J Math Chem (2012) 50:1747–1764
DOI 10.1007/s10910-012-0004-z

ORIGINAL PAPER

D-MORPH regression for modeling with fewer unknown parameters than observation data

Genyuan Li · Roberto Rey-de-Castro · Herschel Rabitz

Received: 24 February 2012 / Accepted: 28 March 2012 / Published online: 19 April 2012

Abstract D-MORPH regression is a procedure for the treatment of a model prescribed as a linear superposition of basis functions with *less* observation data than the number of expansion parameters. In this case, there is an infinite number of solutions exactly fitting the data. D-MORPH regression provides a practical systematic means to search over the solutions seeking one with desired ancillary properties while preserving fitting accuracy. This paper extends D-MORPH regression to consider the common case where there is *more* observation data than unknown parameters. This situation is treated by utilizing a proper subset of the normal equation of least-squares regression to judiciously reduce the number of linear algebraic equations to be *less* than the number of unknown parameters, thereby permitting application of D-MORPH regression. As a result, no restrictions are placed on model complexity, and the model with the best prediction accuracy can be automatically and efficiently identified. Ignition data for a H_2/air combustion model as well as laboratory data for quantum-control-mechanism identification are used to illustrate the method.

Keywords D-MORPH regression · Least-squares regression · Regularization · Ridge regression

G. Li · R. Rey-de-Castro · H. Rabitz (✉)
Department of Chemistry, Princeton University, Princeton, NJ 08544, USA
e-mail: hrabitz@princeton.edu

G. Li
e-mail: genyuan@princeton.edu

R. Rey-de-Castro
e-mail: rrey@princeton.edu

1 Introduction

A mathematical model generally is an approximation to the structure of a complex physical system, and the model is often determined from a set of observation data. Consider the target variable y defined by a deterministic function $f(\mathbf{x})$ with additive Gaussian noise so that [1]

$$y = f(\mathbf{x}) + \varepsilon, \tag{1}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ are the inputs, ε is Gaussian noise with expectation $E(\varepsilon) = 0$ and variance $\mathrm{Var}(\varepsilon) = \sigma_\varepsilon^2$. A common special case is to approximate $f(\mathbf{x})$ as a linear combination of m basis functions $\phi_i(\mathbf{x})$

$$f(\mathbf{x}) \approx y(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^{m} w_i \phi_i(\mathbf{x}). \tag{2}$$

Here $\mathbf{w} = (w_1, w_2, \ldots, w_m)^T$ are the parameters to be identified; $\phi_1(\mathbf{x})$ may be 1 if a constant term $w_1\phi_1(\mathbf{x}) = w_1$ is included as an unknown parameter. The parameters $\mathbf{w}$ are often determined from a set of observation data $(\mathbf{x}^{(j)}, y^{(j)})$ $(j = 1, 2, \ldots, N)$ utilizing the equations

$$\phi(\mathbf{x}^{(j)})^T \mathbf{w} = y^{(j)}, \quad (j = 1, 2, \ldots, N) \tag{3}$$

where $\phi(\mathbf{x}^{(j)})^T = (\phi_1(\mathbf{x}^{(j)}), \phi_2(\mathbf{x}^{(j)}), \ldots, \phi_m(\mathbf{x}^{(j)}))$. In matrix form Eq. 3 becomes

$$\mathbf{\Phi w} = \mathbf{y}, \tag{4}$$

where

$$\mathbf{\Phi} = \begin{bmatrix} \phi_1(\mathbf{x}^{(1)}) & \phi_2(\mathbf{x}^{(1)}) & \ldots & \phi_m(\mathbf{x}^{(1)}) \\ \phi_1(\mathbf{x}^{(2)}) & \phi_2(\mathbf{x}^{(2)}) & \ldots & \phi_m(\mathbf{x}^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}^{(N)}) & \phi_2(\mathbf{x}^{(N)}) & \ldots & \phi_m(\mathbf{x}^{(N)}) \end{bmatrix} \tag{5}$$

and

$$\mathbf{y} = (y^{(1)}, y^{(2)}, \ldots, y^{(N)})^T. \tag{6}$$

Importantly, although the unknown coefficients $\mathbf{w}$ enter linearly in the model of Eq. 2, the relationship $\mathbf{x} \rightarrow y$ can be nonlinear through the basis functions $\phi_i(\mathbf{x})$.

Least-squares regression is commonly used to determine $\mathbf{w}$ by minimizing the error $E(\mathbf{w})$ expressed as the residual sum of squares (RSS),

$$E(\mathbf{w}) = \mathrm{RSS} = \|\mathbf{\Phi w} - \mathbf{y}\|^2. \tag{7}$$

The more complex the model (i.e., more basis functions, and consequently more unknown w_i parameters included in Eq. 4), then the more flexibility available to fit the data. However, the goal of modeling is often to make accurate predictions for new as yet unobserved data, and good fitting does not necessarily mean good prediction. Over-fitting may occur for complex models, which can result in unsatisfactory predictions. Least-squares regression is a special case of maximum likelihood, and over-fitting is a general problem in utilizing maximum likelihood [1]. For a given level of model complexity, the over-fitting problem becomes more severe as the amount of data decreases. One rough heuristic is that the number of data points should be no less than some multiple (say 5 or 10) of the number of unknown parameters. This situation poses the unsatisfactory scenario of limiting the number of unknown parameters in a model according to the size of the available data. A more desirable perspective would entail choosing the complexity of the model according to the complexity of the problem [1].

An appropriate level of model complexity may reasonably be determined by dividing the available data into a training set, employed to determine the parameters $\mathbf{w}$, and a separate validation set, used to optimize the model complexity. In many cases, however, this procedure can be wasteful of the valuable data. As an alternative procedure, a statistical test (e.g., F-test) may be used to identify the significant basis functions in Eq. 4 for a given set of data in order to compose a model with proper complexity [2]. However, the significant basis functions will generally depend on the chosen training data. When the size of the training data set is small and not uniformly distributed, the results can be misleading. The predictions based on such a procedure with different distributions may vary and be unsatisfactory.

Another technique that is often utilized to manage over-fitting is *regularization*, which involves adding a penalty term to $E(\mathbf{w})$. For example, in ridge regression [3–8] a regularization term is included in the minimization:

$$E(\mathbf{w}) = \|\mathbf{\Phi w} - \mathbf{y}\|^2 + \|\Gamma \mathbf{w}\|^2 \tag{8}$$

for some suitable Tikhonov matrix, Γ. In many cases, Γ is chosen as proportional to the identity matrix $\Gamma = \lambda^{1/2}\mathrm{I}$ with parameter $\lambda > 0$ giving preference to solutions $\mathbf{w}$ with smaller norms:

$$E(\mathbf{w}) = \|\mathbf{\Phi w} - \mathbf{y}\|^2 + \lambda\|\mathbf{w}\|^2. \tag{9}$$

Smoothing splines [9–11] introduce a second term to penalize the square curvature norm of the model function. These methods generally pose a trade-off with the following shortcomings: (1) in order to improve prediction accuracy, the fitting accuracy often has to be sacrificed to some degree, (2) determination of the optimal value for the parameter λ (e.g., by the discrepancy principle, cross-validation, the L-curve method, restricted maximum likelihood and unbiased predictive risk estimator validation) [12] can require extensive computational effort. This paper will show that D-MORPH regression does not impose restrictions on model complexity even when the number of data points is close to or less than the number of unknown parameters. D-MORPH

regression automatically and efficiently searches over the system complexity drawn from a set of possible models.

The paper is organized as follows. Section 2 summarizes D-MORPH regression methodology. Section 3 presents two illustrations of D-MORPH regression: (a) ignition delay for a H_2/air combustion model and (b) laboratory data for quantum-control-mechanism identification [23,24]. Some concluding remarks are given in Sect. 4.

2 Methodology

2.1 D-MORPH regression

The principles of D-MORPH regression are briefly summarized here. Readers are referred to [13] for further details and to [14–17] for background. When the number m of unknown parameters (w_i's) is larger than the number of observation data points N (i.e., $m > N$) with the provision that $\mathbf{y}$ lies in Ran($\mathbf{\Phi}$) when $\mathbf{\Phi}$ does not have a full row rank, Eq. 4 is consistent and has an infinite number of solutions for $\mathbf{w}$ having the general form

$$\mathbf{w} = \mathbf{\Phi}^{-}\mathbf{y} + (\mathrm{I}_m - \mathbf{\Phi}^{-}\mathbf{\Phi})\mathbf{u}, \tag{10}$$

where I_m is the identity matrix of dimension m and $\mathbf{u}$ is an arbitrary vector in $\mathbb{R}^m$, and $\mathbf{\Phi}^{-}$ is a generalized inverse of $\mathbf{\Phi}$ satisfying the condition

$$\mathbf{\Phi}\mathbf{\Phi}^{-}\mathbf{\Phi} = \mathbf{\Phi}. \tag{11}$$

One choice for $\mathbf{\Phi}^{-}$ in Eq. 10 is $\mathbf{\Phi}^{+}$, i.e.,

$$\mathbf{w} = \mathbf{\Phi}^{+}\mathbf{y} + (\mathrm{I}_m - \mathbf{\Phi}^{+}\mathbf{\Phi})\mathbf{u}, \tag{12}$$

where $\mathbf{\Phi}^{+}$ is the generalized inverse of $\mathbf{\Phi}$ satisfying all of the four Penrose conditions [18]:

$$\begin{array}{ll} (1)\ \mathbf{\Phi}\mathbf{\Phi}^{+}\mathbf{\Phi} = \mathbf{\Phi}, & (2)\ \mathbf{\Phi}^{+}\mathbf{\Phi}\mathbf{\Phi}^{+} = \mathbf{\Phi}^{+}, \\ (3)\ (\mathbf{\Phi}\mathbf{\Phi}^{+})^{T} = \mathbf{\Phi}\mathbf{\Phi}^{+}, & (4)\ (\mathbf{\Phi}^{+}\mathbf{\Phi})^{T} = \mathbf{\Phi}^{+}\mathbf{\Phi}. \end{array} \tag{13}$$

Equation 12 with $\mathbf{u} = \mathbf{0}$ (i.e., $\mathbf{\Phi}^{+}\mathbf{y}$) is the solution from traditional least-squares regression with the smallest norm $\|\mathbf{w}\|$.

All the solutions $\mathbf{w}$ of Eq. 4 given by Eq. 12 compose a completely connected submanifold $\mathcal{M} \subset \mathbb{R}^m$. D-MORPH regression searches for a solution satisfying an extra requirement by considering an exploration path $\mathbf{w}(s)$ within $\mathcal{M}$ with $s \in [0, \infty)$, which satisfies the differential equation

$$\frac{\mathrm{d}\mathbf{w}(s)}{\mathrm{d}s} = P\mathbf{v}(s) = (\mathrm{I}_m - \mathbf{\Phi}^{+}\mathbf{\Phi})\mathbf{v}(s), \tag{14}$$

where $\mathbf{v}(s) = \mathrm{d}\mathbf{u}/\mathrm{d}s$ and P is an orthogonal projector [18] satisfying

$$P^2 = P, \qquad P^T = P, \qquad P = P^2 = P^T P. \tag{15}$$

The function vector $\mathbf{v}(s)$ may be freely chosen to not only enable broad choices for exploring $\mathbf{w}(s)$, but also provide the possibility of continuously reducing a defined cost $\mathcal{K}(\mathbf{w}(s))$ (e.g., the model variance, fitting smoothness, the weighted norm of $\mathbf{w}$, etc.) along the exploration path. If the free function vector is chosen as

$$\mathbf{v}(s) = -\frac{\partial \mathcal{K}(\mathbf{w}(s))}{\partial \mathbf{w}}, \tag{16}$$

then we obtain

$$\begin{aligned} \frac{\mathrm{d}\mathcal{K}(\mathbf{w}(s))}{\mathrm{d}s} &= \left(\frac{\partial \mathcal{K}(\mathbf{w}(s))}{\partial \mathbf{w}}\right)^T \frac{\mathrm{d}\mathbf{w}(s)}{\mathrm{d}s} = \left(\frac{\partial \mathcal{K}(\mathbf{w}(s))}{\partial \mathbf{w}}\right)^T P\mathbf{v}(s) \\ &= -\left(P\frac{\partial \mathcal{K}(\mathbf{w}(s))}{\partial \mathbf{w}}\right)^T \left(P\frac{\partial \mathcal{K}(\mathbf{w}(s))}{\partial \mathbf{w}}\right) \le 0, \end{aligned} \tag{17}$$

i.e., the cost $\mathcal{K}$, used as an additional requirement, will be continuously reduced (systematically refining the model) over the exploration course for $s \ge 0$. Therefore,

$$\mathbf{w}_\infty = \lim_{s\to\infty} \mathbf{w}(s)$$

is the solution with the minimum of $\mathcal{K}$. When the cost function is defined as a quadratic form in $\mathbf{w}$

$$\mathcal{K} = \frac{1}{2}\mathbf{w}^T C\mathbf{w}, \tag{18}$$

where C is symmetric and non-negative definite, the analytical form of $\mathbf{w}_\infty$ may be obtained as

$$\mathbf{w}_\infty = V_{m-r}(U_{m-r}^T V_{m-r})^{-1} U_{m-r}^T \mathbf{\Phi}^+ \mathbf{y}, \tag{19}$$

where U_{m-r}, and V_{m-r} are the last $m-r$ columns of U and V obtained by singular value decomposition of PC [19,20]

$$PC = U\begin{bmatrix} S_r & 0 \\ 0 & 0 \end{bmatrix} V^T \tag{20}$$

with S_r being a diagonal matrix of nonzero singular values.

Equation 19 is the key practical formula for the optimal solution $\mathbf{w}$ obtained by D-MORPH regression. This solution $\mathbf{w}_\infty$ is unique in $\mathcal{M}$ corresponding to the global minimum of the cost function. As evident in Eq. 19, the solution $\mathbf{w}_\infty$ given by D-MORPH regression is a special linear combination of the elements of $\mathbf{w}$ obtained by the least-squares regression (i.e., $\mathbf{\Phi}^+\mathbf{y}$).

2.2 D-MORPH regression for fewer unknown parameters than observation data

Suppose we have a set of data and it can be *satisfactorily* fit by a model as a linear combination of $m_0 (< N)$ basis functions $\phi_i(\mathbf{x})(i = 1, 2, \ldots, m_0)$. When $m_0 < N$, Eq. 4 may not be consistent, but it can be solved by least-squares regression to yield $\mathbf{w}_0$ that minimizes the RSS given in Eq. 7. Setting

$$\frac{\mathrm{d(RSS)}}{\mathrm{d}\mathbf{w}} = \mathbf{\Phi}_0^T(\mathbf{\Phi}_0\mathbf{w}_0 - \mathbf{y}) = \mathbf{0} \tag{21}$$

gives the normal equation of least-squares regression

$$\mathbf{\Phi}_0^T\mathbf{\Phi_0}\mathbf{w}_0 = \mathbf{\Phi}_0^T\mathbf{y}. \tag{22}$$

Since both sides are pre-multiplied with $\mathbf{\Phi}_0^T$, Eq. 22 is consistent and always has a solution $\mathbf{w}_0$.

We now expand the number of basis functions, the $\phi_i(\mathbf{x})$'s, by adding in m_1 new members (consequently including m_1 more coefficients w_i's). However, the total number $m(= m_0 + m_1)$ of basis functions (consequently unknown parameters) is still smaller than the number of data N. The corresponding normal equation for the new model with m basis functions is

$$\begin{bmatrix} \mathbf{\Phi}_0^T \\ \mathbf{\Phi}_1^T \end{bmatrix} [\mathbf{\Phi}_0|\mathbf{\Phi}_1] \begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_0^T \\ \mathbf{\Phi}_1^T \end{bmatrix} \mathbf{y}. \tag{23}$$

The RSS given by the solution of Eq. 23 is smaller than the RSS given by the solution of Eq. 22 (i.e., Eq. 23 has better fitting accuracy) because Eq. 23 corresponds to a more complex model with additional unknown parameters to flexibly fit the data.

Consider only the first m_0 equations in Eq. 23, i.e.,

$$\mathbf{\Phi}_0^T\,[\mathbf{\Phi}_0|\mathbf{\Phi}_1] \begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_1 \end{bmatrix} = \mathbf{\Phi}_0^T\mathbf{y} \tag{24}$$

where $\mathbf{\Phi}_0^T\,[\mathbf{\Phi}_0|\mathbf{\Phi}_1]$ is an $(m_0 \times m)$ matrix and $\mathbf{\Phi}_0^T\mathbf{y}$ is an m_0-dimensional constant vector, i.e., Eq. 24 is composed of $m_0 (< m)$ equations with m unknown variables which has an infinite number of solutions $\mathbf{w}$ constituting a submanifold $\mathcal{M} \subset \mathbb{R}^m$. At first sight it may appear unusual to create a system of equations that is underdetermined, but the added flexibility with m unknowns can be utilized to obtain a better solution to the original problem by recognizing that Eq. 24 is just the form for application of D-MORPH regression. The solution of Eq. 22 belongs to $\mathcal{M}$ if we consider $\mathbf{w}_1$ to be zero. Similarly, the solution of Eq. 23 also belongs to $\mathcal{M}$ because Eq. 24 is a part of Eq. 23 and the solution $\mathbf{w}$ of Eq. 23 must satisfy its partial equations, Eq. 24. D-MORPH regression is employed to find the solution $\mathbf{w} \in \mathcal{M}$ with the smallest value of the cost function related to prediction accuracy, which implies that its prediction accuracy is superior than the solutions of Eqs. 22 and 23. In this way, D-MORPH regression

can be applied to modeling with fewer unknown parameters than observation data by setting

$$\mathbf{\Phi} \equiv \mathbf{\Phi}_0^T \, [\mathbf{\Phi}_0 | \mathbf{\Phi}_1] \tag{25}$$

and treating $\mathbf{\Phi}_0^T \mathbf{y}$ as $\mathbf{y}$ in Eq. 4.

The fitting accuracy of D-MORPH regression of Eq. 24 is certainly not superior to the solution of Eq. 23, which has the best fitting accuracy compared to other solutions in $\mathcal{M}$ when all $\phi_i(\mathbf{x})(i = 1, 2, \ldots, m)$ are used. However, since the solution of Eq. 24 obtained by D-MORPH regression possesses a minimum of the imposed cost function (i.e., it has the best prediction accuracy for any data including the training data), the fitting accuracy of D-MORPH regression for Eq. 24 will be close to that for Eq. 22. This behavior provides a criterion for choosing $\mathbf{\Phi}_0$ (both the number m_0 and the particular basis set in $\mathbf{\Phi}_0$) based on its fitting accuracy by least-squares regression being acceptable. This criterion for choosing the basis set in $\mathbf{\Phi}_0$ is not unique, and other considerations may be included. These prospects are beyond the scope of the present work.

3 Illustrations

Two examples will be used to illustrate the application of D-MORPH regression to the common problem of modeling with fewer unknown parameters than observed data. The first example concerns the generation of a reduced hydrogen-oxygen combustion model using the input-output results from computational studies [21]. In this case the resultant simplified model for the ignition delay time has three input variables (x_1, x_2, x_3), 63 unknown parameters and 100 available data points. The second example utilizes laboratory data for quantum-control-mechanism identification [23,24]. The physical system is atomic rubidium vapor controlled by ultra-fast laser pulses and the resultant model is generated to reveal the control mechanism. In this case there are two input variables (x_1, x_2), 45 unknown parameters and 60 experimental data points. The results in both examples will demonstrate the capability of D-MORPH regression to successfully find excellent quality solutions.

3.1 Application to an ignition delay model of H_2/air combustion

An important characteristic of combustion is the ignition delay time t_{ig}, which depends on several factors including the initial temperature, pressure and the equivalence ratio of the fuel and oxygen. A homogeneous H_2/air combustion numerical simulation is used with 8 species (H_2, O_2, H_2O, H, O, OH, HO_2, H_2O_2) and 19 reactions [21] for testing the capability of D-MORPH regression. The initial temperature ($1,000 < T_0 < 1,500\,\text{K}$), logarithmic value of pressure ($0.1 < P < 1\,\text{atm}$), and logarithmic value of H_2/O_2 equivalence ratio ($0.1 < \phi < 10.0$) are chosen as three input variables denoted respectively by x_1, x_2, x_3, and the logarithmic value of the ignition delay (defined as the time lapse t_{ig} needed to attain an increase of 400 K from the

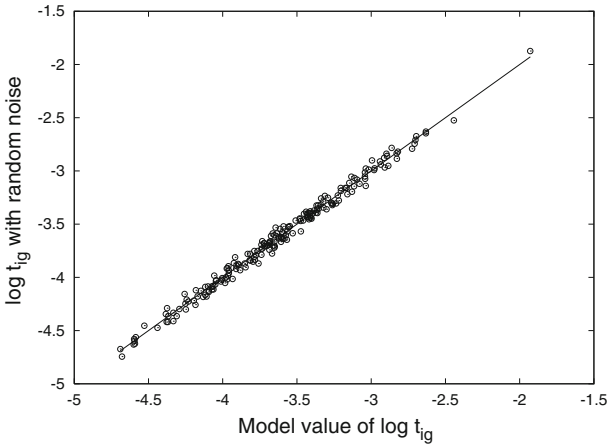

Fig. 1 Comparison between the two data sets for $\log t_{\mathrm{ig}}$, with and without noise

initial temperature) is the output. Two hundred random data points of $\mathbf{x}$ were sampled with a uniform distribution for T_0, $\log P$ and $\log \phi$ within the above ranges, and the corresponding logarithmic values of ignition delay, $\log t_{\mathrm{ig}}$, were calculated from the numerical simulation and defined as $y(\mathbf{x})$. Gaussian white noise was added to the resultant $\log t_{\mathrm{ig}}$ values to simulate the laboratory circumstances. The signal to noise ratio is $\mathrm{Var}(\log t_{\mathrm{ig}})/\mathrm{Var}(\varepsilon) \approx 100$. Comparison between the two data sets for $\log t_{\mathrm{ig}}$, with and without noise, is given in Fig. 1.

All input variables are normalized as $x_i \in [0, 1]$ and the polynomial expansion

$$\begin{aligned} y(\mathbf{x}) - f_0 &= \sum_{i=1}^{36} w_i \phi_i(\mathbf{x}) \\ &= \sum_{i=1}^{3}\sum_{r=1}^{3} \alpha_r^i p_r(x_i) + \sum_{1\le i<j\le 3}\sum_{p=1}^{3}\sum_{q=1}^{3} \beta_{pq}^{ij} p_p(x_i) p_q(x_j) \end{aligned} \tag{26}$$

was used as the approximate model. The lefthand side term $y(\mathbf{x}) - f_0$ is used as $y(\mathbf{x}, \mathbf{w})$ in Eq. 2 with $f_0 = \sum_{s=1}^{N} y^{(s)}/N$; and

$$p_1(x) = \sqrt{3}\,(2x - 1), \tag{27}$$

$$p_2(x) = 6\sqrt{5}\left(x^2 - x + \frac{1}{6}\right), \tag{28}$$

$$p_3(x) = 20\sqrt{7}\left(x^3 - \frac{3}{2}x^2 + \frac{3}{5}x - \frac{1}{20}\right) \tag{29}$$

are orthonormal polynomials for $x \in [0, 1]$. In Eq. 26, polynomials $p_r(x_i)$, and their products $p_p(x_i)p_q(x_j)$ are the basis functions $\phi_i(\mathbf{x})$ in Eq. 2. In Eq. 26 the number m_0 of unknown parameters w_i (i.e., coefficients α_r^i, β_{pq}^{ij}) is 36. The fitting

accuracy of Eq. 26 was found to be quite satisfactory, and thus Eq. 26 was used to construct $\mathbf{\Phi}_0$. Then, additional new basis functions $\phi_i(\mathbf{x})$ (i.e., as third order products $p_p(x_i)p_q(x_j)p_r(x_k)$) were included so that the final model with $m = 63$ unknown parameters,

$$y(\mathbf{x}) - f_0 = \sum_{i=1}^{63} w_i \phi_i(\mathbf{x}) = \sum_{i=1}^{3}\sum_{r=1}^{3} \alpha_r^i p_r(x_i) + \sum_{1\le i<j\le 3}\sum_{p=1}^{3}\sum_{q=1}^{3} \beta_{pq}^{ij} p_p(x_i) p_q(x_j) + \sum_{1\le i<j<k\le 3}\sum_{p=1}^{3}\sum_{q=1}^{3}\sum_{r=1}^{3} \gamma_{pqr}^{ijk} p_p(x_i) p_q(x_j) p_r(x_k) \tag{30}$$

was used to construct $\mathbf{\Phi}$.

The cost function was chosen as the weighted norm of $\mathbf{w}$

$$\mathcal{K} = \frac{1}{2}\sum_{i=1}^{m} C_{ii} w_i^2 = \frac{1}{2}\sum_{i=1}^{m} c_i w_i^2, \tag{31}$$

which reduces C in Eq. 18 to a diagonal matrix with $C_{ii} = c_i$. The weights c_i associated with $p_r(x_i)$, $p_p(x_i)p_q(x_j)$ and $p_p(x_i)p_q(x_j)p_r(x_k)$ were initially set to be r, $p+q$ and $p+q+r$, respectively, i.e., the weight is equal to the sum of the degrees of the polynomials making up each basis function. Therefore, the c_i's take on values between 1 and 9, and the higher degree polynomials have larger weights which makes their contributions more likely to diminish faster under D-MORPH regression. As the contribution of high degree polynomials in the model is repressed, the resultant model is smoother and generally has better prediction accuracy. Figure 2 gives the relationship between $\log t_{ig}$ and x_i for the data. The relation has linear character for x_1 and x_2, but exhibits nonlinear behavior for x_3. Thus, the weights c_i for $p_1(x_i)$, $p_2(x_3)$, $p_1(x_i)p_1(x_j)$ and $p_1(x_i)p_1(x_j)p_1(x_k)$ are set to zero, which removes the coefficients for all linear polynomials and the quadratic function $p_2(x_3)$ from consideration for norm minimization. The remaining weights were set as described above as the sum of the relevant polynomial degrees.

A statistical F-test was also used to identify the significant component functions in Eq. 30 to reduce the number of unknown parameters for least-squares regression. The F-test for the first one hundred data yields the following model with 16 parameters:

$$y(\mathbf{x}) - f_0 = \sum_{i=1}^{16} w_i \phi_i(\mathbf{x}) = \sum_{i=1}^{3} \alpha_1^i p_1(x_i) + \alpha_2^3 p_2(x_3) + \sum_{q=1}^{2} \beta_{1q}^{13} p_1(x_1) p_q(x_3) + \sum_{p=1}^{2}\sum_{q=1}^{3} \beta_{pq}^{23} p_p(x_2) p_q(x_3) + \sum_{q=1}^{2}\sum_{r=1}^{2} \gamma_{1qr}^{123} p_1(x_1) p_q(x_2) p_r(x_3). \tag{32}$$

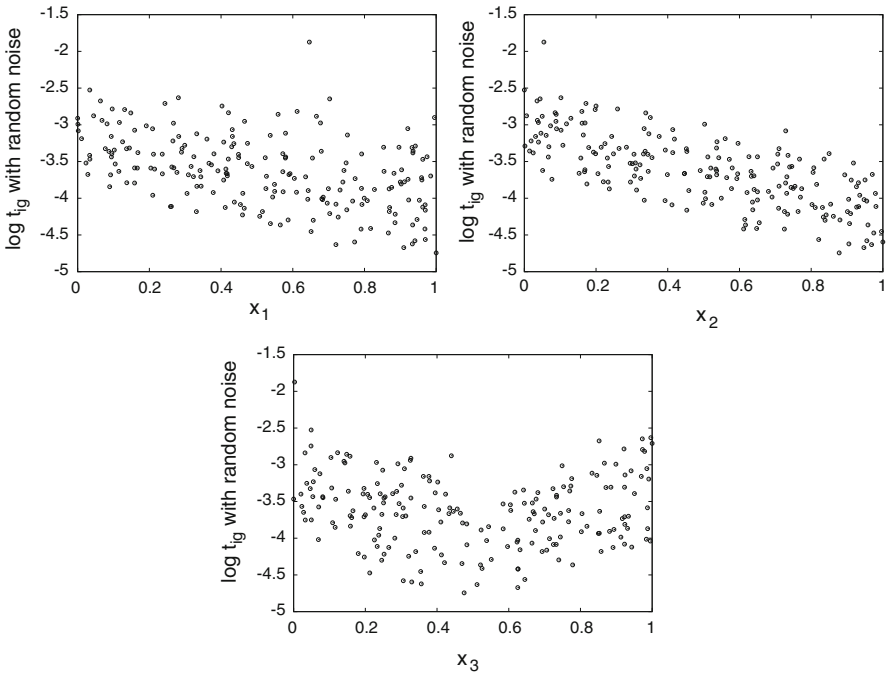

Fig. 2 The relationship between $\log t_{\mathrm{ig}}$ and x_i. These plots are projections from three dimensions to each respective single variable thereby producing the scattered character in each plot

Table 1 Comparison of the average absolute error as well as the relative error obtained by all the methods for the training and testing data on the H_2/air combustion system

Method	Equation	Ave. abs. err.		Ave. rel. err.	
		Training	Testing	Training	Testing
Least-squares	(26)	0.0367	0.0838	0.0107	0.0233
	(30)	0.0187	0.1245	0.0051	0.0341
Least-squares with F-test	(32)	0.0664	0.0975	0.0328	0.0380
Ridge	(30)	0.0267	0.0831	0.0074	0.0231
D-MORPH	(26)(30)*	0.0380	0.0691	0.0106	0.0192

*Eq. (26) for $\mathbf{\Phi}_0$, and Eq. (30) for $\mathbf{\Phi}$

Ridge regression in Eq. 9 with $\lambda = 0.000771$ (obtained by cross validation) was additionally used with Eq. 30.

One hundred data points (i.e., the training data) were used to determine the parameters of the models. An additional 100 points were used to test the resultant models. Note that the test set here is just used to assess the capability of the D-MORPH procedure and not to iteratively refine the model. A comparison of the average absolute error as well as the relative error is given in Table 1 obtained from all these methods for the training and testing data.

The results in Table 1 show that for least-squares regression increasing the model complexity from Eq. 26 to Eq. 30 yields a better fit to the training data, but over-fitting occurs with Eq. 30. Thus, without testing, one cannot correctly select the best model based only on fitting accuracy. For Eq. 32 obtained from the F-test, the fitting of the training data has large errors. Although the prediction accuracy of Eq. 32 is better than Eq. 30, it is not satisfactory and worse than Eq. 26 by least-squares. This behavior demonstrates that the identification of significant basis functions by the F-test from a small set of data can be misleading. Ridge regression improves the prediction accuracy without reducing the model complexity, but the computational effort involved is large due to the need for determining the optimal value of λ. D-MORPH regression has the best prediction accuracy for the most complex model, Eq. 30. As expected, the fitting accuracy of D-MORPH regression is very close to that obtained by least-squares regression for the model used to construct $\mathbf{\Phi}_0$. The fitting accuracy may be used as the main consideration for choosing $\mathbf{\Phi}_0$, and that was done here. Figure 3 gives truth plot comparisons using Eq. 30 treated by least-squares, ridge and D-MORPH regression. In the combustion simulation example with D-MORPH regression, Eq. 30 has 63 unknown parameters with only a slightly larger number of 100 data points.

3.2 Application to quantum-control-mechanism analysis

Quantum control seeks to manipulate dynamical events at the atomic and molecular scale using tailored laser fields [26]. An automatic closed-loop procedure is often used to find laser fields that maximize a signal reflecting the yield in a desired final state. In this way quantum systems can be directed to perform a variety of tasks. In order to gain an understanding of the mechanisms induced by the fields found by quantum control, the technique of Hamiltonian encoding and observable decoding (HE–OD) [23] has been implemented in the laboratory [23,24]. Over a sequence of experiments, HE–OD introduces special encoded signatures into spectral components of the control field, and the outcome appears as a modulated signal. Decoding the modulated signal identifies the hierarchy of correlations between components of the control field. The HE–OD procedure yields the complex amplitudes corresponding to terms in the Dyson expansion for the time evolution operation [23]. An important issue when performing HE–OD is to reduce the amount of data needed to extract the amplitudes, and this need is consistent with the capability of the D-MORPH regression procedure as illustrated here.

In this section we present an application of D-MORPH regression to decoding HE–OD experimental data. The quantum system is atomic Rb and the measurements (details of the experiments are given in [25]) are labeled by the index $s = 1, 2, \ldots, N$. The phases of selected spectral components of the laser field were encoded by adding the functions $x_1(s)$ and $x_2(s)$. The resulting observed laboratory output signal $y(s)$ can be expressed as

$$y(s) = \left\| a_0 + \sum_{j=1}^{m} a_j e^{i(n_{j1}x_1(s) + n_{j2}x_2(s))} \right\|^2, \quad (s = 1, 2, \ldots, N) \tag{33}$$

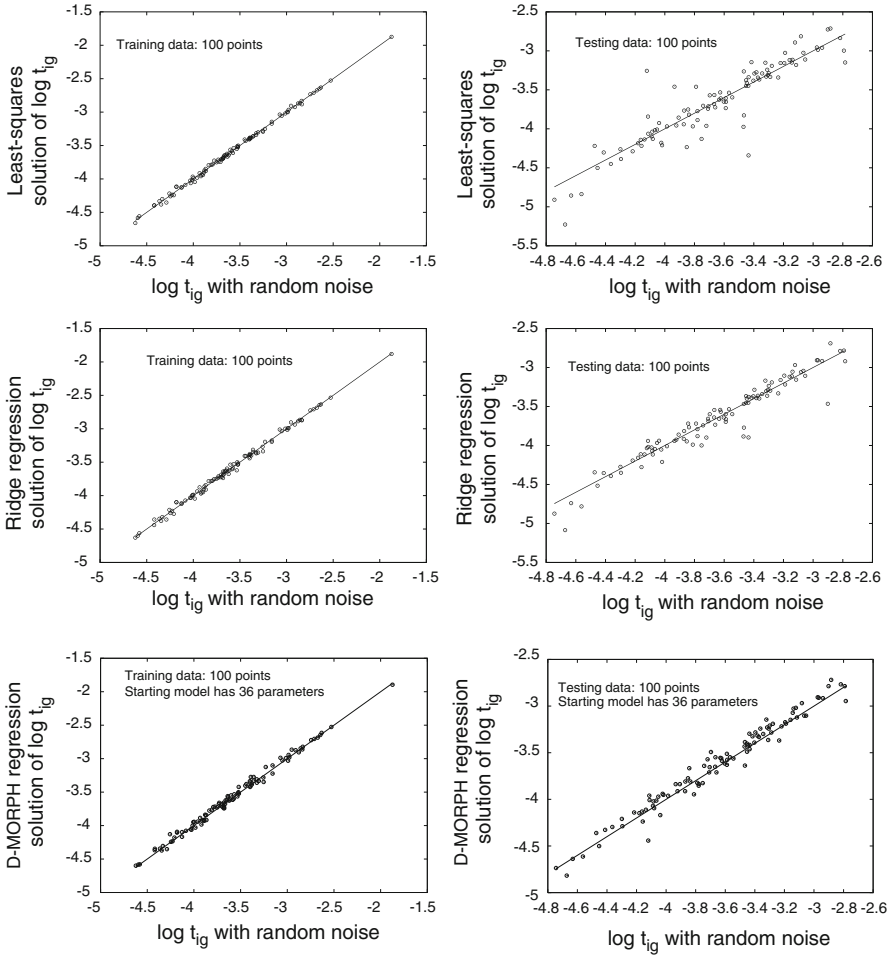

Fig. 3 Truth plots for least-squares, ridge and D-MORPH regression of Eq. 30 for the H_2/air combustion ignition model data with noise. Both the training and testing data sets contain 100 points

where $a_j = r_j e^{i\varphi_j}$ are complex numbers and n_{j1}, n_{j2} are integers representing different transition processes [23]. The actual rubidium experimental system is well described by [25]

$$\begin{aligned} y(s) &= \|a_0 + a_1 e^{ix_1(s)} + a_2 e^{ix_2(s)}\|^2 \\ &= \alpha_0 + \alpha_1 \cos(x_1(s)) + \beta_1 \sin(x_1(s)) + \alpha_2 \cos(x_2(s)) + \beta_2 \sin(x_2(s)) \\ &\quad + \alpha_3 \cos(x_1(s) - x_2(s)) + \beta_3 \sin(x_1(s) - x_2(s)) \end{aligned} \tag{34}$$

with $n_{11} = n_{22} = 1, n_{12} = n_{21} = 0$ in Eq. 33. $\alpha_0, \alpha_j, \beta_j (j = 1, 2, 3)$ are real parameters obtained from $r_j, \varphi_j (j = 0, 1, 2, 3)$.

In practice, an *a priori* physically based model may not be known, thus we expand Eq. 34 into a larger model with the integers $|n_{j1}| + |n_{j2}| = 1, 2, 3$:

$$\begin{aligned} y(s) = \| & a_0 + a_1 e^{ix_1(s)} + a_2 e^{ix_2(s)} + a_3 e^{i(x_1(s)-x_2(s))} + a_4 e^{i(x_1(s)+x_2(s))} \\ & + a_5 e^{i2x_1(s)} + a_6 e^{i2x_2(s)} + a_7 e^{i(2x_1(s)+x_2(s))} + a_8 e^{i(x_1(s)-2x_2(s))} \\ & + a_9 e^{i3x_1(s)} + a_{10} e^{i3x_2(s)} \|^2. \end{aligned} \tag{35}$$

Equation 35 contains the physically motivated "true" mechanism captured by the values of the coefficients in Eq. 34. The goal is to determine the relative magnitudes of the modulus of $a_j (j = 0, 1, \ldots, 10)$. The terms a_j of larger magnitude reveal the important quantum pathways (i.e., the mechanism); see references [23–25] for a detailed physical interpretation of such mechanisms. Taking the square modulus in Eq. 35 results in the following expression:

$$\begin{aligned} y(s) &= \sum_{j=1}^{45} w_j \phi_j(\mathbf{x}(s)) \\ &= \alpha_0 + \alpha_1 \cos(x_1(s)) + \beta_1 \sin(x_1(s)) + \alpha_2 \cos(x_2(s)) + \beta_2 \sin(x_2(s)) \\ &\quad + \alpha_3 \cos(x_1(s) - x_2(s)) + \beta_3 \sin(x_1(s) - x_2(s)) \\ &\quad + \alpha_4 \cos(x_1(s) + x_2(s)) + \beta_4 \sin(x_1(s) + x_2(s)) \\ &\quad + \alpha_5 \cos(2x_1(s)) + \beta_5 \sin(2x_1(s)) + \alpha_6 \cos(2x_2(s)) + \beta_6 \sin(2x_2(s)) \\ &\quad + \alpha_7 \cos(2x_1(s) + x_2(s)) + \beta_7 \sin(2x_1(s) + x_2(s)) \\ &\quad + \alpha_8 \cos(x_1(s) - 2x_2(s)) + \beta_8 \sin(x_1(s) - 2x_2(s)) \\ &\quad + \alpha_9 \cos(3x_1(s)) + \beta_9 \sin(3x_1(s)) + \alpha_{10} \cos(3x_2(s)) + \beta_{10} \sin(3x_2(s)) \\ &\quad + \alpha_{11} \cos(x_2(s) - 2x_1(s)) + \beta_{11} \sin(x_2(s) - 2x_1(s)) \\ &\quad + \alpha_{12} \cos(2x_2(s) + x_1(s)) + \beta_{12} \sin(2x_2(s) + x_1(s)) \\ &\quad + \alpha_{13} \cos(x_1(s) - 3x_2(s)) + \beta_{13} \sin(x_1(s) - 3x_2(s)) \\ &\quad + \alpha_{14} \cos(x_1(s) + 3x_2(s)) + \beta_{14} \sin(x_1(s) + 3x_2(s)) \\ &\quad + \alpha_{15} \cos(x_2(s) - 3x_1(s)) + \beta_{15} \sin(x_2(s) - 3x_1(s)) \\ &\quad + \alpha_{16} \cos(2x_1(s) - 2x_2(s)) + \beta_{16} \sin(2x_1(s) - 2x_2(s)) \\ &\quad + \alpha_{17} \cos(2x_1(s) + 2x_2(s)) + \beta_{17} \sin(2x_1(s) + 2x_2(s)) \\ &\quad + \alpha_{18} \cos(x_1(s) - 4x_2(s)) + \beta_{18} \sin(x_1(s) - 4x_2(s)) \\ &\quad + \alpha_{19} \cos(2x_1(s) - 3x_2(s)) + \beta_{19} \sin(2x_1(s) - 3x_2(s)) \\ &\quad + \alpha_{20} \cos(2x_2(s) - 3x_1(s)) + \beta_{20} \sin(2x_2(s) - 3x_1(s)) \\ &\quad + \alpha_{21} \cos(x_1(s) - 5x_2(s)) + \beta_{21} \sin(x_1(s) - 5x_2(s)) \\ &\quad + \alpha_{22} \cos(3x_1(s) - 3x_2(s)) + \beta_{22} \sin(3x_1(s) - 3x_2(s)). \end{aligned} \tag{36}$$

All together there are 45 unknown parameters α_0, α_j, $\beta_j (j = 1, 2, \ldots, 22)$ corresponding to the coefficients w_j in Eq. 2 with basis functions $\phi_1 = 1$ and the 44 cosine and sine functions as the remaining ϕ_j in Eq. 36. The goal of D-MORPH regression

in this context is to find the set of parameters $\alpha_0, \alpha_j, \beta_j$ that best fits the experimental data, and ideally coincides with the parameters in the true system, Eq. 34.

The two inputs, phases $x_1(s)$ and $x_2(s)$, were uniformly sampled in the laboratory over $[0, 2\pi]$ with 20 points corresponding to $N = 400$ pairs of values for $(x_1(s), x_2(s))$ $(s = 1, 2, \ldots, 400)$. The signal $y(s)$ for each pair of $(x_1(s), x_2(s))$ was measured 10 times, and the mean value of the ten measurements was used as $y(s)$ in D-MORPH regression. The relative error for each measurement of $y(s)$ compared to the mean value represents the experimental error. The average relative error obtained from the 400 points for each measurement is about 0.017. The quality of the identified quantum-control-mechanism by D-MORPH regression may be established by comparing the resultant parameters $\alpha_0, \alpha_j, \beta_j$ with those independently obtained from the experimental data by using the original HE–OD decoding procedure [24,25]. The results below will show that D-MORPH regression is far more efficient than the original HE–OD decoding procedure reflected in the number of data points needed to extract quantum control mechanism information. Moreover, if the relative error of D-MORPH regression for $y(s)$ is close to the experimental error 0.017, the resultant model from the experimental data by D-MORPH regression should be close to the true system.

The 45 basis functions in Eq. 36 are used to construct $\mathbf{\Phi}$ whose normal treatment by least-squares regression has 45 equations. For D-MORPH regression, only a portion of the equations related to $\mathbf{\Phi}_0$ are kept. Since the goal is to determine the relative magnitudes of the modulus of $a_j (j = 0, 1, \ldots, 10)$, it is sufficient to use the first 21 terms in Eq. 36 to construct $\mathbf{\Phi}_0$ because the parameters $\alpha_0, \alpha_j, \beta_j (j = 1, 2, \ldots, 10)$ in these terms already contain all $a_j (j = 0, 1, \ldots, 10)$ in Eq. 35. The coefficient matrix of Eq. 24 in the D-MORPH regression is then rectangular of dimension 21×45.

A proper choice of the cost function $\mathcal{K}$ is important for D-MORPH regression. In this example, two steps were employed to find a suitable cost function. Without using any a priori knowledge about the system, the cost functions is first set to be

$$\begin{aligned} \mathcal{K} &= \frac{1}{2} \sum_{i=1}^{45} C_{ii} w_i^2 \\ &= \frac{1}{2} \left[c_0 \alpha_0^2 + \sum_{j=1}^{22} c_j \left(\alpha_j^2 + \beta_j^2 \right) \right] \end{aligned} \tag{37}$$

with

$$c_0 = 0, \qquad c_j = |n_{j1}| + |n_{j2}| \quad (j = 1 - 22),$$

i.e., the weight c_j is proportional to the sum of absolute values of the coefficients n_{j1} and n_{j2} for x_1 and x_2 in each basis $\phi_i(\mathbf{x})$ (sine and cosine functions), which implies that the basis functions with small values of $|n_{j1}|$ and $|n_{j2}|$ have a priority in the regression. This setting is proper because larger $|n_{j1}|$ and $|n_{j2}|$ cause larger oscillations for the

sine and cosine functions and are anticipated to be less relevant on physical grounds with y then being smoother. The sine and cosine functions with the same $|n_{j1}| + |n_{j2}|$ values are given the same weight. The constant term α_0 will not shrink in D-MORPH regression as $c_0 = 0$.

Sixty points randomly chosen from the 400 data were used for training and the other 340 points were used for testing. The resultant coefficients α_j and β_j in Eq. 36 determined by D-MORPH regression are given in the first part of Table 2 labeled as "Cost function Eq. 37". The magnitudes of α_j (similarly, β_j) are measures of their significance in the model. Thus, to readily observe the significant terms in Eq. 36, all the α_j's are divided by the largest one, α_3, which is given in the column labeled as α_j/α_3. The results show that $\alpha_j (j = 1, 2, 3)$ are most significant (α_0 is the sum of all square moduli r_j^2 of a_j, and is always significant). Other α_j are small and may result from experimental noise. Using this information on the significant terms in Eq. 36, the weight c_j's in Eq. 37 are reset to be

$$c_j = 0 \quad (j = 0, 1, 2, 3), \qquad c_j = |n_{j1}| + |n_{j2}| \quad (j = 4 - 22). \tag{38}$$

Setting $c_j = 0 (j = 0, 1, 2, 3)$ avoids pressure on the parameters $\alpha_0, \alpha_j, \beta_j (j = 1, 2, 3)$ in the D-MORPH exploration. The resultant coefficients α_j and β_j are given in the second part of Table 2 labeled as "Cost function Eq. 38".

The comparison of $\alpha_0, \alpha_j, \beta_j (j = 1, 2, 3)$ obtained from r_j and φ_j determined by the original HE–OD procedure [24,25] and D-MORPH regression with the cost function in Eq. 38 is given in Table 3. The accuracy of the model obtained by D-MORPH regression with the cost function given in Eq. 38 for the training and testing data is given in Fig. 4 and Table 4. For comparison, the results obtained by least-squares regression for Eq. 36 are also given.

Table 3 shows that D-MORPH regression correctly identified the quantum-control-mechanism by accurately finding the value of the parameters $\alpha_0, \alpha_j, \beta_j (j = 1, 2, 3)$ (see Eq. 34). The other nonzero $\alpha_j, \beta_j (j > 3)$ coefficients in Table 2 may arise from experimental noise. In this example the original HE–OD procedure results reported in Table 3 utilized 400 data points, while D-MORPH regression produced comparable results with 60 data points. The advantage offered by D-MORPH regression is very important, as taking the data is an expensive laboratory operation. This capability will be useful in high-duty-cycle experiments. Another feature of D-MORPH regression is the use of iterative learning/testing methodology which permits assessing the credibility of the extracted information. This prospect may also allow the real-time performance of quantum control experiments to 'react' automatically and increase or decrease the number of required data points according to circumstances. The fitting accuracy of D-MORPH regression is also very satisfactory for both the training and testing data. In Fig. 4 and Table 4 the fitting error of least-squares regression is even smaller than the experimental error, which implies the presence of overfitting that yields poor prediction accuracy. The fitting and prediction error of D-MORPH regression is very similar to the experimental error. These results suggest that the model obtained by D-MORPH regression is close to capturing the true system structure.

Table 2 The coefficients α_j and β_j of Eq. 36 obtained by D-MORPH regression with the cost function defined by Eqs. 37 and 38

j	Cost function Eq. 37			Cost function Eq. 38		
	α_j	β_j	α_j/α_3	α_j	β_j	α_j/α_3
0	137.1068	–	–	135.4722	–	–
1	11.6228	9.3743	0.5704	16.3589	11.3782	0.6794
2	16.4405	2.2488	0.8069	18.3712	−0.7507	0.7630
3	20.3756	−3.6916	1.0000	24.0775	−6.6595	1.0000
4	3.7826	−0.9390	0.1856	3.4776	1.8688	0.1444
5	1.1755	1.3808	0.0577	−0.7344	−0.1154	−0.0305
6	−1.2945	2.2191	−0.0635	0.1729	2.1594	0.0072
7	0.7113	0.3344	0.0349	0.8015	−0.5435	0.0333
8	1.1860	−0.1619	0.0582	0.6048	−0.0181	0.0251
9	−1.4324	−1.4181	−0.0703	−0.4788	0.0714	−0.0199
10	−0.3565	−1.8157	−0.0175	−0.7085	−0.7274	−0.0294
11	2.4927	3.1882	0.1223	0.4764	0.5283	0.0198
12	1.3780	−1.8097	0.0676	−0.1687	−0.4521	−0.0070
13	−2.4412	0.5949	−0.1198	−0.3444	−0.3660	−0.0143
14	−1.8357	−1.7843	−0.0901	−0.5999	−0.3334	−0.0249
15	−0.9096	−1.7953	−0.0446	−0.1541	−0.1883	−0.0064
16	−1.8265	−2.5400	−0.0896	0.0587	0.2358	0.0024
17	2.4214	−0.8380	0.1188	−0.1760	−0.0549	−0.0073
17	−1.0266	0.4116	−0.0504	0.0358	−0.0736	0.0015
19	−1.3562	−0.9846	−0.0666	−0.5182	0.0188	−0.0215
20	−0.9586	−2.0583	−0.0470	0.4765	0.1641	0.0198
21	−0.4613	0.4905	−0.0226	−0.1918	0.0417	−0.0080
22	−0.6656	0.7299	−0.0327	−0.1907	0.4528	−0.0079

Table 3 Comparison of $\alpha_0, \alpha_j, \beta_j$ ($j = 1, 2, 3$) obtained by the normal HE–OD treatment of the experimental data to that obtained by D-MORPH regression with cost function Eq. 38

j	HE–OD*		D-MORPH	
	α_j	β_j	α_j	β_j
0	136.8830	–	135.4722	–
1	16.6667	11.0113	16.3589	11.3782
2	18.3780	−0.2510	18.3712	−0.7507
3	23.5137	−7.1320	24.0775	−6.6595

*Procedure in [24,25]

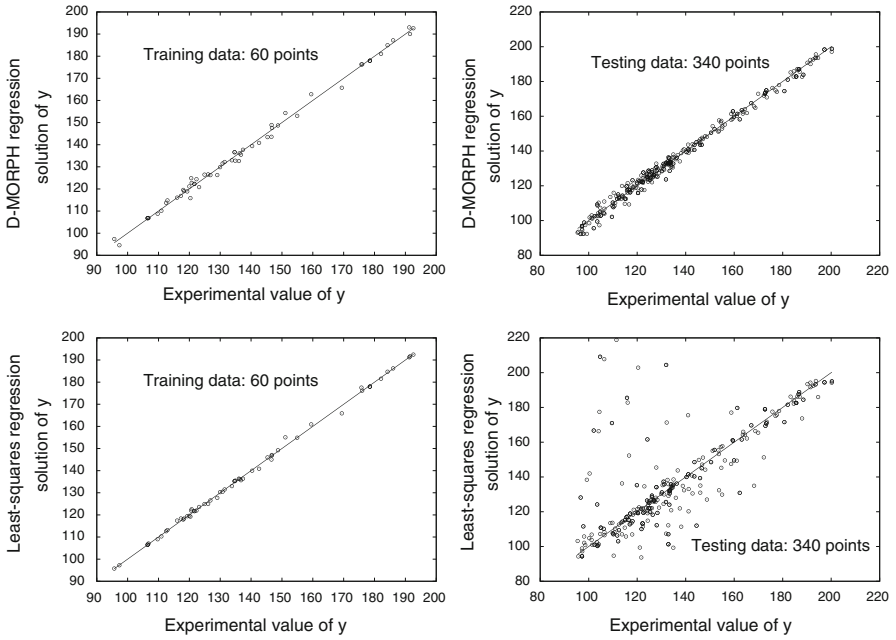

Fig. 4 The accuracy of D-MORPH and least-squares regression for Eq. 36 in the illustration of quantum-control-mechanism analysis

Table 4 The average absolute and relative errors from solving Eq. 36 obtained by D-MORPH and least-squares regression for training and testing data

	Training data		Testing data	
	Abs. err.	Rel. err.	Abs. err.	Rel. err.
D-MORPH	1.3627	0.0102	1.9615	0.0155
Least-squares	0.6405	0.0046	9.9131	0.0810

4 Conclusion

D-MORPH regression is extended to treat the common case where there is more observation data than unknown parameters. A proper subset of the normal equation for least-squares regression with an enhanced number of basis functions is used such that the final number of linear algebraic equations is less than the number of unknown parameters. D-MORPH regression is utilized to exploit this circumstance. The choice of the subset of normal equation depends on the goal. A reasonable criterion is that the fitting accuracy of the model composed of the subset of basis functions be acceptable by least-squares regression. The combustion model example used this criterion. In some cases, the choice of the subset may be determined based on physical considerations which was the case for the treatment of HE–OD experimental data. D-MORPH regression only uses matrix operations including determination of the generalized

inverse $\mathbf{\Phi}^{+}$, the inverse $(U_{m-r}^{T} V_{m-r})^{-1}$ and singular value decomposition of PC. The computation is simple, fast, and Matlab is easy to implement for these purposes [22]. The illustrations in the paper show that D-MORPH regression is superior to normal least-squares regression, least-squares regression combined with the F-test and ridge regression in terms (1) prediction accuracy, (2) required sample size, and (3) computational efficiency.

Acknowledgments Support for this work has been provided by ONR and ARO.

References

1. C.M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, NY, 2007)
2. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, NY, 2001)
3. A.N. Tikhonov, Dokl. Akad. Nauk. SSSR **39**, 195–198 (1943)
4. A.N. Tikhonov, Soviet Math. Dokl. **4**, 1035–1038 (1963). English translation of Dokl. Akad. Nauk. SSSR **151**, 501–504 (1963)
5. A.N. Tikhonov, V.A. Arsenin, *Solution of Ill-Posed Problems*. (Winston & Sons, Washington, 1977). ISBN 0-470-99124-0
6. P.C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems* (SIAM, Philadelphia, PA, 1998)
7. A.E. Hoerl, Chem. Eng. Prog. **58**, 54–59 (1962)
8. A.E. Hoerl, R. Kennard, Technometrics **12**, 55–67 (1970)
9. G. Wahba, *Spline models for observational data* (SIAM, Philadelphia, PA, 1990)
10. G. Wahba, Ann. Stat. **13**, 1378–1402 (1985)
11. G. Wahba, Y.D. Wang, C. Gu, R. Klein, B. Klein, Ann. Stat. **23**, 1865–1895 (1995)
12. Categories: linear algebra—estimation theory views, http://en.wikipedia.org/wiki/Tikhonov_regularization
13. G. Li, H. Rabitz, J. Math. Chem. **48**(4), 1010–1035 (2010)
14. N. Danielson, V. Beltrani, J. Dominy, H. Rabitz, Manuscript in preparation
15. A. Rothman, T.-S. Ho, H. Rabitz, Phys. Rev. A **72**, 023416 (2005)
16. A. Rothman, T.-S. Ho, H. Rabitz, J. Chem. Phys. **123**, 134104 (2005)
17. A. Rothman, T.-S. Ho, H. Rabitz, Phys. Rev. A **73**, 053401 (2006)
18. C.R. Rao, S.K. Mitra, *Generalized Inverse of Matrix and Its Applications* (Willey, New York, NY, 1971)
19. R. Bellman, *Introduction to Matrix Analysis*, 2nd edn. (McGraw-hill Book Co., New York, NY, 1970), p. 118
20. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in* FORTRAN—*The Art of Science Computing*, 2nd edn. (Cambridge university press, New York, NY, 1992), p. 51
21. J. Li, Z.W. Zhao, A. Kazakov, F.L. Dryer, Int. J. Chem. Kinet. **36**, 566–575 (2004)
22. Matlab [7.0R14], MathWorks, Inc. (2004)
23. A. Mitra, H. Rabitz, Phys. Rev. A **67**(3), 33407 (2003)
24. R. Rey-de-Castro, H. Rabitz, Phys. Rev. A **81**, 063422 (2010)
25. R. Rey-de-Castro, Z. Leghtas, H. Rabitz, Phys. Rev. Lett. (submitted)
26. C. Brif, R. Chakrabarti, H. Rabitz, New J. Phys. **12**, 075008 (2010)