 0 0 1 2] for m and
[1 1 2 3 3 3] for t.

See Also   brk2knt | spapi | spcol

Purpose	New break distribution

Syntax	`newknots = newknt(f,newl)`
	`newknt(f)`
	`[...,distfn] = newknt(...)`

Description	`newknots = newknt(f,newl)` returns the knot sequence whose interior knots cut the basic interval of `f` into `newl` pieces, in such a way that a certain piecewise linear monotone function related to the high derivative of `f` is equidistributed.

The intent is to choose a knot sequence suitable to the fine approximation of a function *g* whose rough approximation in `f` is assumed to contain enough information about *g* to make this feasible.

`newknt(f)` uses for `newl` its default value, namely the number of polynomial pieces in `f`.

`[...,distfn] = newknt(...)` also returns, in `distfn`, the ppform of that piecewise linear monotone function being equidistributed.

Examples	If the error in the least-squares approximation `sp` to some data `x,y` by a spline of order `k` seems uneven, you might try for a more equitable distribution of knots by using

`spap2(newknt(sp),k,x,y);`

For another example, see the last part of the example "Solving an ODE by Collocation".

Algorithms	This is the Fortran routine NEWNOT in PGS. With *k* the order of the piecewise-polynomial function *f* in pp, the function $|D^k f|$ is approximated by a piecewise constant function obtained by local, discrete, differentiation of the variation of $D^{k-1}f$. The new break sequence is chosen to subdivide the basic interval of the piecewise-polynomial *f* in such a way that

$$\int_{\text{newknots}(i)}^{\text{newknots}(i+1)} |\, D^k f\, |^{1/k} = \text{const}, \ \text{i} = k : k + \text{newl} - 1$$

Purpose   Number of input arguments of cfit, sfit, or fittype object

Syntax   nargs = numargs(fun)

Description   nargs = numargs(fun) returns the number of input arguments nargs
                   of the cfit, sfit, or fittype object fun.

Examples   f = fittype('a*x^2+b*exp(n*x)');
                   nargs = numargs(f)
                   nargs =
                        4
                   args = argnames(f)
                   args =
                        'a'
                        'b'
                        'n'
                        'x'

See Also   fittype | formula | argnames

numcoeffs

Purpose	Number of coefficients of cfit, sfit, or fittype object
Syntax	ncoeffs = numcoeffs(fun)
Description	ncoeffs = numcoeffs(fun) returns the number of coefficients ncoeffs of the cfit, sfit, or fittype object fun.

Examples

```
f = fittype('a*x^2+b*exp(n*x)');
ncoeffs = numcoeffs(f)
ncoeffs =
     3
coeffs = coeffnames(f)
coeffs =
    'a'
    'b'
    'n'
```

See Also   fittype | formula | coeffnames

Purpose   Knot distribution "optimal" for interpolation

Syntax   knots = optknt(tau,k,maxiter)
                     optknt(tau,k)

Description   knots = optknt(tau,k,maxiter) provides the knot sequence t that
                     is *best* for interpolation from $S_{k,t}$ at the site sequence tau, with 10
                     the default for the optional input maxiter that bounds the number
                     of iterations to be used in this effort. Here, *best* or *optimal* is used
                     in the sense of Micchelli/Rivlin/Winograd and Gaffney/Powell, and
                     this means the following: For any *recovery scheme R* that provides
                     an interpolant *Rg* that matches a given *g* at the sites tau(1),
                     ..., tau(n), we may determine the smallest constant const_R for
                     which $||g - Rg|| \leq \text{const}_R ||D^k_g||$ for all smooth functions *g*.

                     Here, $||f||:=\sup_{\text{tau}(1) < x < \text{tau(n)}}|f(x)|$. Then we may look for the optimal
                     recovery scheme as the scheme *R* for which const_R is as small as
                     possible. Micchelli/Rivlin/Winograd have shown this to be interpolation
                     from $S_{k,t}$, with t uniquely determined by the following conditions:

                     1 t(1) = ... = t(k) = tau(1);

                     2 t(n+1) = ... = t(n+k) = tau(n);

                     3 Any absolutely constant function *h* with sign changes at the sites
                        t(k+1), ..., t(n) and nowhere else satisfies

                        $$\int_{\text{tau}(1)}^{\text{tau}(n)} f(x)h(x)dx = 0 \text{ for all } f \in S_{k,t}$$

                     Gaffney/Powell called this interpolation scheme *optimal* since it
                     provides the *center* function in the band formed by all interpolants to
                     the given data that, in addition, have their *k*th derivative between *M*
                     and –*M* (for large *M*).

                     optknt(tau,k) is the same as optknt(tau,k,10).

optknt

Examples
See the last part of the example "Spline Interpolation" for an illustration. For the following highly nonuniform knot sequence

```
t = [0, .0012+[0, 1, 2+[0,.1], 4]*1e-5, .002, 1];
```

the command `optknt(t,3)` will fail, while the command `optknt(t,3,20)`, using a high value for the optional parameter `maxiter,` will succeed.

Algorithms
This is the Fortran routine SPLOPT in PGS. It is based on an algorithm described in , for the construction of that sign function *h* mentioned above. It is essentially Newton's method for the solution of the resulting nonlinear system of equations, with `aveknt(tau,k)` providing the first guess for `t(k+1)`, ...,`t(n)`, and some damping used to maintain the Schoenberg-Whitney conditions.

References
[1]C. de Boor, "Computational aspects of optimal recovery", in *Optimal Estimation in Approximation Theory*, C.A. Micchelli & T.J. Rivlin eds., Plenum Publ., New York, 1977, 69-91.

[2]P.W. Gaffney & M.J.D. Powell, "Optimal interpolation", in *Numerical Analysis*, G.A. Watson ed., *Lecture Notes in Mathematics, No. 506*, Springer-Verlag, 1976, 90-99.

[3]C.A. Micchelli, T.J. Rivlin & S. Winograd, "The optimal recovery of smooth functions", *Numer. Math.* **80**, (1974), 903-906.

See Also
aptknt | aveknt | newknt

Purpose   Plot `cfit` or `sfit` object

Syntax
```
plot(sfit)
plot(sfit, [x, y], z)
plot(sfit, [x, y], z, 'Exclude', ExcludeData)
H = plot(sfit, ..., 'Style', Style)
H = plot(sfit, ..., 'Level', Level)
H = plot(sfit, ..., 'XLim', XLIM)
H = plot(sfit, ..., 'YLim', YLIM)
H = plot(sfit, ...)
H = plot(sfit, ..., 'Parent', HAXES )
plot(cfit)
plot(cfit,x,y)
plot(cfit,x,y,DataLineSpec)
plot(cfit,FitLineSpec,x,y,DataLineSpec)
plot(cfit,x,y,outliers)
plot(cfit,x,y,outliers,OutlierLineSpec)
plot(...,ptype,...)
plot(...,ptype,level)
h = plot(...)
```

Description   For **surfaces**:

- `plot(sfit)` plots the `sfit` object over the range of the current axes, if any, or otherwise over the range stored in the fit.

- `plot(sfit, [x, y], z)` plots z versus x and y and plots `sfit` over the range of x and y.

- `plot(sfit, [x, y], z, 'Exclude', ExcludeData)` plots the excluded data in a different color. `ExcludeData` is a logical array where `true` represents an outlier.

- `H = plot(sfit, ..., 'Style', Style)` selects which way to plot the surface fit object `sfit`.

  `Style` may be any of the following strings

  - `'Surface'` Plot the fit object as a surface (default)

- ▪ `'PredFunc'` Surface with prediction bounds for function

- ▪ `'PredObs'` Surface with prediction bounds for new observation

- ▪ `'Residuals'` Plot the residuals (fit is the plane Z=0)

- ▪ `'Contour'` Make a contour plot of the surface

- `H = plot(sfit, ..., 'Level', Level)` sets the confidence level to be used in the plot. `Level` is a positive value less than 1, and has a default of 0.95 (for 95% confidence). This option only applies to the `'PredFunc'` and `'PredObs'` plot styles.

- `H = plot(sfit, ..., 'XLim', XLIM)` and `H = plot(sfit, ..., 'YLim', YLIM)` sets the limits of the axes used for the plot. By default the axes limits are taken from the data, `XY`. If no data is given, then the limits are taken from the surface fit object, `sfit`.

- `H = plot(sfit, ...)` returns a vector of handles of the plotted objects.

- `H = plot(sfit, ..., 'Parent', HAXES )` plots the fit object `sfit` in the axes with handle HAXES rather than the current axes. The range is taken from these axes rather than from the fit or the data.

For **curves**:

- `plot(cfit)` plots the `cfit` object over the domain of the current axes, if any. If there are no current axes, and `fun` is an output from the `fit` function, the plot is over the domain of the fitted data.

- `plot(cfit,x,y)` plots `cfit` together with the predictor data `x` and the response data `y`.

- `plot(cfit,x,y,DataLineSpec)` plots the predictor and response data using the color, marker symbol, and line style specified by the `DataLineSpec` formatting string. `DataLineSpec` strings take the same values as `LineSpec` strings used by the MATLAB `plot` function.

- `plot(cfit,FitLineSpec,x,y,DataLineSpec)` plots `fun` using the color, marker symbol, and line style specified by the `FitLineSpec` formatting string, and plots `x` and `y` using the color, marker symbol, and line style specified by the `DataLineSpec` formatting string.

FitLineSpec and DataLineSpec strings take the same values as LineSpec strings used by the MATLAB plot function.

- plot(cfit,x,y,outliers) plots data indicated by outliers in a different color. outliers is a logical array the same size as x and y. outliers can be computed with the excludedata function.

- plot(cfit,x,y,outliers,OutlierLineSpec) plots outliers using the color, marker symbol, and line style specified by the OutlierLineSpec. OutlierLineSpec strings take the same values as LineSpec strings used by the MATLAB plot function.

  plot(...,ptype,...) uses the plot type specified by ptype. Supported plot types are:

  - 'fit' — Data and fit (default)
  - 'predfunc' — Data and fit with prediction bounds for the fit
  - 'predobs' — Data and fit with prediction bounds for new observations
  - 'residuals' — Residuals
  - 'stresiduals' — Standardized residuals (residuals divided by their standard deviation).
  - 'deriv1' — First derivative of the fit
  - 'deriv2' — Second derivative of the fit
  - 'integral' — Integral of the fit

- plot(...,ptype,level) plots prediction intervals with a confidence level specified by level. level must be between 0 and 1. The default value of level is 0.95.

For both curves and surfaces:

- Plot types can be single or multiple, with multiple plot types specified as a cell array of strings. With a single plot type, plot draws to the current axes and can be used with commands like hold and subplot. With multiple plot types, plot creates subplots for each plot type.

- h = plot(...) returns a vector of handles to the plotted objects.

Examples   Create a baseline sinusoidal signal:

```
xdata = (0:0.1:2*pi)';
y0 = sin(xdata);
```

Add noise to the signal with non-constant variance:

```
% Response-dependent Gaussian noise
gnoise = y0.*randn(size(y0));

% Salt-and-pepper noise
spnoise = zeros(size(y0));
p = randperm(length(y0));
sppoints = p(1:round(length(p)/5));
spnoise(sppoints) = 5*sign(y0(sppoints));

ydata = y0 + gnoise + spnoise;
```

Fit the noisy data with a baseline sinusoidal model:

```
f = fittype('a*sin(b*x)');
fit1 = fit(xdata,ydata,f,'StartPoint',[1 1]);
```

Identify "outliers" as points at a distance greater than 1.5 standard deviations from the baseline model, and refit the data with the outliers excluded:

```
fdata = feval(fit1,xdata);
I = abs(fdata - ydata) > 1.5*std(ydata);
outliers = excludedata(xdata,ydata,'indices',I);

fit2 = fit(xdata,ydata,f,'StartPoint',[1 1],...
           'Exclude',outliers);
```

Compare the effect of excluding the outliers with the effect of giving them lower bisquare weight in a robust fit:

```
fit3 = fit(xdata,ydata,f,'StartPoint',[1 1],'Robust','on');
```

Plot the data, the outliers, and the results of the fits:

```
plot(fit1,'r-',xdata,ydata,'k.',outliers,'m*')
hold on
plot(fit2,'c--')
plot(fit3,'b:')
xlim([O 2*pi])
```


Plot the residuals for the two fits considering outliers:

```
figure
plot(fit2,xdata,ydata,'co','residuals')
hold on
plot(fit3,xdata,ydata,'bx','residuals')
```

See Also   cftool | excludedata | fit | differentiate | integrate

Purpose   Put together spline in ppform

Syntax   ppmak(breaks,coefs)
ppmak
ppmak(breaks,coefs,d)
ppmak(breaks,coefs,sizec)

Description   The command ppmak(...) puts together a spline in ppform from
minimal information, with the rest inferred from that information.
fnbrk provides any or all of the parts of the completed description. In
this way, the actual data structure used for the storage of the ppform is
easily modified without any effect on the various fn... commands that
use this construct. However, the casual user is not likely to use ppmak
explicitly, relying instead on the various spline construction commands
in the toolbox to construct particular splines.

ppmak(breaks,coefs) returns the ppform of the spline specified by the
break information in breaks and the coefficient information in coefs.
How that information is interpreted depends on whether the function
is univariate or multivariate, as indicated by breaks being a sequence
or a cell array.

If breaks is a sequence, it must be nondecreasing, with its first entry
different from its last. Then the function is assumed to be univariate,
and the various parts of its ppform are determined as follows:

1 The number l of polynomial pieces is computed as
length(breaks)-1, and the basic interval is, correspondingly, the
interval [breaks(1) .. breaks(l+1)].

2 The dimension d of the function's target is taken to be the number
of rows in coefs. In other words, each column of coefs is taken to
be one coefficient. More explicitly, coefs(:,i*k+j) is assumed to
contain the jth coefficient of the (i+1)st polynomial piece (with the
first coefficient the highest and the kth coefficient the lowest, or
constant, coefficient). Thus, with kl the number of columns of coefs,
the order k of the piecewise-polynomial is computed as fix(kl/l).

After that, the entries of `coefs` are reordered, by the command

```
 coefs = reshape(permute(reshape(coefs,[d,k,l]),[1 3 2]),[d*l,k])
```

in order to conform with the internal interpretation of the coefficient array in the ppform for a univariate spline. This only applies when you use the syntax `ppmak(breaks,coefs)` where `breaks` is a sequence (row vector), not when it is a cell-array. The permutation is not made when you use the three-argument forms of `ppmak`. For the three-argument forms only a reshape is done, not a permute.

If `breaks` is a cell array, of length `m`, then the function is assumed to be m-variate (tensor product), and the various parts of its ppform are determined from the input as follows:

1 The m-vector l has `length(breaks{i})-1` as its `ith` entry and, correspondingly, the m-cell array of its basic intervals has the interval `[breaks{i}(1) .. breaks{i}(end)]` as its `ith` entry.

2 The dimension `d` of the function's target and the m-vector `k` of (coordinate-wise polynomial) orders of its pieces are obtained directly from the size of `coefs`, as follows.

   a If `coefs` is an m-dimensional array, then the function is taken to be scalar-valued, i.e., `d` is 1, and the m-vector `k` is computed as `size(coefs)./l`. After that, `coefs` is reshaped by the command `coefs = reshape(coefs,[1,size(coefs)]).`

   b If `coefs` is an (r+m)-dimensional array, with `sizec = size(c)` say, then `d` is set to `sizec(1:r)`, and the vector `k` is computed as `sizec(r+(1:m))./l`. After that, `coefs` is reshaped by the command `coefs = reshape(coefs,[prod(d),sizec(r+(1:m))]).`

Then, `coefs` is interpreted as an equivalent array of size `[d,l(1),k(1),l(2),k(2),...,l(m),k(m)]`, with its `(:,i(1),r(1),i(2),r(2),...,i(m),r(m))th` entry the coefficient of

$$\prod_{\mu=1}^{m} \left(x(\mu) - \text{breaks} \,|\, \mu](i(\mu)) \right)^{(k(\mu) - r(\mu))}$$

in the local polynomial representation of the function on the (hyper)rectangle with sides

$$[\text{breaks} \,|\, \mu](i(\mu)) \,..\, \text{breaks} \,|\, \mu](i(\mu)+1)], \quad \mu = 1:m$$

This is, in fact, the internal interpretation of the coefficient array in the ppform of a multivariate spline.

`ppmak` prompts you for `breaks` and `coefs`.

`ppmak(breaks,coefs,d)` with `d` a positive integer, also puts together the ppform of a spline from the information supplied, but expects the function to be univariate. In that case, `coefs` is taken to be of size `[d*l,k]`, with `l` obtained as `length(breaks)-1`, and this determines the order, `k`, of the spline. With this, `coefs(i*d+j,:)` is taken to be the `j`th components of the coefficient vector for the (`i+1`)st polynomial piece.

`ppmak(breaks,coefs,sizec)` with `sizec` a row vector of positive integers, also puts together the ppform of a spline from the information supplied, but interprets `coefs` to be of size `sizec` (and returns an error when `prod(size(coefs))` differs from `prod(sizec)`). This option is important only in the rare case that the input argument `coefs` is an array with one or more trailing singleton dimensions. For, MATLAB suppresses trailing singleton dimensions, hence, without this explicit specification of the intended size of `coefs`, `ppmak` would interpret `coefs` incorrectly.

Examples   The two splines

```
p1 = ppmak([1 3 4],[1 2 5 6;3 4 7 8]);
p2 = ppmak([1 3 4],[1 2;3 4;5 6;7 8],2);
```

have exactly the same ppform (2-vector-valued, of order 2). But the second command provides the coefficients in the arrangement used internally.

ppmak

`ppmak([0:2],[1:6])` constructs a piecewise-polynomial function with basic interval [0..2] and consisting of two pieces of order 3, with the sole interior break 1. The resulting function is scalar, i.e., the dimension d of its target is 1. The function happens to be continuous at that break since the first piece is $x|\rightarrow x^2 + 2x + 3$, while the second piece is $x|\rightarrow 4(x-1)^2 + 5(x-1) + 6$.

When the function is univariate and the dimension d is not explicitly specified, then it is taken to be the row number of `coefs`. The column number should be an integer multiple of the number l of pieces specified by `breaks`. For example, the statement `ppmak([0:2],[1:3;4:6])` leads to an error, since the break sequence `[0:2]` indicates two polynomial pieces, hence an even number of columns are expected in the coefficient matrix. The modified statement `ppmak([0:1],[1:3;4:6])` specifies the parabolic curve $x|\rightarrow(1,4)x^2 + (2,5)x + (3,6)$. In particular, the dimension d of its target is 2. The differently modified statement `ppmak([0:2],[1:4;5:8])` also specifies a planar curve (i.e., d is 2), but this one is piecewise linear; its first polynomial piece is $x|\rightarrow(1,5)x + (2,6)$.

Explicit specification of the dimension d leads, in the univariate case, to a different interpretation of the entries of `coefs`. Now the column number indicates the polynomial order of the pieces, and the row number should equal d times the number of pieces. Thus, the statement `ppmak([0:2],[1:4;5:8],2)` is in error, while the statement `ppmak([0:2],[1:4;5:8],1)` specifies a scalar piecewise cubic whose first piece is $x|\rightarrow x^3 + 2x^2 + 3x + 4$.

If you wanted to make up the constant polynomial, with basic interval [0..1] say, whose value is the matrix eye(2), then you would have to use the full optional third argument, i.e., use the command

```
pp = ppmak(0:1,eye(2),[2,2,1,1]);
```

Finally, if you want to construct a 2-vector-valued bivariate polynomial on the rectangle [–1 .. 1] x [0 .. 1], linear in the first variable and constant in the second, say

```
coefs = zeros(2,2,1); coefs(:,:,1) = [1 0; 0 1];
```

then the straightforward

```
pp = ppmak({[-1 1],[0 1]},coefs);
```

will fail, producing a scalar-valued function of order 2 in each variable, as will

```
pp = ppmak({[-1 1],[0 1]},coefs,size(coefs));
```

while the following command will succeed:

```
pp = ppmak({[-1 1],[0 1]},coefs,[2 2 1]);
```

See the example "Intro to ppform" for other examples.

See Also   fnbrk

predint

Purpose	Prediction intervals for `cfit` or `sfit` object

Syntax	`ci = predint(fitresult,x)`
	`ci = predint(fitresult,x,level)`
	`ci = predint(fitresult,x,level,`*`intopt`*`,`*`simopt`*`)`
	`[ci,y] = predint(...)`

Description	`ci = predint(fitresult,x)` returns upper and lower 95% prediction bounds for response values associated with the `cfit` object `fitresult` at the new predictor values specified by the vector `x`. `fitresult` must be an output from the `fit` function to contain the necessary information for `ci`. `ci` is an n-by-2 array where `n = length(x)`. The left column of `ci` contains the lower bound for each coefficient; the right column contains the upper bound.

`ci = predint(fitresult,x,level)` returns prediction bounds with a confidence level specified by `level`. `level` must be between 0 and 1. The default value of `level` is 0.95.

`ci = predint(fitresult,x,level,`*`intopt`*`,`*`simopt`*`)` specifies the type of bounds to compute.

intopt is one of

- `'observation'` — Bounds for a new observation (default)

- `'functional'` — Bounds for the fitted curve

simopt is one of

- `'off'` — Non-simultaneous bounds (default)

- `'on'` — Simultaneous bounds

Observation bounds are wider than functional bounds because they measure the uncertainty of predicting the fitted curve plus the random variation in the new observation. Non-simultaneous bounds are for individual elements of `x`; simultaneous bounds are for all elements of `x`.

`[ci,y] = predint(...)` returns the response values `y` predicted by `fitresult` at the predictors in `x`.

Note predint cannot compute prediction intervals for non-parametric regression methods such as Interpolant, Lowess, and Spline.

Examples   Generate data with an exponential trend:

```
x = (0:0.2:5)';
y = 2*exp(-0.2*x) + 0.5*randn(size(x));
```

Fit the data using a single-term exponential:

```
fitresult = fit(x,y,'exp1');
```

Compute prediction intervals:

```
p11 = predint(fitresult,x,0.95,'observation','off');
p12 = predint(fitresult,x,0.95,'observation','on');
p21 = predint(fitresult,x,0.95,'functional','off');
p22 = predint(fitresult,x,0.95,'functional','on');
```

Plot the data, fit, and prediction intervals:

```
subplot(2,2,1)
plot(fitresult,x,y),hold on,plot(x,p11,'m--'),xlim([0 5])
title('Nonsimultaneous observation bounds','Color','m')
subplot(2,2,2)
plot(fitresult,x,y),hold on,plot(x,p12,'m--'),xlim([0 5])
title('Simultaneous observation bounds','Color','m')
subplot(2,2,3)
plot(fitresult,x,y),hold on,plot(x,p21,'m--'),xlim([0 5])
title('Nonsimultaneous functional bounds','Color','m')
subplot(2,2,4)
plot(fitresult,x,y),hold on,plot(x,p22,'m--'),xlim([0 5])
title('Simultaneous functional bounds','Color','m')
```

See Also   confint | fit | plot

prepareSurfaceData

Purpose   Prepare data inputs for surface fitting

Syntax   [XOut, YOut, ZOut] = prepareSurfaceData(XIn, YIn, ZIn)
                     [XOut, YOut, ZOut, WOut] = prepareSurfaceData(XIn, YIn, ZIn,
                         WIn)

Description   [XOut, YOut, ZOut] = prepareSurfaceData(XIn, YIn, ZIn)
                     transforms data, if necessary, for surface fitting with the fit function.
                     The function transforms data as follows:

- For table data, transform row (YIn) and column (XIn) headers into
  arrays YOut and XOut that are the same size as ZIn. Warn if XIn
  and YIn are reversed.

- Return data as columns regardless of the input shapes. Error if the
  number of elements do not match. Warn if the number of elements
  match, but the sizes are different.

- Convert complex to real (remove imaginary parts) and warn of this
  conversion.

- Remove NaN or Inf from data and warn of this removal.

- Convert nondouble to double and warn of this conversion.

[XOut, YOut, ZOut, WOut] = prepareSurfaceData(XIn, YIn,
ZIn, WIn) transforms data including weights (WIn) for surface fitting
with the fit function.

See Also   fit | excludedata

probnames

Purpose	Problem-dependent parameter names of `cfit`, `sfit`, or `fittype` object
Syntax	`pnames = probnames(fun)`
Description	`pnames = probnames(fun)` returns the names of the problem-dependent (fixed) parameters of the `cfit`, `sfit`, or `fittype` object `fun` as a cell array of strings.

Examples

```
f = fittype('(x-a)^n + b','problem',{'a','b'});
coeffnames(f)
ans =
    'n'
probnames(f)
ans =
    'a'
    'b'

load census

c = fit(cdate,pop,f,'problem',{cdate(1),pop(1)},...
        'StartPoint',2);
coeffvalues(c)
ans =
    0.9877
probvalues(c)
ans =
  1.0e+003 *
    1.7900    0.0039
```

See Also   `fittype` | `coeffnames` | `probvalues`

probvalues

Purpose   Problem-dependent parameter values of `cfit` or `sfit` object

Syntax   pvals = probvalues(fun)

Description pvals = probvalues(fun) returns the values of the problem-dependent (fixed) parameters of the `cfit` object `fun` as a row vector.

Examples
```
f = fittype('(x-a)^n + b','problem',{'a','b'});
coeffnames(f)
ans =
    'n'
probnames(f)
ans =
    'a'
    'b'

load census

c = fit(cdate,pop,f,'problem',{cdate(1),pop(1)},...
        'StartPoint',2);
coeffvalues(c)
ans =
    0.9877
probvalues(c)
ans =
  1.0e+003 *
    1.7900    0.0039
```

See Also   fit | fittype | probnames

quad2d

Purpose	Numerically integrate sfit object
Syntax	Q = quad2d(FO, a, b, c, d) [Q,ERRBND] = quad2d(...) [Q,ERRBND] = QUAD2D(FO,a,b,c,d,PARAM1,VAL1,PARAM2,VAL2,...)
Description	Q = quad2d(FO, a, b, c, d) approximates the integral of the surface fit object FO over the planar region $a \le x \le b$ and $c(x) \le y \le d(x)$. The bounds c and d can each be a scalar, a function handle, or a curve fit (cfit) object. [Q,ERRBND] = quad2d(...) also returns an approximate upper bound on the absolute error, ERRBND. [Q,ERRBND] = QUAD2D(FO,a,b,c,d,PARAM1,VAL1,PARAM2,VAL2,...) performs the integration with specified values of optional parameters. See the MATLAB function quad2d for details of the upper bound and the optional parameters.
See Also	quad2d | fit | sfit | cfit

Purpose   Put together rational spline

Syntax
```
rp = rpmak(breaks,coefs)
rp = rpmak(breaks,coefs,d)
rpmak(breaks,coefs,sizec)
rs = rsmak(knots,coefs)
rs = rsmak(shape,parameters)
```

Description   Both `rpmak` and `rsmak` put together a rational spline from minimal information. `rsmak` is also equipped to provide rational splines that describe standard geometric shapes. A rational spline must be scalar- or vector-valued.

`rp = rpmak(breaks,coefs)` has the same effect as the command `ppmak(breaks, coefs)` except that the resulting ppform is tagged as a rational spline, i.e., as a rpform.

To describe what this means, let R be the piecewise-polynomial put together by the command `ppmak(breaks,coefs)`, and let $r(x) = s(x)/w(x)$ be the rational spline put together by the command `rpmak(breaks,coefs)`. If `v` is the value of R at x, then `v(1:end-1)/v(end)` is the value of r at x. In other words, $R(x) = [s(x);w(x)]$. Correspondingly, the dimension of the target of r is one less than the dimension of the target of R. In particular, the dimension (of the target) of R must be at least 2, i.e., the coefficients specified by `coefs` must be `d`-vectors with `d > 1`. See `ppmak` for how the input arrays `breaks` and `coefs` are being interpreted, hence how they are to be specified in order to produce a particular piecewise-polynomial.

`rp = rpmak(breaks,coefs,d)` has the same effect as `ppmak(breaks,coefs,d+1)`, except that the resulting ppform is tagged as being a rpform. Note that the desire to have that optional third argument specify the dimension of the target requires different values for it in `rpmak` and `ppmak` for the same coefficient array `coefs`.

`rpmak(breaks,coefs,sizec)` has the same effect as `ppmak(breaks,coefs,sizec)` except that the resulting ppform is tagged as being a rpform, and the target dimension is taken to be `sizec(1)-1`.

`rs = rsmak(knots,coefs)` is similarly related to `spmak(knots,coefs)`, and `rsmak(knots,coefs,sizec)` to `spmak(knots,coefs,sizec)`. In particular, `rsmak(knots,coefs)` puts together a rational spline in B-form, i.e., it provides a rBform. See `spmak` for how the input arrays `knots` and `coefs` are being interpreted, hence how they are to be specified in order to produce a particular piecewise-polynomial.

`rs = rsmak(shape,parameters)` provides a rational spline in rBform that describes the shape being specified by the string `shape` and the optional additional `parameters`. Specific choices are:

```
rsmak('arc',radius,center,[alpha,beta])
rsmak('circle',radius,center)
rsmak('cone',radius,halfheight)
rsmak('cylinder',radius,height)
rsmak('southcap',radius,center)
rsmak('torus',radius,ratio)
```

with `1` the default value for `radius`, `halfheight` and `height`, and the origin the default for `center`, and the arc running through all the angles from `alpha` to `beta` (default is `[-pi/2,pi/2]`), and the cone, cylinder, and torus centered at the origin with their major circle in the (x,y)-plane, and the minor circle of the torus having radius `radius*ratio`, the default for `ratio` being `1/3`.

From these, one may generate related shapes by affine transformations, with the help of `fncmb(rs,transformation)`.

All `fn...` commands except `fnint`, `fnder`, `fndir` can handle rational splines.

Examples   The commands

```
runges = rsmak([-5 -5 -5 5 5 5],[1 1 1; 26 -24 26]);
rungep = rpmak([-5 5],[0 0 1; 1 -10 26],1);
```

both provide a description of the rational polynomial $r(x) = 1/(x^2 + 1)$ on the interval [-5 .. 5]. However, outside the interval [-5 .. 5], the function

given by runges is zero, while the rational spline given by rungep agrees with $1/(x^2 + 1)$ for every x.

The figure of a rotated cone is generated by the commands

```
fnplt(fncmb(rsmak('cone',1,2),[0 0 -1;0 1 0;1 0 0]))
axis equal, axis off, shading interp
```


A Rotated Cone Given by a Rational Quadratic Spline

A Helix on page 13-166, showing a helix with several windings, is generated by the commands

```
arc = rsmak('arc',2,[1;-1],[0 7.3*pi]);
[knots,c] = fnbrk(arc,'k','c');
helix = rsmak(knots, [c(1:2,:);aveknt(knots,3).*c(3,:);
c(3,:)]);
fnplt(helix)
```

rpmak

A Helix

For further illustrated examples, see "NURBS and Other Rational Splines" on page 10-30

See Also   rsmak | fnbrk | ppmak | spmak

Purpose   Piecewise biarc Hermite interpolation

Syntax   c = rscvn(p,u)
                    c = rscvn(p)

Description   c = rscvn(p,u) returns a planar piecewise biarc curve (in quadratic
                    rBform) that passes, in order, through the given points p(:,j) and
                    is constructed in the following way (see Construction of a Biarc on
                    page 13-169). Between any two distinct points p(:,j) and p(:,j+1),
                    the curve usually consists of two circular arcs (including straight-line
                    segments) which join with tangent continuity, with the first arc starting
                    at p(:,j) and normal there to u(:,j), and the second arc ending at
                    p(:,j+1) and normal there to u(:,j+1), and with the two arcs written
                    as one whenever that is possible. Thus the curve is tangent-continuous
                    everywhere except, perhaps, at repeated points, where the curve may
                    have a corner, or when the angle, formed by the two segments ending
                    at p(:,j), is unusually small, in which case the curve may have a
                    cusp at that point.

                    p must be a real matrix, with two rows, and at least two columns,
                    and any column must be different from at least one of its neighboring
                    columns.

                    u must be a real matrix with two rows, with the same number of columns
                    as p (for two exceptions, see below), and can have no zero column.

                    c = rscvn(p) chooses the normals in the following way. For j=2:end-1,
                    u(:,j) is the average of the (normalized, right-turning) normals to the
                    vectors p(:,j)-p(:,j-1) and p(:,j+1)-p(:,j). If p(:,1)==p(:,end),
                    then both end normals are chosen as the average of the normals to
                    p(:,2)-p(:,1) and p(:,end)-p(:,end-1), thus preventing a corner in
                    the resulting closed curve. Otherwise, the end normals are so chosen
                    that there is only one arc over the first and last segment (not-a-knot
                    end condition).

                    rscvn(p,u), with u having exactly two columns, also chooses the
                    interior normals as for the case when u is absent but uses the two
                    columns of u as the end-point normals.

Examples   **Example 1.** The following code generates a description of a circle, using just four pieces. Except for a different scaling of the knot sequence, it is the same description as is supplied by `rsmak('circle',1,[1;1])`.

```
p = [1 0 -1 0 1; 0 1 0 -1 0]; c = rscvn([p(1,:)+1;p(2,:)+1],p);
```

The same circle, but using just two pieces, is provided by

```
c2 = rscvn([0,2,0; 1,1,1]);
```

Example 2. The following code plots two letters. Note that the second letter is the result of interpolation to just four points. Note also the use of translation in the plotting of the second letter.

```
p = [-1 .8 -1 1 -1 -1 -1; 3 1.75 .5 -1.25 -3 -3  3];
i = eye(2); u = i(:,[2 1 2 1 2 1 1]); B = rscvn(p,u);
S = rscvn([1 -1 1 -1; 2.5 2.5 -2.5 -2.5]);
fnplt(B), hold on, fnplt(fncmb(S,[3;0])), hold off
axis equal, axis off
```


Two Letters Composed of Circular Arcs

Example 3. The following code generates the Construction of a Biarc on page 13-169, of use in the discussion below of the biarc construction

used here. Note the use of fntlr to find the tangent to the biarc at the beginning, at the point where the two arcs join, and at the end.

```
p = [0 1;0 0];  u = [.5 -.1;-.25 .5];
plot(p(1,:),p(2,:),'k'), hold on
biarc = rscvn(p,u); breaks = fnbrk(biarc,'b');
fnplt(biarc,breaks(1:2),'b',3), fnplt(biarc,breaks(2:3),'r',3)
vd = fntlr(biarc,2,breaks);
quiver(vd(1,:),vd(2,:),vd(4,:),-vd(3,:)), hold off
```


Construction of a Biarc

Algorithms   Given two distinct points, p1 and p2, in the plane and, correspondingly, two nonzero vectors, u1 and u2, there is a one-parameter family of biarcs (i.e., a curve consisting of two arcs with common tangent at their join) starting at p1 and normal there to u1 and ending at p2 and normal there to u2. One way to parametrize this family of biarcs is by the normal direction, v, at the point q at which the two arcs join. With a nonzero v chosen, there is then exactly one choice of q, hence the entire

biarc is then determined. In the construction used in rscvn, v is chosen as the reflection, across the perpendicular to the segment from p1 to p2, of the average of the vectors u1 and u2, -- after both vectors have been so normalized that their length is 1 and that they both point to the right of the segment from p1 to p2. This choice for v seems natural in the two standard cases: (i) u2 is the reflection of u1 across the perpendicular to the segment from p1 to p2; (ii) u1 and u2 are parallel. This choice of v is validated by Biarcs as a Function of the Left Normal on page 13-170 which shows the resulting biarcs when p1, p2, and u2 = [.809;.588] are kept fixed and only the normal at p1 is allowed to vary.

Biarcs as a Function of the Left Normal

But it is impossible to have the interpolating biarc depend continuously at all four data, p1, p2, u1, u2. There has to be a discontinuity as the normal directions, u1 and u2, pass through the direction from p1 to p2. This is illustrated in Biarcs as a Function of One Endpoint on page 13-171 which shows the biarcs when one point, p1 = [0;0], and the

two normals, u1 = [1;1] and u2 = [1;-1], are held fixed and only the other point, p2, moves, on a circle around p1.

Biarcs as a Function of One Endpoint

See Also   rsmak | cscvn

rsmak

Purpose	Put together rational spline for standard geometric shapes
Syntax	rs = rsmak(shape,parameters)

Description   rs = rsmak(shape,parameters) provides a rational spline in rBform that describes the shape being specified by the string shape and the optional additional parameters. Specific choices for shape are:

```
rsmak('arc',radius,center,[alpha,beta])
rsmak('circle',radius,center)
rsmak('cone',radius,halfheight)
rsmak('cylinder',radius,height)
rsmak('southcap',radius,center)
rsmak('torus',radius,ratio)
```

with 1 the default value for radius, halfheight and height, and the origin the default for center, and the arc running through all the angles from alpha to beta (default is [-pi/2,pi/2]), and the cone, cylinder, and torus centered at the origin with their major circle in the (x,y)-plane, and the minor circle of the torus having radius radius*ratio, the default for ratio being 1/3.

From these, one may generate related shapes by affine transformations, with the help of fncmb(rs,transformation).

See rpmak for more information on other options.

See Also   rpmak

Purpose   Assign values in fit options structure

Syntax   set(options)
                     s = set(options)
                     set(options,*fld1*,val1,*fld2*,val2,...)
                     set(options,*flds*,vals)

Description   set(options) displays all property names of the fit options structure
                    options. If a property has a finite list of possible string values, these
                    values are also displayed.

                    s = set(options) returns a structure s with the same property names
                    as options. If a property has a finite list of possible string values, the
                    value of the property in s is a cell array containing the possible string
                    values. If a property does not have a finite list of possible string values,
                    the value of the property in s is an empty cell array.

                    set(options,*fld1*,val1,*fld2*,val2,...) sets the properties specified
                    by the strings *fld1*, *fld2*, ... to the values val1, val2, ..., respectively.

                    set(options,*flds*,vals) sets the properties specified by the cell array
                    of strings *flds* to the corresponding values in the cell array vals.

Examples   Create a custom nonlinear model, and create a default fit options
                    structure for the model:

                    ```
                    f = fittype('a*x^2+b*exp(n*c*x)','problem','n');
                    options = fitoptions(f);
                    ```

                    Set the Robust and Normalize properties of the fit options structure
                    using property name/value pairs:

                    ```
                    set(options,'Robust','LAR','Normalize','On')
                    ```

                    Set the Display, Lower, and Algorithm properties of the fit options
                    structure using cell arrays of property names/values:

                    ```
                    set(opts,{'Disp','Low','Alg'},...
                            {'Final',[O O O],'Levenberg'})
                    ```

See Also   fitoptions | get

Purpose   Set model fit options

Syntax   FT = setoptions(FT, options)

Description   FT = setoptions(FT, options) sets the fit options of FT to options, where FT is a fittype, cfit, or sfit object. The FT output argument must match the FT input argument.

See Also   fitoptions | fit | fittype

sfit

Purpose   Constructor for `sfit` object

Syntax   `surfacefit = sfit(fittype,coeff1,coeff2,...)`

Description An `sfit` object encapsulates the result of fitting a surface to data. They are normally constructed by calling the `fit` function, or interactively by exporting a fit from the Curve Fitting Tool to the workspace. You can get and set coefficient properties of the `sfit` object.

You can treat an `sfit` object as a function to make predictions or evaluate the surface at values of X and Y.

Like the `cfit` class, `sfit` inherits all `fittype` methods.

`surfacefit = sfit(fittype,coeff1,coeff2,...)` constructs the `sfit` object `surfacefit` using the model type specified by the `fittype` object and the coefficient values `coeff1`, `coeff2`, etc.

Note `sfit` is called by the `fit` function when fitting `fittype` objects to data. To create a `sfit` object that is the result of a regression, use `fit`.

You should only call `sfit` directly if you want to assign values to coefficients and problem parameters of a `fittype` object *without* performing a fit.

Methods of `sfit` objects:

argnames	Input argument names of `cfit`, `sfit`, or `fittype` object
category	Category of fit of `cfit`, `sfit`, or `fittype` object
coeffnames	Coefficient names of `cfit`, `sfit`, or `fittype` object

coeffvalues	Coefficient values of `cfit` or `sfit`, object
confint	Confidence intervals for fit coefficients of `cfit` or `sfit` object
dependnames	Dependent variable of `cfit`, `sfit`, or `fittype` object
differentiate	Differentiate `cfit` or `sfit` object
feval	Evaluate `cfit`, `sfit`, or `fittype` object
formula	Formula of `cfit`, `sfit`, or `fittype` object
indepnames	Independent variable of `cfit`, `sfit`, or `fittype` object
islinear	Determine if `cfit`, `sfit`, or `fittype` object is linear
numargs	Number of input arguments of `cfit`, `sfit`, or `fittype` object
numcoeffs	Number of coefficients of `cfit`, `sfit`, or `fittype` object
plot	Plot `cfit` or `sfit` object
predint	Prediction intervals for `cfit` or `sfit` object
probnames	Problem-dependent parameter names of `cfit`, `sfit`, or `fittype` object
probvalues	Problem-dependent parameter values of `cfit` or `sfit` object
quad2d	Numerically integrate `sfit` object
setoptions	Set model fit options

sfit	Constructor for sfit object
type	Name of cfit, sfit, or fittype object

Examples   You can treat an sfit object as a function to make predictions or evaluate the surface at values of X and Y, e.g.,

```
x = 3 - 6 * rand( 49, 1 );
y = 3 - 6 * rand( 49, 1 );
z = peaks( x, y );
sf = fit( [x, y], z, 'poly32' );
zhat = sf( mean( x ), mean( y ) )
```

See Also   fit | fittype | feval | cfit

Purpose   Open Curve Fitting Tool

Syntax   sftool
                sftool(x,y,z)
                sftool(x,y,z,w)
                sftool(*filename*)

Description

> **Note** sftool will be removed in a future release. Use cftool instead.

sftool opens Curve Fitting Tool or brings focus to the tool if it is already open.

sftool(x,y,z) creates a fit to x and y inputs (or predictor data) and z output (or response data). sftool opens Curve Fitting Tool if necessary.

x, y, and z must be numeric, have two or more elements, and have compatible sizes. Sizes are compatible if either:

- x, y, and z all have the same number of elements, or

- x and y are vectors, z is a 2D matrix, where length(x) = n, length(y) = m, and [m,n] = size(z).

sftool(x,y,z,w) creates a fit with weights w. w must be numeric and have the same number of elements as z.

sftool(*filename*) loads the surface fitting session in *filename* into Curve Fitting Tool. The *filename* should have the extension .sfit.

Infs, NaNs, and imaginary parts of complex numbers are ignored in the data.

Curve Fitting Tool provides a flexible and intuitive graphical user interface where you can interactively fit curves and surfaces to data and view plots. You can:

- Create, plot, and compare multiple fits

- Use linear or nonlinear regression, interpolation, local smoothing regression, or custom equations

sftool

- View goodness-of-fit statistics, display confidence intervals and residuals, remove outliers and assess fits with validation data

- Automatically generate code for fitting and plotting surfaces, or export fits to workspace for further analysis

See Also   cftool

How To   · "Interactive Curve and Surface Fitting" on page 2-2

Purpose   Smooth response data

Syntax
```
yy = smooth(y)
yy = smooth(y,span)
yy = smooth(y,method)
yy = smooth(y,span,method)
yy = smooth(y,'sgolay',degree)
yy = smooth(y,span,'sgolay',degree)
yy = smooth(x,y,...)
```

Description   yy = smooth(y) smooths the data in the column vector y using a
                  moving average filter. Results are returned in the column vector yy.
                  The default span for the moving average is 5.

                  The first few elements of yy are given by

```
yy(1) = y(1)
yy(2) = (y(1) + y(2) + y(3))/3
yy(3) = (y(1) + y(2) + y(3) + y(4) + y(5))/5
yy(4) = (y(2) + y(3) + y(4) + y(5) + y(6))/5
...
```

                  Because of the way endpoints are handled, the result differs from the
                  result returned by the filter function.

                  yy = smooth(y,span) sets the span of the moving average to span.
                  span must be odd.

                  yy = smooth(y,*method*) smooths the data in y using the method
                  method and the default span. Supported values for *method* are listed
                  in the table below.

method	Description
'moving'	Moving average (default). A lowpass filter with filter coefficients equal to the reciprocal of the span.
'lowess'	Local regression using weighted linear least squares and a 1st degree polynomial model
'loess'	Local regression using weighted linear least squares and a 2nd degree polynomial model
'sgolay'	Savitzky-Golay filter. A generalized moving average with filter coefficients determined by an unweighted linear least-squares regression and a polynomial model of specified degree (default is 2). The method can accept nonuniform predictor data.
'rlowess'	A robust version of 'lowess' that assigns lower weight to outliers in the regression. The method assigns zero weight to data outside six mean absolute deviations.
'rloess'	A robust version of 'loess' that assigns lower weight to outliers in the regression. The method assigns zero weight to data outside six mean absolute deviations.

yy = smooth(y,span,*method*) sets the span of *method* to span. For the loess and lowess methods, span is a percentage of the total number of data points, less than or equal to 1. For the moving average and Savitzky-Golay methods, span must be odd (an even span is automatically reduced by 1).

yy = smooth(y,'sgolay',degree) uses the Savitzky-Golay method with polynomial degree specified by degree.

yy = smooth(y,span,'sgolay',degree) uses the number of data points specified by span in the Savitzky-Golay calculation. span must be odd and degree must be less than span.

yy = smooth(x,y,...) additionally specifies x data. If x is not provided, methods that require x data assume x = 1:length(y). You should specify x data when it is not uniformly spaced or sorted. If x is not uniform and you do not specify method, lowess is used. If the smoothing method requires x to be sorted, the sorting occurs automatically.

Tips   Another way to generate smoothed data is to fit it with a smoothing spline. Refer to the fit function for more information.

Examples   Load the data in count.dat:

```
load count.dat
```

The 24-by-3 array count contains traffic counts at three intersections for each hour of the day.

First, use a moving average filter with a 5-hour span to smooth all of the data at once (by linear index) :

```
c = smooth(count(:));
C1 = reshape(c,24,3);
```

Plot the original data and the smoothed data:

```
subplot(3,1,1)
plot(count,':');
hold on
plot(C1,'-');
title('Smooth C1 (All Data)')
```

Second, use the same filter to smooth each column of the data separately:

```
C2 = zeros(24,3);
for I = 1:3,
    C2(:,I) = smooth(count(:,I));
end
```

smooth

Again, plot the original data and the smoothed data:

```
subplot(3,1,2)
plot(count,':');
hold on
plot(C2,'-');
title('Smooth C2 (Each Column)')
```

Plot the difference between the two smoothed data sets:

```
subplot(3,1,3)
plot(C2 - C1,'o-')
title('Difference C2 - C1')
```


Note the additional end effects from the 3-column smooth.

Examples   Create noisy data with outliers:

```
x = 15*rand(150,1);
y = sin(x) + 0.5*(rand(size(x))-0.5);
```

```
y(ceil(length(x)*rand(2,1))) = 3;
```

Smooth the data using the `loess` and `rloess` methods with a span of 10%:

```
yy1 = smooth(x,y,0.1,'loess');
yy2 = smooth(x,y,0.1,'rloess');
```

Plot original data and the smoothed data.

```
[xx,ind] = sort(x);
subplot(2,1,1)
plot(xx,y(ind),'b.',xx,yy1(ind),'r-')
set(gca,'YLim',[-1.5 3.5])
legend('Original Data','Smoothed Data Using ''loess''',...
       'Location','NW')
subplot(2,1,2)
plot(xx,y(ind),'b.',xx,yy2(ind),'r-')
set(gca,'YLim',[-1.5 3.5])
legend('Original Data','Smoothed Data Using ''rloess''',...
       'Location','NW')
```

Note that the outliers have less influence on the robust method.

See Also   fit | sort

Purpose	Solve almost block-diagonal linear system
Syntax	`x = slvblk(blokmat,b)` `x = slvblk(blockmat,b,w)`
Description	`x = slvblk(blokmat,b)` returns the solution (if any) of the linear system `Ax = b`, with the matrix `A` stored in `blokmat` in the spline almost block-diagonal form. At present, only the command `spcol` provides such a description, of the matrix whose typical entry is the value of some derivative (including the 0th derivative, i.e., the value) of a B-spline at some site. If the linear system is overdetermined (i.e., has more equations than unknowns but is of full rank), then the least-squares solution is returned.
	The right side `b` may contain several columns, and is expected to contain as many rows as there are rows in the matrix described by `blokmat`.
	`x = slvblk(blockmat,b,w)` returns the vector `x` that minimizes the *weighted* sum $\Sigma_j w(j)((Ax - b)(j))^2$.
Examples	`sp=spmak(knots,slvblk(spcol(knots,k,x,1),y.'))` provides in `sp` the B-form of the spline s of order `k` with knot sequence `knots` that matches the given data `(x,y)`, i.e., for which `s(x)` equals `y`.
Algorithms	The command `bkbrk` is used to obtain the essential parts of the coefficient matrix described by `blokmat` (in one of two available forms).
	A QR factorization is made of each diagonal block, after it was augmented by the equations not dealt with when factoring the preceding block. The resulting factorization is then used to solve the linear system by backsubstitution.
See Also	`bkbrk` | `spap2` | `spapi` | `spcol`

sorted

Purpose Locate sites with respect to mesh sites

Syntax pointer = sorted(meshsites,sites)

Description Various commands in this toolbox need to determine the index j for which a given x lies in the interval $[t_j .. t_{j+1}]$, with (t_i) a given nondecreasing sequence, e.g., a knot sequence. This job is done by sorted in the following fashion.

pointer = sorted(meshsites,sites) is the integer row vector whose j-th entry equals the number of entries in meshsites that are $\leq$ ssites(j), with ssites the vector sort(sites). Thus, if both meshsites and sites are nondecreasing, then

meshsites(pointer(j))   sites(j) < meshsites(pointer(j)+1)

with the obvious interpretations when

pointer(j) < 1   or   length(meshsites) < pointer(j) + 1

Specifically, having pointer(j) < 1 then corresponds to having sites(j) strictly to the left of meshsites(1), while having length(meshsites) < pointer(j)+1 then corresponds to having sites(j) at, or to the right of, meshsites(end).

Examples The statement

sorted([1 1 1 2 2 3 3 3],[0:4])

will generate the output 0 3 5 8 8, as will the statement

sorted([3 2 1 1 3 2 3 1],[2 3 0 4 1])

Algorithms The indexing output from sort([meshsites(:).',sites(:).']) is used.

Purpose　　Least-squares spline approximation

Syntax　　　spap2(knots,k,x,y)
spap2(l,k,x,y)
sp = spap2(...,x,y,w)
spap2({knorl1,...,knorlm},k,{x1,...,xm},y)
spap2({knorl1,...,knorlm},k,{x1,...,xm},y,w)

Description　spap2(knots,k,x,y) returns the B-form of the spline *f* of order k with
the given knot sequence knots for which

(*)　　y(:,j) = *f*(x(j)), all j

in the weighted mean-square sense, meaning that the sum

$$\sum_j w(j)\,|\,y(:,j) - f(x(j))\,|^2$$

is minimized, with default weights equal to 1. The data values y(:,j)
may be scalars, vectors, matrices, even ND-arrays, and $|z|^2$ stands for
the sum of the squares of all the entries of *z*. Data points with the same
site are replaced by their average.

If the sites x satisfy the (Schoenberg-Whitney) conditions

$$\text{knots}(j) < x(j) < \text{knots}(j+k)$$

(**)　　　　　　　　$j = 1,...,\text{length}(x) = \text{length}(\text{knots}) - k$

then there is a unique spline (of the given order and knot sequence)
satisfying (*) exactly. No spline is returned unless (**) is satisfied for
some subsequence of x.

spap2(l,k,x,y), with l a positive integer, returns the B-form of a
least-squares spline approximant, but with the knot sequence chosen
for you. The knot sequence is obtained by applying aptknt to an
appropriate subsequence of x. The resulting piecewise-polynomial
consists of l polynomial pieces and has k-2 continuous derivatives.

If you feel that a different distribution of the interior knots might do a better job, follow this up with

```
sp1 = spap2(newknt(sp),k,x,y));
```

`sp = spap2(...,x,y,w)` lets you specify the weights `w` in the error measure (given above). `w` must be a vector of the same size as `x`, with nonnegative entries. All the weights corresponding to data points with the same site are summed when those data points are replaced by their average.

`spap2({knorl1,...,knorlm},k,{x1,...,xm},y)` provides a least-squares spline approximation to *gridded* data. Here, each `knorli` is either a knot sequence or a positive integer. Further, `k` must be an m-vector, and `y` must be an (r+m)-dimensional array, with `y(:,i1,...,im)` the datum to be fitted at the `site [x{1}(i1),...,x{m}(im)]`, all i1, ..., im. However, if the spline is to be scalar-valued, then, in contrast to the univariate case, `y` is permitted to be an m-dimensional array, in which case `y(i1,...,im)` is the datum to be fitted at the `site [x{1}(i1),...,x{m}(im)]`, all i1, ..., im.

`spap2({knorl1,...,knorlm},k,{x1,...,xm},y,w)` also lets you specify the weights. In this m-variate case, `w` must be a cell array with m entries, with `w{i}` a nonnegative vector of the same size as `xi`, or else `w{i}` must be empty, in which case the default weights are used in the `ith` variable.

Examples   `sp = spap2(augknt([a,xi,b],4),4,x,y)`

is the least-squares approximant to the data `x`, `y`, by cubic splines with two continuous derivatives, basic interval [a..b], and interior breaks `xi`, provided `xi` has all its entries in (`a..b`) and the conditions (**) are satisfied in some fashion. In that case, the approximant consists of `length(xi)+1` polynomial pieces. If you do not want to worry about the conditions (**) but merely want to get a cubic spline approximant consisting of `l` polynomial pieces, use instead

```
sp = spap2(l,4,x,y);
```

If the resulting approximation is not satisfactory, try using a larger l. Else use

```
sp = spap2(newknt(sp),4,x,y);
```

for a possibly better distribution of the knot sequence. In fact, if that helps, repeating it may help even more.

As another example, `spap2(1, 2, x, y);` provides the least-squares straight-line fit to data x,y, while

```
w = ones(size(x)); w([1 end]) = 100; spap2(1,2, x,y,w);
```

forces that fit to come very close to the first data point and to the last.

Algorithms   spcol is called on to provide the almost block-diagonal collocation matrix $(B_{j,k}(x_i))$, and slvblk solves the linear system (*) in the (weighted) least-squares sense, using a block QR factorization.

Gridded data are fitted, in tensor-product fashion, one variable at a time, taking advantage of the fact that a univariate weighted least-squares fit depends linearly on the values being fitted.

See Also   slvblk | spapi | spcol

spapi

Purpose	Spline interpolation

Syntax

```
spline = spapi(knots,x,y)
spapi(k,x,y)
spapi({knork1,...,knorkm},{x1,...,xm},y)
spapi(...,'noderiv')
```

Description   `spline = spapi(knots,x,y)` returns the spline f (if any) of order

```
k = length(knots) - length(x)
```

with knot sequence `knots` for which

$$(*)\quad f(x(j)) = y(:,j),\ \text{all }j.$$

If some of the entries of `x` are the same, then this is taken in the osculatory sense, i.e., in the sense that $D^{m(j)}f(x(j)) = y(:, j)$, with `m(j)` : = #{ i < j : x(i) = x(j) }, and $D^m f$ the mth derivative of f. Thus r-fold repetition of a site z in `x` corresponds to the prescribing of value and the first $r - 1$ derivatives of f at z. If you don't want this, call `spapi` with an additional (fourth) argument, in which case, at each data site, the average of all data values with the same data site is matched.

The data values, `y(:,j)`, may be scalars, vectors, matrices, or even ND-arrays.

`spapi(k,x,y)` , with k a positive integer, merely specifies the desired spline order, k, in which case `aptknt` is used to determine a workable (though not necessarily optimal) knot sequence for the given sites `x`. In other words, the command `spapi(k,x,y)` has the same effect as the more explicit command `spapi(aptknt(x,k),x,y)`.

`spapi({knork1,...,knorkm},{x1,...,xm},y)` returns the B-form of a tensor-product spline interpolant to *gridded* data. Here, each `knorki` is either a knot sequence, or else is a positive integer specifying the polynomial order to be used in the ith variable, thus leaving it to `spapi` to provide a corresponding knot sequence for the ith variable. Further, y must be an (r+m)-dimensional array, with `y(:,i1,...,im)` the datum to be fitted at the site `[x{1}(i1),...,x{m}(im)]`, all i1, ..., im ,

unless the spline is to be scalar-valued, in which case, in contrast to the univariate case, y is permitted to be an m-dimensional array.

spapi(...,'noderiv') with the string 'noderiv' as a fourth argument, has the same effect as spapi(...) except that data values sharing the same site are interpreted differently. With the fourth argument present, the average of the data values with the same data site is interpolated at such a site. Without it, data values with the same data site are interpreted as values of successive derivatives to be matched at such a site, as described above, in the first paragraph of this Description.

Examples   spapi([0 0 0 0 1 2 2 2 2],[0 1 1 1 2],[2 0 1 2 -1]) produces the unique cubic spline *f* on the interval [0..2] with exactly one interior knot, at 1, that satisfies the five conditions

$$f(0+) = 2, f(1) = 0, Df(1) = 1, D^2f(1) = 2, f(2-) = -1$$

These include 3-fold matching at 1, i.e., matching there to prescribed values of the function and its first two derivatives.

Here is an example of osculatory interpolation, to values y and slopes s at the sites x by a quintic spline:

```
sp = spapi(augknt(x,6,2),[x,x,min(x),max(x)],[y,s,ddy0,ddy1]);
```

with ddy0 and ddy1 values for the second derivative at the endpoints.

As a related example, if the function sin(x) is to be interpolated at the distinct data sites x by a cubic spline, and its slope is also to be matched at a subsequence x(s), then this can be accomplished by the command

```
sp = spapi(4,[x x(s)], [sin(x) cos(x(s))]);
```

in which a suitable knot sequence is supplied with the aid of aptknt. In fact, if you wanted to interpolate the same data by quintic splines, simply change the 4 to 6.

As a bivariate example, here is a bivariate interpolant.

```
x = -2:.5.2; y = -1:.25:1; [xx, yy] = ndgrid(x,y);
z = exp(-(xx.^2+yy.^2));
sp = spapi({3,4},{x,y},z); fnplt(sp)
```

As an illustration of osculatory interpolation to gridded data, here is complete bicubic interpolation, with the data explicitly derived from the bicubic polynomial $g(u,v) = u^3v^3$, to make it easy for you to see exactly where the slopes and slopes of slopes (i.e., cross derivatives) must be placed in the data values supplied. Since our g is a bicubic polynomial, its interpolant, f, must be g itself. We test this.

```
sites = {[0,1],[0,2]}; coefs = zeros(4,4); coefs(1,1) = 1;
g = ppmak(sites,coefs);
Dxg = fnval(fnder(g,[1,0]),sites);
Dyg = fnval(fnder(g,[0,1]),sites);
Dxyg = fnval(fnder(g,[1,1]),sites);
f = spapi({4,4}, {sites{1}([1,2,1,2]),sites{2}([1,2,1,2])}, ...
          [fnval(g,sites), Dyg ; ...
           Dxg.'          , Dxyg]);
if any( squeeze( fnbrk(fn2fm(f,'pp'), 'c') ) - coefs )
'something went wrong', end
```

Algorithms spcol is called on to provide the almost-block-diagonal collocation matrix $(B_{j,k}(x))$ (with repeats in x denoting derivatives, as described above), and slvblk solves the linear system (*), using a block QR factorization.

Gridded data are fitted, in tensor-product fashion, one variable at a time, taking advantage of the fact that a univariate spline fit depends linearly on the values being fitted.

Limitations The given (univariate) knots and sites must satisfy the Schoenberg-Whitney conditions for the interpolant to be defined. Assuming the site sequence x to be nondecreasing, this means that we must have

$$\text{knots}(j) < x(j) < \text{knots}(j+k), \text{ all } j$$

(with equality possible at knots(1) and knots(end)). In the multivariate case, these conditions must hold in each variable separately.

See Also   csapi | spap2 | spaps | spline

spaps

Purpose   Smoothing spline

Syntax
```
sp = spaps(x,y,tol)
[sp,values] = spaps(x,y,tol)
[sp,values,rho] = spaps(x,y,tol)
[...] = spaps(x,y,tol,arg1,arg2,...)
[...] = spaps({x1,...,xr},y,tol,...)
```

Description   `sp = spaps(x,y,tol)` returns the B-form of the smoothest function *f* that lies within the given tolerance `tol` of the given data points `(x(j), y(:,j))`, `j=1:length(x)`. The data values `y(:,j)` may be scalars, vectors, matrices, even ND-arrays. Data points with the same data site are replaced by their weighted average, with its weight the sum of the corresponding weights, and the tolerance `tol` is reduced accordingly.

`[sp,values] = spaps(x,y,tol)` also returns the smoothed values, i.e., `values` is the same as `fnval(sp,x)`.

Here, the distance of the function *f* from the given data is measured by

$$E(f) = \sum_{j=1}^{n} w(j) \,|\, (y(:,j) - f(x(j))) \,|^2$$

with the default choice for the weights `w` making $E(f)$ the composite trapezoidal rule approximation to $\int_{\min(x)}^{\max(x)} | \, y - f \, |^2$, and $|z|^2$ denoting the sum of squares of the entries of *z*.

Further, *smoothest* means that the following roughness measure is minimized:

$$F(D^m f) = \int_{\min(x)}^{\max(x)} \lambda(t) \big| D^m f(t) \big|^2 \, dt$$

where $D^m f$ denotes the mth derivative of f. The default value for m is 2, the default value for the roughness measure weight λ is the constant 1, and this makes f a cubic smoothing spline.

When tol is nonnegative, then the spline f is determined as the unique minimizer of the expression $\rho E(f) + F(D^m f)$, with the smoothing parameter ρ (optionally returned) so chosen that $E(f)$ equals tol. Hence, when m is 2, then, after conversion to ppform, the result should be the same (up to roundoff) as obtained by csaps($x,y,\rho/(\rho + 1)$). Further, when tol is zero, then the "natural" or variational spline interpolant of order $2m$ is returned. For large enough tol, the least-squares approximation to the data by polynomials of degree <m is returned.

When tol is negative, then ρ is taken to be -tol.

The default value for the weight function λ in the roughness measure is the constant function 1. But you may choose it to be, more generally, a piecewise constant function, with breaks only at the data sites. Assuming the vector x to be strictly increasing, you specify such a piecewise constant λ by inputting tol as a vector of the same size as x. In that case, tol(i) is taken as the constant value of λ on the interval (x(i-1) .. x(i)), i=2:length(x), while tol(1) continues to be used as the specified tolerance.

[sp,values,rho] = spaps(x,y,tol) also returns the actual value of ρ used as the third output argument.

[...] = spaps(x,y,tol,arg1,arg2,...) lets you specify the weight vector w and/or the integer m, by supplying them as an argi. For this, w must be a nonnegative vector of the same size as x; m must be 1 (for a piecewise linear smoothing spline), or 2 (for the default cubic smoothing spline), or 3 (for a quintic smoothing spline).

If the resulting smoothing spline, sp, is to be evaluated outside its basic interval, it should be replaced by fnxtr(sp,m) to ensure that its m-th derivative is zero outside that interval.

[...] = spaps({x1,...,xr},y,tol,...) returns the B-form of an r-variate tensor-product smoothing spline that is roughly within the specified tolerance to the given *gridded*

spaps

data. (For *scattered* data, use `tpaps`.) Now `y` is expected to supply the corresponding gridded values, with `size(y)` equal to `[length(x1),...,length(xr)]` in case the function is scalar-valued, and equal to `[d,length(x1),...,length(xr)]` in case the function is `d`-valued. Further, `tol` must be a cell array with `r` entries, with `tol{i}` the tolerance used during the `i`-th step when a univariate (but vector-valued) smoothing spline in the `i`-th variable is being constructed. The optional input for `m` must be an `r`-vector (with entries from the set `{1,2,3}`), and the optional input for `w` must be a cell array of length `r`, with `w{i}` either empty (to indicate that the default choice is wanted) or else a positive vector of the same length as `xi`.

Examples   The statements

```
w = ones(size(x)); w([1 end]) = 100;
sp = spaps(x,y, 1.e-2, w, 3);
```

give a quintic smoothing spline approximation to the given data that close to interpolates the first and last datum, while being within about `1.e-2` of the rest.

```
x = -2:.2:2; y=-1:.25:1; [xx,yy] = ndgrid(x,y); rand('seed',39);
z = exp(-(xx.^2+yy.^2)) + (rand(size(xx))-.5)/30;
sp = spaps({x,y},z,8/(60^2));  fnplt(sp), axis off
```

produces the figure below, showing a smooth approximant to noisy data from a smooth bivariate function. Note the use of `ndgrid` here; use of `meshgrid` would have led to an error.

Algorithms   Reinsch's approach is used (including his clever way of choosing the equation for the optimal smoothing parameter in such a way that a good initial guess is available and Newton's method is guaranteed to converge and to converge fast).

References   [1] C. Reinsch, "Smoothing by spline functions", *Numer. Math.* 10 (1967), 177–183.

See Also   csaps | spap2 | spapi | tpaps

spcol

Purpose   B-spline collocation matrix

Syntax
```
colmat = spcol(knots,k,tau)
colmat = spcol(knots,k,tau,arg1,arg2,...)
```

Description   `colmat = spcol(knots,k,tau)` returns the matrix, with `length(tau)` rows and `length(knots)-k` columns, whose (*i,j*)th entry is

$$D^{m(i)}B_j(\text{tau}(i))$$

This is the value at tau(*i*) of the *m*(*i*)th derivative of the *j*th B-spline of order `k` for the knot sequence `knots`. Here, `tau` is a sequence of sites, assumed to be *nondecreasing*, and *m* = knt2mlt(tau), i.e., *m*(*i*) is #{*j* < *i*:tau(*j*) = tau(*i*)}, all *i*.

`colmat = spcol(knots,k,tau,arg1,arg2,...)` also returns that matrix, but gives you the opportunity to specify some aspects.

If one of the `argi` is a string with the same first two letters as in `'slvblk'`, the matrix is returned in the almost block-diagonal format (specialized for splines) required by `slvblk` (and understood by `bkbrk`).

If one of the `argi` is a string with the same first two letters as in `'sparse'`, then the matrix is returned in the `sparse` format of MATLAB.

If one of the `argi` is a string with the same first two letters as in `'noderiv'`, multiplicities are ignored, i.e., *m*(*i*) is taken to be 1 for all *i*.

Examples   To solve approximately the non-standard second-order ODE

$$D^2y(t) = 5 \cdot (y(t) - \sin(2t))$$

on the interval [0..π], using cubic splines with 10 polynomial pieces, you can use `spcol` in the following way:

```
tau = linspace(0,pi,101); k = 4;
knots = augknt(linspace(0,pi,11),k);
colmat = spcol(knots,k,brk2knt(tau,3));
```

```
coefs = (colmat(3:3:end,:)/5-colmat(1:3:end,:))\(-sin(2*tau).');
sp = spmak(knots,coefs.');
```

You can check how well this spline satisfies the ODE by computing and plotting the residual, $D^2y(t) - 5 \cdot (y(t) - \sin(2t))$, on a fine mesh:

```
t = linspace(0,pi,501);
yt = fnval(sp,t);
D2yt = fnval(fnder(sp,2),t);
plot(t,D2yt - 5*(yt-sin(2*t)))
title(['residual error; to be compared to max(abs(D^2y)) = ',...
        num2str(max(abs(D2yt)))])
```

The statement `spcol([1:6],3,.1+[2:4])` provides the matrix

```
ans =

    0.5900    0.0050         0
    0.4050    0.5900    0.0050
         0    0.4050    0.5900
```

in which the typical row records the values at 2.1, or 3.1, or 4.1, of all B-splines of order 3 for the knot sequence `1:6`. There are three such B-splines. The first one has knots 1,2,3,4, and its values are recorded in the first column. In particular, the last entry in the first column is zero since it gives the value of that B-spline at 4.1, a site to the right of its last knot.

If you add the string `'sl'` as an additional input to `spcol`, then you can ask `bkbrk` to extract detailed information about the block structure of the matrix encoded in the resulting output from `spcol`. Thus, the statement `bkbrk(spcol(1:6,3,.1+2:4,'sl'))` gives:

```
block 1 has 2 row(s)
    0.5900    0.0050         0
    0.4050    0.5900    0.0050
next block is shifted over 1 column(s)
block 2 has 1 row(s)
```

spcol

```
      0.4050   0.5900   0.0050
next block is shifted over 2 column(s)
```

Algorithms   This is the most complex command in this toolbox since it has to deal with various ordering and blocking issues. The recurrence relations are used to generate, simultaneously, the values of all B-splines of order k having anyone of the `tau(i)` in their support.

A separate calculation is carried out for the (presumably few) sites at which derivative values are required. These are the sites `tau(i)` with $m(i) > 0$. For these, and for every order $k - j$, $j = j_0, j_0 - 1,...,0$, with j_0 equal to max(m), values of all B-splines of that order are generated by recurrence and used to compute the jth derivative at those sites of all B-splines of order k.

The resulting rows of B-spline values (each row corresponding to a particular `tau(i)`) are then assembled into the overall (usually rather sparse) matrix.

When the optional argument `'sl'` is present, these rows are instead assembled into a convenient almost block-diagonal form that takes advantage of the fact that, at any site `tau(i)`, at most k B-splines of order k are nonzero. This fact (together with the natural ordering of the B-splines) implies that the collocation matrix is almost block-diagonal, i.e., has a staircase shape, with the individual blocks or steps of varying height but of uniform width k.

The command `slvblk` is designed to take advantage of this storage-saving form available when used, in `spap2`, `spapi`, or `spaps`, to help determine the B-form of a piecewise-polynomial function from interpolation or other approximation conditions.

Limitations   The sequence `tau` is assumed to be nondecreasing.

See Also   slvblk | spap2 | spapi

Purpose   Spline curve by uniform subdivision

Syntax   spcrv(c,k)
               spcrv(c)
               spcrv(c,k,maxpnt)

Description spcrv(c,k) provides a dense sequence *f*(tt) of points on the uniform
               B-spline curve *f* of order k with B-spline coefficients c. Explicitly, this
               is the curve

$$f : t \mapsto \sum_{j=1}^{n} B(t - k/2 \mid j,...,j+k)\, c(j), \ \frac{k}{2} \le t \le n + \frac{k}{2}$$

with $B(\cdot \mid a,...,z)$ the B-spline with knots $a,...,z$, and n the number of
coefficients in c, i.e., [d,n] equals size(c).

spcrv(c) chooses the order k to be 4.

spcrv(c,k,maxpnt) makes sure that at least maxpnt points are
generated. The default value for the maximum number of sites tt to
be generated is 100.

The parameter interval that the site sequence tt fills out uniformly is
the interval [k/2 .. (n-k/2)].

The output consists of the array *f*(tt).

Examples   The following would show a questionable broken line and its smoothed
               version:

```
points = [0 0 1 1 0 -1 -1 0 0 ;
          0 0 0 1 2 1 0 -1 -2];
plot(points(1,:),points(2,:),':')
values = spcrv(points,3);
hold on, plot(values(1,:),values(2,:)), hold off
```

Algorithms Repeated midpoint knot insertion is used until there are at least maxpnt
               sites. There are situations where use of fnplt would be more efficient.

See Also   fnplt

Purpose   Experiment with some spline approximation methods

Syntax   splinetool
                   splinetool(x,y)

Description   splinetool is a graphical user interface (GUI), whose initial menu
                   provides you with various choices for data including the option of
                   importing some data from the workspace.

                   splinetool(x,y) brings up the GUI with the specified data x and y,
                   which are vectors of the same length.

Tips   The Spline Tool is shown in the following figure comparing cubic spline
                   interpolation with a smoothing spline on sample data created by adding
                   noise to the cosine function.

splinetool

Approximation Methods

The approximation methods and options supported by the GUI are shown below.

Approximation Method	Option
Cubic Interpolating Spline	Adjust the type and values of the end conditions.
Smoothing Spline	Choose between cubic (order 4) and quintic (order 6) splines. Adjust the value of the tolerance and/or smoothing parameter. Adjust the weights in the error and roughness measures.
Least-Squares Approximation	Vary the order from 1 to 14. The default order is 4, which gives cubic approximating splines. Modify the number of polynomial pieces. Add and move knots to improve the fit. Adjust the weights in the error measure.
Spline Interpolation	Vary the order from 2 to 14. The default order is 4, which gives cubic spline interpolants. If the default knots supplied are not satisfactory, you can move them around to vary the fit.

Graphs

You can generate and compare several approximations to the same data. One of the approximations is always marked as "current" using a thicker line width. The following displays are available:

- Data graph. It shows:

  - The data

  - The approximations chosen for display in **List of approximations**

  - The current knot sequence or the current break sequence

- Auxiliary graph (if viewed) for the current approximation. You can invoke this graph by selecting any one of the items in the **View** menu. It shows one of the following:

  - The first derivative

  - The second derivative

  - The error

By default, the error is the difference between the given data values and the value of the approximation at the data sites. In particular, the error is zero (up to round-off) when the approximation is an interpolant. However, if you provide the data values by specifying a function, then the error displayed is the difference between that function and the current approximation. This also happens if you change the y-label of the data graph to the name of a function.

Menu Options

You can annotate and print the graphs with the **File > Print to Figure** menu.

You can export the data and approximations to the workspace for further use or analysis with the **File > Export Data** and **File > Export Spline** menus, respectively.

You can create, with the **File > Generate Code** menu, a function file that you can use to generate, from the original data, any or all graphs currently shown. This file also provides you with a written record of the commands used to generate the current graph(s).

You can save, with the **Replicate** button, the current approximation before you experiment further. If, at a later time, you click on the approximation so saved, splinetool restores everything to the way it was, including the data used in the construction of the saved approximation. This is true even if, since saving this approximation, you have edited the data while working on other approximations.

You can add, delete, or move data, knots, and breaks by right-clicking in the graph, or selecting the appropriate item in the **Edit** menu.

You can toggle the grid or the legend in the graph(s) with the **Tools** menu.

Examples

- "Exploring End Conditions For Cubic Spline Interpolation" on page 13-209
- "Estimating the Second Derivative at an Endpoint" on page 13-212
- "Least-Squares Approximation" on page 13-214
- "Smoothing Spline" on page 13-217

Exploring End Conditions For Cubic Spline Interpolation

The purpose of this example is to explore the various end conditions available with cubic spline interpolation:

1 Type splinetool at the command line.

2 Select **Import your own data** from the initial screen, and accept the default function. You should see the following display.

The default approximation shown is the cubic spline interpolant with the not-a-knot end condition.

The vector x of data sites is linspace(0,2*pi,31) and the values are cos(x). This differs from simply providing the vector y of values in that the cosine function is explicitly recorded as the underlying function. Therefore, the error shown in the graph is the error in the spline as an approximation to the cosine rather than as an approximation to the given values. Notice the resulting relatively large error, about 5e-5, near the endpoints.

3 For comparison, follow these steps:

- Click on **New** in the **List of approximations**.

- In **Approximation method**, select **complete** from the list of **End conditions**.

- Since the first derivative of the cosine function is sine, adjust the first-derivative values to their known values of zero at both the left end and the right end.

This procedure results in the display shown below (after the mouse is used to move the Legend further down). Note that the right end slope is zero only up to round-off. **Bottomline** tells you that the toolbox function `csape` was used to create the spline.

Be impressed by the improvement in the error, which is only about `5e-6`.

4 For further comparison, follow these steps:

- Click on **New** in the **List of approximations**.

- In **Approximation method**, select **natural** from the list of **End conditions**.

Note the deterioration of the approximation near the ends, an error of about `2e-3`, which is much worse than with the not-a-knot end conditions.

5 For a final comparison, follow these steps:

- Click on **New** in the **List of approximations**.

- Since we know that the cosine function is periodic, in **Approximation method**, select **periodic** from the list of **End conditions**.

Note the dramatic improvement in the approximation, back to an error of about `5e-6`, particularly compared to the **natural** end conditions.

Estimating the Second Derivative at an Endpoint

This example uses cubic spline interpolation and least-squares approximation to determine an estimate of the initial acceleration for a drag car:

1 Type `splinetool` at the command line or if the GUI is already running, click on **File > Restart**.

2 Choose **Richard Tapia's drag racing data**. These data show the distance traveled by a drag car as a function of time. The message window asks you to estimate the initial acceleration by setting the initial speed to zero. Click on **OK,** or use **Space** or **Enter,** to remove the message window.

3 In **Approximation method**, select **complete** from the list of **End conditions**.

4 Adjust the initial speed by changing the first derivative at the left endpoint to zero.

5 Look for the value of the initial acceleration, which is given by the value of the second derivative at the left endpoint. You can toggle between the first derivative and the second derivative at this endpoint by clicking on the **left end** button. The value of the second derivative should be around 187 in the units chosen. Choose **View > Show 2nd Derivative** to see this graphically.

6 For comparison, click on **New**, then choose **Least-Squares Approximation** as the **Approximation method**. With this method, you can no longer specify end conditions. Instead, you may vary the order of the method. Verify that the initial acceleration is close to the cubic interpolation value.

The results of this procedure are shown below.

splinetool

Least-Squares Approximation

This example encourages you to place five interior knots in such a way that the least-squares approximation to these data by cubic splines has an absolute error no bigger than .04 everywhere:

1 Type splinetool at the command line or if the GUI is already running, click on **File > Restart**.

2 Choose **Titanium heat data**.

3 Select **Least-Squares Approximation** as the **Approximation method.**

4 Notice how poorly this approximates the data since there are no interior knots. To view the current knots and add new knots, choose **knots** from **Data, breaks/knots, weights**. The knots are now listed in **knots**, and also displayed in the data graph as vertical lines. Notice that there are just the two end knots, each with multiplicity 4.

5 Right-click in the data graph and choose **Add Knot**. This brings up crosshairs for you to move with the mouse. Its precise horizontal location is shown in the edit field below the list of knots. A mouse click places a new knot at the current location of the crosshairs. One possible strategy is to place the additional knot at the place of maximum absolute error, as shown in the auxiliary graph below the data graph.

If you right-click and choose **Replicate Knot**, you will increase the multiplicity of the current knot, which is shown by its repeated occurrence in **Knots**. If you don't like a particular knot, you can delete it. To delete a specific knot, you must first select it in either the list of knots or the data graph, and then right-click in the graph and choose **Delete Knot**.

6 You could also ask for an approximation using six polynomial pieces, which corresponds to five interior knots. To do this, enter 6 as **# pieces** in **Data, breaks/knots, weights**.

7 After you have the five interior knots, try to make the error even smaller by moving the knots. To do this, select the knot you want to move by clicking on its vertical line in the graph, then use the

interface control below **Knots** in **Data, breaks/knots, weights** and observe how the error changes with the movement of the knot. You can also use the edit field to overwrite the current knot location. You could also try **adjust**, which redistributes the current knot sequence.

8 Use **Replicate** in **List of approximations** to save any good knot distribution for later use. Rename the replicated approximation to lstsqr using **Rename**. To return to the original approximation, click on its name in **List of approximations**.

Smoothing Spline

This example experiments with smoothing splines:

1 Type splinetool at the command line or, if the GUI is already running, click on **File > Restart**.

2 Choose **Titanium heat data**.

3 In **Approximation method**, choose **Smoothing Spline**.

4 Vary **Parameter** between 0 and 1, which changes the approximation from the least-squares straight-line approximation to the "natural" cubic spline interpolant.

5 Vary **Tolerance** between 0 and some large value, even inf. The approximation changes from the best possible one, the "natural" cubic spline interpolant, to the least-squares straight-line approximation.

6 As you increase the **Parameter** value or decrease the **Tolerance** value, the error decreases. However, a smaller error corresponds to more roughness, as measured by the size of the second derivative. To see this, choose **View > Show 2nd Derivative** and vary the **Parameter** and **Tolerance** values once again.

7 This step modifies the weights in the error measure to force the approximation to pass through a particular data point.

- Set **Tolerance** to `0.2`. Notice that the approximation does not pass through the highest data point. To see the large error at this site, choose **View > Error**.

- To force the smoothing spline to go through this point, choose **Error Weights** from **Data, breaks/knots, weights**.

- Click on the highest data point in the graph and notice its site, which is indicated in **Sites and Values**.

- Use the edit field beneath the list of weights to change the current weight to `1000`. Notice how much closer the smoothing spline now comes to that highest data point, and the decrease in the error at that site. Turn on the grid, by **Tools > Grid**, to locate the error at that site more readily.

8 This step modifies the weights in the roughness measure to permit a more accurate but less smooth approximation in the peak area while insisting on a smoother, hence less accurate, approximation away from the peak area.

- Choose **Jumps in Roughness Weight** from **Data, breaks/knots, weights**.

- Choose **View > Show 2nd Derivative**

- Select any data point to the left of the peak in the data.

- Set the jump at the selected site to `-1` by changing its value in the edit field below it. Since the roughness weight for the very first site interval is 1, you have just set the roughness weight to the right of the highlighted site to 0. Correspondingly, the second derivative has become relatively small to the left of that site.

- Select any data point to the right of the peak in the data.

- Set the jump across the selected site to `1`. Since the roughness weight just to the left of the highlighted site is 0, you have just set the roughness weight to the right of the highlighted site to 1. Correspondingly, the second derivative has become relatively small to the right of that site. The total effect is a very smooth but not very accurate fit away from the peak, while in the peak area,

the spline fit is much better but the second derivative is much larger, as is shown in the auxiliary graph below.

At the sites where there is a jump in the roughness weight, there is a corresponding jump in the second derivative. If you increase the **Parameter** value, the error across the peak area decreases but the second derivative remains quite large, while the opposite holds true away from the peak area.

See Also   csape | csapi | csaps | spap2 | spapi | spaps

splpp, sprpp

Purpose   Taylor coefficients from local B-coefficients

Syntax   [v,b] = splpp(tx,a)
                [v,b] = sprpp(tx,a)

Description These are utility commands of use in the conversion from B-form to
                ppform (and in certain evaluations), but of no interest to the casual user.

[v,b] = splpp(tx,a) provides the matrices v and b, both of the same
size [r,k] as a, and related to the input in the following way.

For i=1:r, b(i,:) are the B-coefficients, with respect to the knot
sequence [tx(i,1:k-1),0,...,0], of the polynomial of order k on
the interval [tx(i,k-1) .. tx(i,k)] whose k B-spline coefficients,
with respect to the knot sequence tx(i,:), are in a(i,:). This is
done by repeated knot insertion (of the knot 0). It is assumed that
tx(i,k-1)<0<=tx(i,k).

For i=1:r, v(i,:) are the polynomial coefficients for that polynomial,
i.e., v(i,j) is the number $D^{k-j}s(0-)/k-j)!$, j=1:k, with *s* having the
knots tx(i,:) and the B-coefficients a(i,:).

[v,b] = sprpp(tx,a) carries out exactly the same job, except
that now b(i,:) are the B-coefficients for that polynomial with
respect to the knot sequence [0,...,0,tx(i,k: 2*(k-1))], and,
correspondingly, v(i,j) is $D^{k-j}s(0+)/k-j)!$, j=1:k. Also, now it is
assumed that tx(i,k-1)<=0<tx(i,k).

Examples   The statement [v,b]=splpp([-2 -1 0 1],[0 1 0]) provides the
                sequence

        v = -1.0000 -1.0000 0.5000 = $D^2s(0-)/2, Ds(0-), s(0-)$

with *s* the B-spline with knots -2, -1, 0, 1. This is so because the l in
splpp indicates the limit from the left, and the second argument, [0 1
0], indicates the spline *s* in question to be

$$s = 0 \times B(\cdot\,|\,[?,-2,-1,0]) + 1 \times B(\cdot\,|\,[-2,-1,0,1]) + 0 \times B(\cdot\,|\,[-1,0,1,?])$$

i.e., this particular linear combination of the third-order B-splines for the knot sequence ..., -2, -1,0,1,... (Note that the values calculated do not depend on the knots marked ?.) The above statement also provides the sequence b = 0 1.0000 0.5000 of B-spline coefficients for the polynomial piece of *s* on the interval [-1. .0], and with respect to the knot sequence ?, -2, -1, 0, 0, ?.

In other words, on the interval [-1. .0], the B-spline with knots 2, -1, 0, 1 can be written

$$0 \times B(\cdot \,|\, [?,-2,-1,0]) + 1 \times B(\cdot \,|\, [-2,-1,0,0]) + 5 \times B(\cdot \,|\, [-1,0,0,?])$$

The statement [v,b]=sprpp([-1 0 1 2],[1 0 0]) provides the sequence

v = [0.5000 -1.0000 0.5000] = $D^2 s(0+)/2, Ds(0+), s(0+)$

with *s* the B-spline with knots ?,-1,0,1. Its polynomial piece on the interval [0..1] is independent of the choice of ?, so we might as well think of ? as -2, i.e., we are dealing with the same B-spline as before. Note that the last two numbers agree with the limits from the left computed above, while the first number does not. This reflects the fact that a quadratic B-spline with simple knots is continuous with continuous first, but discontinuous second, derivative. (It also reflects the fact that the leftmost knot of a B-spline is irrelevant for its right-most polynomial piece.) The sequence b = 0.5000 0 0 also provided states that, on the interval [0. .1], the B-spline $B(\cdot \,|\, [?,-1,0,1])$ can be written

$$0.5 \times B(\cdot \,|\, [0,0,0,1]) + 0 \times B(\cdot \,|\, [0,0,1,2]) + 0 \times B(\cdot \,|\, [0,1,2,?])$$

spmak

Purpose	Put together spline in B-form

Syntax

```
spmak(knots,coefs)
spmak(knots,coefs,sizec)
spmak
sp = spmak(knots,coeffs)
```

Description   The command spmak(...) puts together a spline function in B-form, from minimal information, with the rest inferred from the input. fnbrk returns all the parts of the completed description. In this way, the actual data structure used for the storage of this form is easily modified without any effect on the various fn... commands that use this construct.

spmak(knots,coefs) returns the B-form of the spline specified by the knot information in knots and the coefficient information in coefs.

The action taken by spmak depends on whether the function is univariate or multivariate, as indicated by knots being a sequence or a cell array. For the description, let sizec be size(coefs).

If knots is a sequence (required to be non-decreasing), then the spline is taken to be univariate, and its order k is taken to be length(knots)-sizec(end). This means that each 'column' coefs(:,j) of coefs is taken to be a B-spline coefficient of the spline, hence the spline is taken to be sizec(1:end-1)-valued. The basic interval of the B-form is [knots(1) .. knots(end)].

Knot multiplicity is held to be ≤ k. This means that the coefficient coefs(:,j) is simply ignored in case the corresponding B-spline has only one distinct knot, i.e., in case knots(j) equals knots(j+k).

If knots is a cell array, of length m, then the spline is taken to be m-variate, and coefs must be an (r+m)-dimensional array, – except when the spline is to be scalar-valued, in which case, in contrast to the univariate case, coefs is permitted to be an m-dimensional array, but sizec is reset by

```
sizec = [1, sizec]; r = 1;
```

The spline is `sizec(1:r)`-valued. This means the output of the spline is an array with `r` dimensions, e.g., if `sizec(1:2) = [2, 3]` then the output of the spline is a 2-by-3 matrix.

The spline is `sizec(1:r)`-valued, the `ith` entry of the m-vector `k` is computed as `length(knots{i}) - sizec(r+i), i=1:m`, and the `ith` entry of the cell array of basic intervals is set to `[knots{i}(1), knots{i}(end)]`.

`spmak(knots,coefs,sizec)` lets you supply the intended size of the array `coefs`. Assuming that `coefs` is correctly sized, this is of concern only in the rare case that `coefs` has one or more trailing singleton dimensions. For, MATLAB suppresses trailing singleton dimensions, hence, without this explicit specification of the intended size of `coefs`, `spmak` would interpret `coefs` incorrectly.

`spmak` prompts you for `knots` and `coefs`.

`sp = spmak(knots,coeffs)` returns the spline `sp`.

Examples   `spmak(1:6,0:2)` constructs a spline function with basic interval [1. .6], with 6 knots and 3 coefficients, hence of order 6 - 3 = 3.

`spmak(t,1)` provides the B-spline $B(\cdot|t)$ in B-form.

The coefficients may be d-vectors (e.g., 2-vectors or 3-vectors), in which case the resulting spline is a curve or surface (in R^2 or R^3).

If the intent is to construct a 2-vector-valued bivariate polynomial on the rectangle [–1..1] × [0..1], linear in the first variable and constant in the second, say

```
 coefs = zeros([2 2 1]); coefs(:,:,1) = [1 0;0 1];
```

then the straightforward

```
sp = spmak({[-1 -1 1 1],[0 1]},coefs);
```

will result in the error message `'There should be no more knots than coefficients'`, because the trailing singleton dimension of

coefs will not be perceived by spmak, while proper use of that third argument, as in

```
sp = spmak({[-1 -1 1 1],[0 1]},coefs,[2 2 1]);
```

will succeed. Replacing here [2 2 1] by size(coefs) would not work.

See the example "Intro to B-form" for other examples.

Diagnostics   There will be an error return if the proposed knot sequence fails to be nondecreasing, or if the coefficient array is empty, or if there are not more knots than there are coefficients. If the spline is to be multivariate, then this last diagnostic may be due to trailing singleton dimensions in coefs.

See Also   fnbrk

Purpose   Explain spline terms

Syntax   spterms(term)
                expl = spterms(term)
                [...,term] = spterms(...)

Description   spterms(term) provides, in a message box, an explanation of the
                technical term indicated by the string term as used in the Curve Fitting
                Toolbox spline functions and, specifically, in the GUI splinetool. Only
                the first few (but at least two) letters of the desired term need to be
                specified, and the full term is shown in the title of the message box.

                expl = spterms(term) returns, in expl, the string, or cell array of
                strings, comprising the explanation of the desired term.

                [...,term] = spterms(...) also returns, in term, the fully
                spelled-out term actually used.

Examples   spterms('sp') gives an explanation of the term 'spline', while
                spterms('spline i') explains the terms 'spline interpolation'.

                help spterms provides the list of all available terms.

See Also   splinetool

How To   • "List of Terms" on page A-3

stcol

Purpose   Scattered translates collocation matrix

Syntax
```
colmat = stcol(centers,x,type)
colmat = stcol(...,'tr')
```

Description   `colmat = stcol(centers,x,type)` is the matrix whose (i,j)th entry is

$$\psi_j\big(x(:,i)\big), \quad i=1:\text{size}(x,2), j=1:n$$

with the bivariate functions ψ_j and the number `n` depending on the `centers` and the string `type`, as detailed in the description of `stmak`.

`centers` and `x` must be matrices with the same number of rows.

The default for `type` is the string `'tp'`, and for this default, `n` equals `size(centers,2)`, and the functions ψ_j are given by

$$\psi_j(x)=\psi\big(x-\text{centers}(:,j)\big), \quad j=1:n$$

with ψ the thin-plate spline basis function

$$\psi(x)=|x|^2\log|x|^2$$

and with $|x|$ denoting the Euclidean norm of the vector x.

Note See `stmak` for a description of other possible values for `type`.

The matrix `colmat` is the coefficient matrix in the linear system

$$\sum_j a_j\psi_j(x(:,i))=y_i, \quad i=1:\text{size}(x,2)$$

that the coefficients a_j of the function $f=\Sigma_j a_j\psi_j$ must satisfy in order that f interpolate the value y_i at the site $x(:,i)$, all `i`.

colmat = stcol(...,'tr') returns the transpose of the matrix returned by stcol(...).

Examples   **Example 1.** The following evaluates and plots the function

$$f(x) = \psi(x - c_1) + \psi(x - c_2) + \psi(x - c_3) - 3.5\psi(x)$$

on a regular mesh, with ψ the above thin-plate basis function, and with c_1, c_2, c_3 three points on the unit circle; see the figure below.

```
a = [0,2/3*pi,4/3*pi]; centers = [cos(a), 0; sin(a), 0];
[xx,yy] = ndgrid(linspace(-2,2,45));
xy = [xx(:) yy(:)].';
coefs = [1 1 1 -3.5];
zz = reshape( coefs*stcol(centers,xy,'tr') , size(xx));
surf(xx,yy,zz), view([240,15]), axis off
```


Example 2. The following also evaluates, on the same mesh, and plots the length of the gradient of the function in Example 1.

```
zz = reshape( sqrt(...
        ([coefs,0]*stcol(centers,xy,'tp10','tr')).^2 + ...
        ([coefs,0]*stcol(centers,xy,'tr','tp01')).^2),
size(xx));
figure, surf(xx,yy,zz), view([220,-15]), axis off
```

See Also   spcol | stmak

Purpose   Put together function in stform

Syntax
```
stmak(centers,coefs)
st = stmak(centers,x,type)
st = stmak(centers,coefs,type,interv)
```

Description   `stmak(centers,coefs)` returns the stform of the function *f* given by

$$f(x) = \sum_{j=1}^{n} \text{coefs}(:,j) \cdot \psi(x - \text{centers}(:,j))$$

with

$$\psi(x) = |x|^2 \log |x|^2$$

the thin-plate spline basis function, and with $|x|$ denoting the Euclidean norm of the vector *x*.

`centers` and `coefs` must be matrices with the same number of columns.

`st = stmak(centers,x,type)` stores in `st` the stform of the function *f* given by

$$f(x) = \sum_{j=1}^{n} \text{coefs}(:,j) \cdot \psi_j(x)$$

with the ψ_j as indicated by the string `type`, which can be one of the following:

- `'tp00'`, for the thin-plate spline;
- `'tp10'`, for the first derivative of a thin-plate spline wrto its first argument;
- `'tp01'`, for the first derivative of a thin-plate spline wrto its second argument;
- `'tp'`, the default.

Here are the details.

`'tp00'`	$\psi_j(x) = \varphi(\mid x - c_j \mid^2)$, c_j =`centers(:,j)`, `j=1:n-3`
	with $\varphi(t) = t\log(t)$
	$\psi_{n-2}(x) = x(1)$
	$\psi_{n-1}(x) = x(2)$
	$\psi_n(x) = 1$
`'tp10'`	$\psi_j(x) = \varphi(\mid x - c_j \mid^2)$, c_j =`centers(:,j)`, `j=1:n-1`
	with $\varphi(t) = (D_1 t)(\log t + 1)$, and $D_1 t$ the partial derivative of $t = t(x) = \mid x - c_j \mid^2$
	wrto $x(1)$
	$\psi_n(x) = 1$
`'tp01'`	$\psi_j(x) = \varphi(\mid x - c_j \mid^2)$, c_j =`centers(:,j)`, `j=1:n-1`
	with $\varphi(t) = (D_2 t)(\log t + 1)$, and $D_2 t$ the partial derivative of $t = t(x) = \mid x - c_j \mid^2$
	wrto $x(2)$
	$\psi_n(x) = 1$
`'tp'` (default)	$\psi_j(x) = \varphi(\mid x - c_j \mid^2)$, c_j =`centers(:,j)`, `j=1:n`
	with $\varphi(t) = t\log(t)$

`st = stmak(centers,coefs,type,interv)` also specifies the basic interval for the stform, with `interv{j}` specifying, in the form `[a,b]`, the range of the jth variable. The default for `interv` is the smallest such box that contains all the given centers.

Examples   **Example 1.** The following generates the figure below, of the thin-plate spline basis function, $\psi(x) = \mid x \mid^2 \log \mid x \mid^2$, but suitably restricted to show that this function is negative near the origin. For this, the extra lines are there to indicate the zero level.

```
inx = [-1.5 1.5]; iny = [0 1.2];
fnplt(stmak([0;0],1),{inx,iny})
hold on, plot(inx,repmat(linspace(iny(1),iny(2),11),2,1),'r')
```

```
view([25,20]),axis off, hold off
```


Example 2. We now also generate and plot, on the very same domain, the first partial derivative $D_2\psi$ of the thin-plate spline basis function, with respect to its second argument.

```
inx = [-1.5 1.5]; iny = [0 1.2];
fnplt(stmak([0;0],[1 0],'tp01',{inx,iny}))
view([13,10]),shading flat,axis off
```

Note that, this time, we have explicitly set the basic interval for the stform.

The resulting figure, below, shows a very strong variation near the origin. This reflects the fact that the *second* derivatives of ψ have a logarithmic singularity there.

See Also   stcol

Purpose   Positive part

Syntax   xp = subplus(x)

Description xp = subplus(x) returns $(x)_+$, i.e., the positive part of x, which is x
                if x is nonnegative and 0 if x is negative. In other words, xp equals
                max(x,0). If x is an array, this operation is applied entry by entry.

Examples   **Example 1.** Here is a plot of the essential part of the subplus function,
                as generated by

                x = -2:2; plot(x,subplus(x),'linew',4), axis([-2,2,-.5,2.5])

                Example 2. The following anonymous function describes the so-called
                hat function:

                hat = @(x) subplus(x) - 2*subplus(x-1) + subplus(x-2);

                i.e., the spline also given by spmak(0:2,1), as the following plot shows.

                x = -.5:.5:2.5; plot(x,hat(x),'linew',4), set(gca,'Fontsize',16)

subplus

Purpose   Titanium test data

Syntax   `[x,y] = titanium`

Description `[x,y] = titanium` returns measurements of a certain property of
titanium as a function of temperature. Since their use in , these data
have become a standard test for data fitting since they are hard to fit by
classical techniques and have a significant amount of noise.

Examples   The plot of the data shown below is generated by the following
commands:

```
[x,y] = titanium; plot(x,y,'ok'), set(gca,'Fontsize',16)
```


References   C. de Boor and J. R. Rice, Least squares cubic spline approximation II -
Variable knots, CSD TR 21, Comp.Sci., Purdue Univ., April 1968.

tpaps

Purpose   Thin-plate smoothing spline

Syntax
```
tpaps(x,y)
tpaps(x,y,p)
[...,p] = tpaps(...)
```

Description   `tpaps(x,y)` is the stform of a thin-plate smoothing spline *f* for the given data sites `x(:,j)` and the given data values `y(:,j)`. The `x(:,j)` must be distinct points in the plane, the values can be scalars, vectors, matrices, even ND-arrays, and there must be exactly as many values as there are sites.

The thin-plate smoothing spline *f* is the unique minimizer of the weighted sum

$$pE(f) + (1-p)R(f)$$

with *E*(*f*) the error measure

$$E(f) = \sum_j \left| y(:,j) - f\left(x(:,j)\right) \right|^2$$

and *R*(*f*) the roughness measure

$$R(f) = \int (\left| D_1 D_1 f \right|^2 + 2 \left| D_1 D_2 f \right|^2 + \left| D_2 D_2 f \right|^2)$$

Here, the integral is taken over all of R^2, $|z|^2$ denotes the sum of squares of all the entries of *z*, and $D_i f$ denotes the partial derivative of *f* with respect to its *i*th argument, hence the integrand involves second partial derivatives of *f*. The smoothing parameter `p` is chosen so that `(1-p)/p` equals the average of the diagonal entries of the matrix `A`, with `A + (1-p)/p*eye(n)` the coefficient matrix of the linear system for the `n` coefficients of the smoothing spline to be determined. This choice of `p` is meant to ensure that we are in between the two extremes, of interpolation (when `p` is close to `1` and the coefficient matrix is essentially `A`) and complete smoothing (when `p` is close to `0` and the

coefficient matrix is essentially a multiple of the identity matrix). This should serve as a good first guess for p.

tpaps(x,y,p) also inputs the *smoothing parameter*, p, a number between 0 and 1. As the smoothing parameter varies from 0 to 1, the smoothing spline varies, from the least-squares approximation to the data by a linear polynomial when p is 0, to the thin-plate spline interpolant to the data when p is 1.

[...,p] = tpaps(...)   also returns the smoothing parameter actually used.

Examples   **Example 1.** The following code obtains values of a smooth function at 31 randomly chosen sites, adds some random noise to these values, and then uses tpaps to recover the underlying exact smooth values. To illustrate how well tpaps does in this case, the code plots, in addition to the smoothing spline, the exact values (as black balls) as well as each arrow leading from a smoothed value to the corresponding noisy value.

```
rand('seed',23); nxy = 31;
xy = 2*(rand(2,nxy)-.5); vals = sum(xy.^2);
noisyvals = vals + (rand(size(vals))-.5)/5;
st = tpaps(xy,noisyvals); fnplt(st), hold on
avals = fnval(st,xy);
plot3(xy(1,:),xy(2,:),vals,'wo','markerfacecolor','k')
quiver3(xy(1,:),xy(2,:),avals,zeros(1,nxy),zeros(1,nxy), ...
        noisyvals-avals,'r'), hold off
```

Example 2. The following code uses an interpolating thin-plate spline to vector-valued data values to construct a map, from the plane to the plane, that carries the unit square $\{x : |x(j)| \leq 1, j = 1{:}2\}$ approximately onto the unit disk $\{x : x(1)^2 + x(2)^2 \leq 1\}$, as shown by the picture generated.

```
n = 64; t = linspace(0,2*pi,n+1); t(end) = [];
values = [cos(t); sin(t)];
centers = values./repmat(max(abs(values)),2,1);
st = tpaps(centers, values, 1);
fnplt(st), axis equal
```

Note the choice of 1 for the smoothing parameter here, to obtain interpolation.

Limitations   The determination of the smoothing spline involves the solution of a linear system with as many unknowns as there are data points. Since the matrix of this linear system is full, the solving can take a long time even if, as is the case here, an iterative scheme is used when there are more than 728 data points. The convergence speed of that iteration is strongly influenced by p, and is slower the larger p is. So, for large problems, use interpolation, i.e., p equal to 1, only if you can afford the time.

See Also   csaps | spaps

type

Purpose	Name of `cfit`, `sfit`, or `fittype` object
Syntax	`name = type(fun)`
Description	`name = type(fun)` returns the custom or library name `name` of the `cfit`, `sfit`, or `fittype` object `fun` as a character array.
| **Examples** | ```
f = fittype('a*x^2+b*exp(n*x)');
category(f)
ans =
custom
type(f)
ans =
customnonlinear

g = fittype('fourier4');
category(g)
ans =
library
type(g)
ans =
fourier4
``` |
| **See Also** | `fittype` \| `category` |
| **How To** | • "List of Library Models for Curve and Surface Fitting" on page 4-13 |

Bibliography

[1] Barber, C. B., D. P. Dobkin, and H. T. Huhdanpaa. "The Quickhull Algorithm for Convex Hulls." *ACM Transactions on Mathematical Software*. Vol. 22, No. 4, 1996, pp. 469–483.

[2] Bevington, P. R., and D. K. Robinson. *Data Reduction and Error Analysis for the Physical Sciences*. 2nd ed. London: McGraw-Hill, 1992.

[3] Branch, M. A., T. F. Coleman, and Y. Li. "A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems." *SIAM Journal on Scientific Computing*. Vol. 21, No. 1, 1999, pp. 1–23.

[4] Carroll, R. J., and D. Ruppert. *Transformation and Weighting in Regression*. London: Chapman & Hall, 1988.

[5] Chambers, J., W. S. Cleveland, B. Kleiner, and P. Tukey. *Graphical Methods for Data Analysis*. Belmont, CA: Wadsworth International Group, 1983.

[6] Cleveland, W. S. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association*. Vol. 74, 1979, pp. 829–836.

[7] Cleveland, W. S., and S. J. Devlin. "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting." *Journal of the American Statistical Association*. Vol. 83, 1988, pp. 596–610.

[8] Daniel, C., and F. S. Wood. *Fitting Equations to Data*. Hoboken, NJ: Wiley-Interscience, 1980.

[9] DeAngelis, D. J., J. R. Calarco, J. E. Wise, H. J. Emrich, R. Neuhausen, and H. Weyand. "Multipole Strength in ^{12}C from the (e,e'α) Reaction for Momentum Transfers up to 0.61 fm^{-1}." *Physical Review C*. Vol. 52, No. 1, 1995, pp. 61–75.

[10] de Boor, C. *A Practical Guide to Splines*. Berlin: Springer-Verlag, 1978.

[11] Draper, N. R., and H. Smith. *Applied Regression Analysis*. 3rd ed. Hoboken, NJ: Wiley-Interscience, 1998.

[12] DuMouchel, W., and F. O'Brien. "Integrating a Robust Option into a Multiple Regression Computing Environment." *Computing Science and Statistics: Proceedings of the 21st Symposium on the Interface*. (K. Berk and L. Malone, eds.) Alexandria, VA: American Statistical Association, 1989, pp. 297–301.

[13] Goodall, C. "A Survey of Smoothing Techniques." *Modern Methods of Data Analysis*. (J. Fox and J. S. Long, eds.) Newbury Park, CA: Sage Publications, 1990, pp. 126–176.

[14] Holland, P. W., and R. E. Welsch. "Robust Regression Using Iteratively Reweighted Least-Squares." *Communications in Statistics—Theory and Methods*. Vol. A6, 1977, pp. 813–827.

[15] Huber, P. J. *Robust Statistics*. Hoboken, NJ: Wiley-Interscience, 1981.

[16] Hutcheson, M. C. "Trimmed Resistant Weighted Scatterplot Smooth." Master's Thesis. Cornell University, Ithaca, NY, 1995.

[17] Levenberg, K. "A Method for the Solution of Certain Problems in Least Squares." *Quarterly of Applied Mathematics*. Vol. 2, 1944, pp. 164–168.

[18] Marquardt, D. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters." *SIAM Journal on Applied Mathematics*. Vol. 11, 1963, pp. 431–441.

[19] Orfanidis, S. J. *Introduction to Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[20] Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge, UK: Cambridge University Press, 1993.

[21] Street, J. O., R. J. Carroll, and D. Ruppert. "A Note on Computing Robust Regression Estimates Via Iteratively Reweighted Least Squares." *The American Statistician.* Vol. 42, 1988, pp. 152–154.

[22] Watson, David E. *Contouring: A Guide to the Analysis and Display of Spatial Data.* Tarrytown, NY: Pergamon, 1992.

Index

W