Generic smooth connection functions: a new analytic approach to Hermite interpolation

2002 J. Phys. A: Math. Gen. 35 3877

(http://iopscience.iop.org/0305-4470/35/17/305)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **35** (2002) 3877–3898 PII: S0305-4470(02)25746-7

Generic smooth connection functions: a new analytic approach to Hermite interpolation

Alex Alon[1] **and Sven Bergmann**[2,3,4]

[1] Blurbusters Company, Masarik, Tel Aviv 64351, Israel
[2] Department of Particle Physics, Weizmann Institute of Science, Rehovot 76100, Israel

E-mail: Sven.Bergmann@weizmann.ac.il

Received 12 June 2001, in final form 4 February 2002
Published 19 April 2002
Online at stacks.iop.org/JPhysA/35/3877

Abstract

We present a new analytic approach to Hermite's interpolation problem in two dimensions. The interpolating curves are the exact solutions of a variational problem that is invariant under translations and rotations. We study the general case of functionals that are given by the integral of the curvature raised to some power ν along the curve. The parameter ν determines the importance of minimal curvature with respect to minimal length. The boundary conditions are given by the initial and final points of the curve and the tangent vectors at these points. In order to find the family of functions that obtain the minimal weight, we use extensively notions that are well known in classical mechanics. The minimization of the weight functional via the Euler–Lagrange formalism leads to a highly non-trivial differential equation. Using the symmetries of the problem it is possible to find conserved quantities that help to simplify the problem to a level where the solution functions can be written in a closed form for any given ν.

PACS numbers: 02.60.Ed, 02.30.Xx, 45.20.−d

(Some figures in this article are in colour only in the electronic version)

1. Introduction

In this work we consider the classical Hermite interpolation problem of finding a curve that goes through two points in a plane with given tangent vectors at these points [1]. We investigate

[3] Author to whom any correspondence should be addressed.
[4] Present address: Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel.

a very broad class of variational curves [2][5] which result from minimizing the integral of the curvature κ raised to some power ν along the curve:

$$\int_{t_i}^{t_f} [\kappa(t)]^{\nu} \, \mathrm{d}s(t). \tag{1}$$

The parameter ν determines the importance of minimal curvature with respect to minimal length. Solutions to the above problems have been studied for specific choices of ν. For $\nu = 0$ and $\nu = 2$ the integral is known as the *stretch* and *bend energy*, respectively. However, the exact solutions for arbitrary values of ν have not been determined yet.

Our approach to solve this problem analytically is motivated by its physical flavour. Terms such as bending or stretching already have connotations to physics, but in fact we take the analogy much further and use extensively concepts from analytical mechanics in order to obtain the curves that minimize the integral in equation (1). The crucial point is to exploit the *translational* and *rotational* symmetries of the problem that are explicit in the parametric formulation in equation (1). A well known theorem by Noether [3][6] states that each continuous symmetry is associated with a conserved quantity. In order to apply this theorem we reformulate the problem in a way that is not manifestly invariant under those symmetries, but allows us to extract the conserved *linear* and *angular momenta*. These *constants of motion* allow us to provide explicit solutions to the variational problem.

The quality of interpolation curves is relevant to many fields that deal with numerical data. In particular the importance of Hermite interpolation is well recognized in the field of computer aided geometrical design (CAGD) [4]. For computational reasons many applications related to interpolation rely on simple functions, such as Bezier curves [5][7]. However, there exists both practical and theoretical motivation to explore the exact solutions of the functional minimization of equation (1).

Obviously there are cost-related applications where the quality of the interpolation is more essential than the time needed to obtain the solution. Consider for example a spacecraft that has been launched to explore some distant planets and which is flying with constant velocity v through the interstellar space. When approaching a certain planet, one has to adjust carefully the trajectory of the spaceship, say to enter the atmosphere of the planet at a precise angle or to use its gravitational field to accelerate the spacecraft to its next destination. Suppose that for this purpose the spaceship has thrusters that exert a force F perpendicular to the direction of motion. Then at any given time t between the ignition (at t_i) and the end of the manoeuvre (at t_f) the instantaneous curvature radius $r = \kappa^{-1}$ is inversely proportional to F. Let us assume that the fuel consumption per unit time $\mathcal{W}$ is governed by some potential law, i.e. $\mathcal{W} \propto F^{\nu} \propto \kappa^{\nu}$, where ν is an empirical parameter. Then the total fuel consumption, $\mathcal{C} = \int_{t_i}^{t_f} \mathcal{W} \, \mathrm{d}t$, is proportional to the action in equation (1), since $\mathrm{d}t = \mathrm{d}s/v$. Note that once we have found the solution for the trajectory we can obtain the curvature as a function of t, which indicates how much fuel should be burned at each point.

Additional insight into standard interpolation functions can be obtained by comparison with the exact solutions presented in this paper. While a detailed analysis of this kind is beyond the scope of this paper, we present here numerical results that focus on the cubic Bezier function as an approximation to the variational curves we study.

Finally the generality of our approach makes it readily applicable to different forms of the weight functional as long as it shares the translational and rotational symmetries. In fact we believe that the physics concepts and language introduced in our work might prove useful to

[5] For the general concept and definition of variational curves see also [8].
[6] See also [7].
[7] See also chapters 1 and 4 in [4] and references therein.

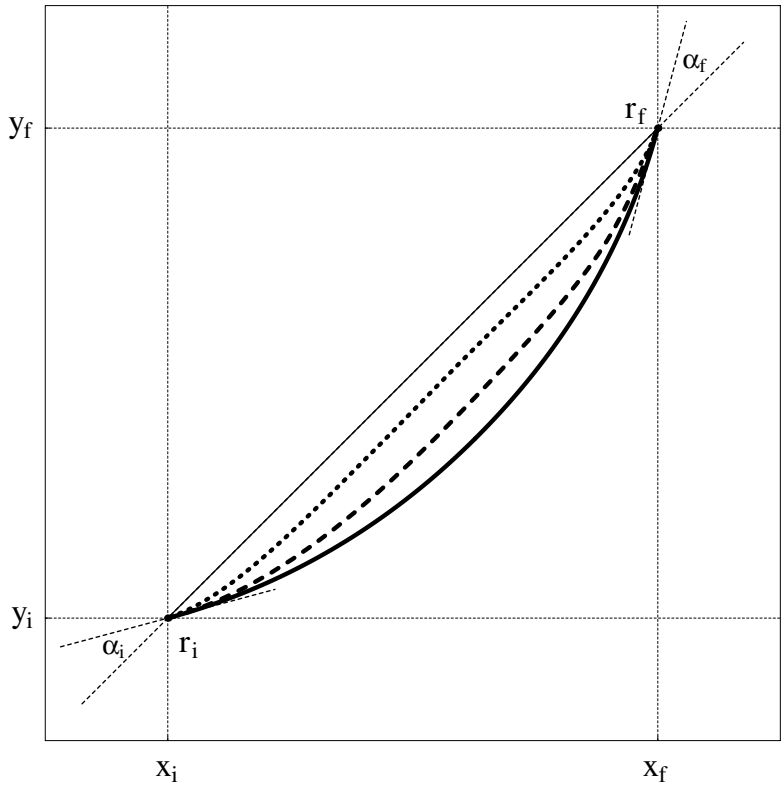

Figure 1. Different 'smooth connection functions' that connect the initial point $\boldsymbol{r}_i = (x_i, y_i)$ with the final point $\boldsymbol{r}_f = (x_f, y_f)$. At $\boldsymbol{r}_i$ and $\boldsymbol{r}_f$ the curves are tangent to the associated rays (dashed), which are specified by their inclination angles, α_i and α_f, with respect to the vector pointing from $\boldsymbol{r}_i$ to $\boldsymbol{r}_f$. The dotted curve (resulting from $\nu = 1.2$) is the shortest, but has the strongest curvature. Conversely, the solid curve ($\nu = 20$) is the longest, but almost corresponds to an arc of a circle with minimal curvature. The dashed curve ($\nu = 2$) presents a more balanced compromise between minimal length and curvature.

analyse related geometric problems and could provide powerful tools in the field of CAGD. Conversely, for the physicist it should be interesting to see some fundamental physics principles in the context of a very intuitive and visual, yet non-trivial problem.

2. Formalism

We start by introducing some notations: the boundary conditions of the Hermite interpolation problem are determined at the initial point $\boldsymbol{r}_i$ and the final point $\boldsymbol{r}_f$ of the interpolating curve. For their coordinates

$$\boldsymbol{r}_i = (x_i, y_i) \qquad \text{and} \qquad \boldsymbol{r}_f = (x_f, y_f), \tag{2}$$

we demand that $x_i < x_f$. Here the coordinates $\boldsymbol{r} = (x, y)$ are a set of two real numbers specifying any point in the plane. We use simple Euclidean geometry. The tangent vectors at $\boldsymbol{r}_i$ and $\boldsymbol{r}_f$ are described by their inclination angles, α_i and α_f, with respect to the vector pointing from $\boldsymbol{r}_i$ to $\boldsymbol{r}_f$ (cf figure 1).

We call the function $y(x)$ that interpolates between $\boldsymbol{r}_i$ and $\boldsymbol{r}_f$ such that the integral in equation (1) is minimal the 'smooth connection function' (SCF). It has to satisfy the boundary conditions

$$y(x_i) = y_i \qquad \text{and} \qquad y(x_f) = y_f, \tag{3}$$

$$y'(x_i) = \tan(\alpha_i + \alpha_0) \qquad \text{and} \qquad y'(x_f) = \tan(\alpha_f + \alpha_0), \tag{4}$$

where the prime denotes a derivative with respect to x and α_0 is the angle between the positive x-axis and the vector pointing from $\boldsymbol{r}_i$ to $\boldsymbol{r}_f$. Note that our parametrization of the curve, with one coordinate being a function of the other, breaks explicitly the equivalence of the two coordinates. Moreover, the functional description precludes looped curves. In fact the built-in limitation to curves without loops is a major advantage with respect to parametric curves. This is because without this restriction in general there is no well defined solution corresponding to the absolute minimum of the integral in equation (1) [6].

In general, any weight functional that depends only on the local curvature[8]

$$\kappa[y(x)] = \frac{1}{r[y(x)]} = \frac{|y''(x)|}{[1 + y'(x)^2]^{3/2}}, \tag{5}$$

and the infinitesimal length element along the curve

$$\mathrm{d}s(x) = \sqrt{1 + y'(x)^2}\,\mathrm{d}x, \tag{6}$$

or any other translational and rotational invariant expression, will itself be invariant under translations and rotations.

For the functional parametrization of the interpolating curve the integral introduced in equation (1) reads

$$\tilde{\mathcal{S}}[y(x)] \equiv \int_{x_i}^{x_f} (\kappa[y(x)])^{\nu}\,\mathrm{d}s(x) = \int_{x_i}^{x_f} \frac{|y''(x)|^{\nu}}{[1 + y'(x)^2]^{(3\nu-1)/2}}\,\mathrm{d}x. \tag{7}$$

We shall refer to $\tilde{\mathcal{S}}[y(x)]$ as the *action* [7]. For a given function $y(x)$ it returns the weighted sum of $\mathrm{d}s(x)$ where x goes from x_i to x_f. The weight at position x is given by the local curvature κ raised to some power ν. *A priori* this parameter is an arbitrary real number that determines how to choose the compromise between minimal curvature and minimal length. For example, for $\nu = 0$ the curvature does not play a role at all and $\tilde{\mathcal{S}}[y(x)]_{\nu=0}$ is just the length of the curve. For $|\nu| \to \infty$ the situation is exactly the opposite, since $\tilde{\mathcal{S}}[y(x)]_{|\nu|\to\infty} \simeq \int (1/r)^{\nu}\,\mathrm{d}x$ only depends on the curvature in this case. A special situation arises for $\nu = 1$, where $\tilde{\mathcal{S}}[y(x)] = \int_{x_i}^{x_f} (\mathrm{d}s/r) = \int_{x_i}^{x_f} \mathrm{d}\alpha = \alpha_f - \alpha_i$. Since the action gives just the difference between the inclination angles at the boundary, independent of the particular choice of $y(x)$, it is impossible to determine the optimal path for $\nu = 1$ and we exclude this case from our subsequent discussion.

In order to find the optimal curve we have to minimize the action under the boundary conditions in equations (3) and (4). The integrand in equation (7), usually referred to as the *Lagrangian* [7]

$$\tilde{\mathcal{L}}(y', y'') \equiv \frac{|y''(x)|^{\nu}}{[1 + y'(x)^2]^{(3\nu-1)/2}}, \tag{8}$$

only depends on $y'(x)$ and $y''(x)$, while there is no explicit dependence on $y(x)$ and x. Thus y and x enter the action $\tilde{\mathcal{S}}[y(x)]$ only via the boundary conditions. From equation (3) it follows that

$$y_f - y_i = y(x_f) - y(x_i) = \int_{x_i}^{x_f} y'(x)\,\mathrm{d}x = \text{constant}. \tag{9}$$

[8] To show the relation in equation (5) it is enough to verify that the curves that describe a circle, i.e. $y(x) = \pm\sqrt{r^2 - (x - x_0)^2} + y_0$, where r is the radius and (x_0, y_0) denotes the centre of the circle, give $r(x) = r$.

This constraint can be absorbed into the action functional by introducing a Lagrange multiplier λ, resulting in a new action

$$\mathcal{S}[y'(x)] \equiv \tilde{\mathcal{S}}[y'(x)] + \lambda \int_{x_i}^{x_f} y'(x)\,\mathrm{d}x = \int_{x_i}^{x_f} [\tilde{\mathcal{L}} + \lambda\, y'(x)]\,\mathrm{d}x. \tag{10}$$

Since there is no longer any reference to $y(x)$, but only to its first and second derivative, we can change variables,

$$q(x) = y'(x) \qquad \text{and} \qquad q'(x) = y''(x), \tag{11}$$

and write the Lagrangian corresponding to the action $\mathcal{S}[q(x)]$ in equation (10) as

$$\mathcal{L}(q, q') \equiv \tilde{\mathcal{L}}(q, q') + \lambda\, q(x) = \frac{|q'|^{\nu}}{(1+q^2)^{(3\nu-1)/2}} + \lambda\, q. \tag{12}$$

Now, to minimize $\tilde{\mathcal{S}}[y(x)]$ under the boundary condition in equation (3) is equivalent to the minimization of $\mathcal{S}[q(x)]$. A necessary condition for $\mathcal{S}[q(x)]$ to be extremal is that its first functional derivative $\delta\mathcal{S}[q(x)]/\delta q(x)$ vanishes [8]. Since $\mathcal{S}[q(x)]$ only depends on q and q', this is a well known problem. Its solution is given by the *Euler–Lagrange equation* [7]:

$$\frac{\partial \mathcal{L}}{\partial q} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial \mathcal{L}}{\partial q'} = 0. \tag{13}$$

Computing

$$\frac{\partial \mathcal{L}}{\partial q} = \frac{(1-3\nu)|q'|^{\nu} q}{(1+q^2)^{(3\nu+1)/2}} + \lambda, \tag{14}$$

$$\frac{\partial \mathcal{L}}{\partial q'} = \frac{\nu\,\sigma|q'|^{\nu-1}}{(1+q^2)^{(3\nu-1)/2}}, \tag{15}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial \mathcal{L}}{\partial q'} = \frac{\nu(\nu-1)|q'|^{\nu-2}q''}{(1+q^2)^{(3\nu-1)/2}} + \frac{\nu(1-3\nu)|q'|^{\nu}q}{(1+q^2)^{(3\nu+1)/2}}, \tag{16}$$

where $\sigma \equiv \mathrm{sign}(q')$, we obtain

$$\lambda = \frac{(\nu-1)|q'|^{\nu-2}}{(1+q^2)^{(3\nu+1)/2}}[\nu(1+q^2)q'' + (1-3\nu)(q')^2 q]. \tag{17}$$

The above equation still depends on the parameter λ. We can eliminate this dependence by differentiating equation (17) with respect to x, yielding the following *equation of motion* (EOM) [7]:

$$\begin{aligned} 0 = \frac{(\nu-1)|q'|^{\nu-3}}{(1+q^2)^{\frac{3\nu+3}{2}}}&\{(1-3\nu)[(1-3\nu q^2)(q')^4 + 2\nu q(1+q^2)(q')^2 q''] \\ &+ \nu(1+q^2)^2[(\nu-2)(q'')^2 + q'q''']\}. \end{aligned} \tag{18}$$

For all functions $q(x)$ that solve the above EOM the action $\mathcal{S}[q(x)]$ in equation (10) is *stationary*, i.e. $\delta\mathcal{S}[q(x)]/\delta q(x) = 0$. However, in order to *minimize* $\mathcal{S}[q(x)]$ the so-called Legendre condition [8],

$$0 < \frac{\partial^2 \mathcal{L}}{(\partial q')^2} = \frac{\nu(\nu-1)|q'|^{\nu-2}}{(1+q^2)^{(3\nu-1)/2}}, \tag{19}$$

also has to be satisfied. This condition arises from the second functional variation of the action with respect to $q(x)$. Note that unlike the case for the minimization of usual functions it is only a *necessary* condition for a minimum of $\mathcal{S}[q(x)]$. This is essentially because a small variation in q does not necessarily imply a small variation in q'. However in our case the sign of $\partial^2\mathcal{L}/(\partial q')^2$ only depends on the prefactor $\nu(\nu-1)$, but is independent of q and q'

even if they are taken as *independent* variables. Such a situation is referred to as a 'regular problem' [8] and it implies that the solutions of the EOM indeed give rise to a local minimum of the action if $\nu < 0$ or $\nu > 1$. Conversely for $0 < \nu < 1$ one obtains a local maximum. Thus we rule out any $0 < \nu \leqslant 1$ from our analysis and we shall only consider $\nu \leqslant 0$ and $\nu > 1$ from now on.

We note that negative values for ν are somewhat peculiar, since the corresponding SCFs avoid zero curvature along the entire curve. However, these curves cannot loop in order to minimize the action due to the functional description we use (see remark after equation (3)). As a result of these two constraints we obtain well behaved curves for $\nu < 0$.

Even though we have managed to translate the problem at hand into a differential equation, solving this equation analytically seems like a formidable task given that equation (18) is third order in q and non-linear. However, the integrated form in equation (17) gives us a hint of how to proceed. The fact that λ is a constant is related to the symmetries inherent to the problem. Studying these symmetries systematically one can simplify the problem significantly and find its solutions.

3. Symmetries

Let us examine more carefully the problem with respect to its intrinsic symmetries. The important point to realize is that when introducing the coordinates in equation (2) we have made an explicit choice of

- where to define the origin,
- how to orientate the axes and
- what units of length to use.

These choices are arbitrary and once we have found a solution to our problem we can redefine our coordinate system. Assume that $y(x)$ solves the EOM in equation (18) in some coordinate system Σ and that $\tilde{y}(\tilde{x})$ is obtained from $y(x)$ by a transformation to a different coordinate system $\tilde{\Sigma}$,

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \boldsymbol{R} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}, \tag{20}$$

where

$$\boldsymbol{R} \equiv \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{21}$$

describes a *rotation* by an angle θ, and x_0 and y_0 parametrize *translations* in the x-direction and in the y-direction, respectively. The action in equation (7) is defined in terms of the curvature radius r (cf equation (5)) and the length element $\mathrm{d}s$ in equation (6). Both quantities are manifestly invariant under rotations and translations. Therefore it is clear that the coordinate transformations in equation (20) do not change the action or the EOM derived from it. However the boundary conditions obviously have to change under $\Sigma \to \tilde{\Sigma}$.

Let us consider the infinitesimal change of the coordinate y at a given x,

$$\delta y(x) \equiv \tilde{y}(x) - y(x), \tag{22}$$

for the three types of coordinate transformation. For a translation in the y-direction

$$y_0 = \epsilon_y \ll 1, \qquad x_0 = 0, \qquad \theta = 0, \tag{23}$$

the answer is trivial. The change in y is simply

$$\delta y(x, \epsilon_y) = \epsilon_y, \tag{24}$$

implying that the changes in all the derivatives of $y(x)$ vanish:

$$\begin{aligned} &\delta q(x, \epsilon_y) = 0, \\ &\delta q'(x, \epsilon_y) = 0, \\ &\delta q''(x, \epsilon_y) = 0. \end{aligned} \tag{25}$$

However, a translation in the x-direction

$$y_0 = 0, \qquad x_0 = \epsilon_x \ll 1, \qquad \theta = 0, \tag{26}$$

is a bit more tricky. First, note that

$$\tilde{y}(\tilde{x}) = y(x). \tag{27}$$

This expresses merely the fact that the coordinate transformation in equation (20) leaves the functional behaviour of $y(x)$ invariant. It does not make any difference whether we refer to the function by $y(x)$ in the coordinate system Σ or $\tilde{y}(\tilde{x})$ in the system $\tilde{\Sigma}$. Then expanding $\tilde{y}(\tilde{x})$ around x to first order gives

$$\tilde{y}(\tilde{x}) = \tilde{y}(x) + \left.\frac{\partial y}{\partial x}\right|_x \cdot \epsilon_x, \tag{28}$$

implying that

$$\delta y(x, \epsilon_x) = -\left.\frac{\partial y}{\partial x}\right|_x \cdot \epsilon_x. \tag{29}$$

From equation (29) it follows that the corresponding changes in the derivatives of $y(x)$ are given by

$$\begin{aligned} &\delta q(x, \epsilon_x) = -q'(x)\,\epsilon_x, \\ &\delta q'(x, \epsilon_x) = -q''(x)\,\epsilon_x, \\ &\delta q''(x, \epsilon_x) = -q'''(x)\,\epsilon_x. \end{aligned} \tag{30}$$

Finally, an infinitesimal rotation

$$y_0 = 0, \qquad x_0 = 0, \qquad \theta = \epsilon_\theta, \tag{31}$$

involves both a change in x by $-y\epsilon_\theta$ and in y by $x\epsilon_\theta$. Together this results in a change of $y(x)$ by

$$\delta y(x, \epsilon_\theta) = \left(x + \left.\frac{\partial y}{\partial x}\right|_x \cdot y\right)\epsilon_\theta. \tag{32}$$

From equation (32) it follows that the changes in q, q' and q'' for an infinitesimal rotation are given by

$$\begin{aligned} &\delta q(x, \epsilon_\theta) = (1 + q^2 + yq')\,\epsilon_\theta, \\ &\delta q'(x, \epsilon_\theta) = (3qq' + yq'')\,\epsilon_\theta, \\ &\delta q''(x, \epsilon_\theta) = (3(q')^2 + 4qq'' + yq''')\,\epsilon_\theta. \end{aligned} \tag{33}$$

Let us now investigate the effect of the coordinate transformations in equation (20) on the Lagrangian $\mathcal{L}$ in equation (12). In general an infinitesimal transformation parametrized by $\epsilon \ll 1$ will induce a change

$$\delta\mathcal{L} \equiv \mathcal{L}(\tilde{q}, \tilde{q}') - \mathcal{L}(q, q'). \tag{34}$$

Provided that $\mathcal{L}$ has no explicit dependence on x, expanding $\mathcal{L}(\tilde{q}, \tilde{q}')$ around q to first order gives

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial q}\delta q + \frac{\partial\mathcal{L}}{\partial q'}\delta q' \tag{35}$$

$$= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial\mathcal{L}}{\partial q'}\delta q\right) + \left(\frac{\partial\mathcal{L}}{\partial q} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial\mathcal{L}}{\partial q'}\right)\delta q. \tag{36}$$

The second term in equation (36) vanishes for $q(x)$ that satisfy the EOM (13).

Any transformation that changes the action $\mathcal{S}[q(x)]$ in equation (10) at most by a constant is called a *symmetry transformation*. Such a transformation does not change the extremal condition for the action and therefore leaves the EOM invariant. (Note however that not all transformations that leave the EOM unchanged are symmetry transformations according to our definition.) Then, the corresponding Lagrangian could change only by a total derivative:

$$\delta\mathcal{L} = \epsilon\frac{\mathrm{d}F}{\mathrm{d}x}, \tag{37}$$

where $F(q, q')$ is a function of q and q'. Equating the two results for $\delta\mathcal{L}$ we find that

$$\epsilon\frac{\mathrm{d}F}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial\mathcal{L}}{\partial q'}\delta q\right). \tag{38}$$

We stress that the expression on the left-hand side is correct for a symmetry transformation of *any* function $y(x)$, while the right-hand side is only valid for a transformation of $y(x)$ that solves the EOM. From equation (38) it follows that the *conserved charge* defined as

$$Q = F - \frac{\partial\mathcal{L}}{\partial q'}\frac{\delta q}{\epsilon} \tag{39}$$

is a constant with respect to x, i.e. $\mathrm{d}Q/\mathrm{d}x = 0$. Consequently it has one specific value for all points on a particular solution of the EOM. Obviously Q can only be defined up to a constant. The above argument, that every continuous symmetry transformation implies a conserved charge, is known as the *Noether theorem* [3].

To be explicit let us apply it to a translation in the x direction. Using equation (30) it follows that

$$\delta\mathcal{L}(x, \epsilon_x) = -\left(\frac{\partial\mathcal{L}}{\partial q}q' + \frac{\partial\mathcal{L}}{\partial q'}q''\right)\epsilon_x \tag{40}$$

$$= -\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}x}\epsilon_x. \tag{41}$$

This means that (up to a constant) we can identify $F_x = -\mathcal{L}$. Then it follows that the conserved charge corresponding to translations in the x-direction is given by

$$P_x = \frac{\partial\mathcal{L}}{\partial q'}q' - \mathcal{L} = \frac{(\nu-1)|q'|^\nu}{(1+q^2)^{(3\nu-1)/2}} - \lambda\, q. \tag{42}$$

We call P_x the *conserved momentum in the x direction* [7]. In fact it is nothing more than the Legendre transformation of the Lagrangian with respect to q'. (If the variable of integration in equation (10) had been the time t rather the coordinate x then the equivalent Legendre transformation of the Lagrangian with respect to $\mathrm{d}q/\mathrm{d}t$ is called the *Hamiltonian* and the charge related to time invariance is the *energy* [7].)

One might guess that a similar argument for translations in the y-direction should give another conserved charge P_y, which is the *conserved momentum in the y direction* [7]. However a change in the variable $y(x)$ as in equation (24) has no effect on $q(x)$ and $q'(x)$ (see equation (25)), implying trivially that

$$\delta\mathcal{L}(x, \epsilon_y) = 0. \tag{43}$$

Thus Noether's theorem does not help in this case to derive the conserved charge. However we have already encountered another conserved quantity, which could serve as a candidate for P_y. In order to absorb the boundary condition into the action we introduced the Lagrange multiplier λ, which intuitively is related to changes in y, cf equation (10). Equation (17) states that λ is equal to a complicated function of q, q' and q'' for all x. Therefore this function is a constant of motion. To prove that indeed $P_y = \lambda$ is nontrivial, and we shall show this after discussing the conserved charge related to the rotations in equation (20).

Using equations (14) and (15) and the changes in q and q' under rotations according to equation (33) the change in the Lagrangian under an infinitesimal rotation is

$$\delta\mathcal{L}(x,\epsilon_\theta) = \left\{\left[\frac{(1-3\nu)|q'|^\nu q}{(1+q^2)^{(3\nu+1)/2}} + \lambda\right](1+q^2+yq') + \left[\frac{\nu|q'|^{\nu-1}}{(1+q^2)^{(3\nu-1)/2}}\right](3qq'+yq'')\right\}\epsilon_\theta. \tag{44}$$

It is not difficult to check that this can be rewritten as a total derivative,

$$\delta\mathcal{L}(x,\epsilon_\theta) = \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{y|q'|^\nu}{(1+q^2)^{(3\nu-1)/2}} + \lambda(x+yq)\right]\epsilon_\theta. \tag{45}$$

Thus for infinitesimal rotations the function F can be identified to be (up to a constant)

$$F_\theta = y\mathcal{L} + x\lambda. \tag{46}$$

Then the charge corresponding to the rotation symmetry, which is called the *total angular momentum* [7], is given by

$$J = F_\theta - \frac{\partial\mathcal{L}}{\partial q'}\frac{\delta q(x,\epsilon_\theta)}{\epsilon_\theta} \tag{47}$$

$$= y\mathcal{L} + x\lambda - \frac{\nu\,\sigma\,|q'|^{\nu-1}}{(1+q^2)^{(3\nu-1)/2}}(1+q^2+yq') \tag{48}$$

$$= (xP_y - yP_x) + S. \tag{49}$$

In the last step we have separated the total angular momentum into the orbital contribution

$$L = xP_y - yP_x \tag{50}$$

and the remaining term

$$S = -\frac{\nu\,\sigma\,|q'|^{\nu-1}}{(1+q^2)^{\frac{3\nu-3}{2}}} = -\frac{\nu\,\sigma}{[r(x)]^{\nu-1}}, \tag{51}$$

which we call the *spin* or *intrinsic angular momentum* [7]. The *orbital angular momentum* [7] would be the sole contribution if $y(x)$ were a straight line (as is the case for $\nu = 0$ or if $q' = 0$). A non-vanishing spin S arises for all the other curves due to their curvature. Note that it is sensitive to the sign σ of q'.

We would like to return now to our claim that P_y coincides with λ defined in equation (17). To this end let us compute the change in $P_i (i = x, y)$ induced by rotations. In general

$$\delta P_i(x,\epsilon_\theta) = \left[\frac{\partial P_i}{\partial q}(1+q^2) + \frac{\partial P_i}{\partial q'}(3qq') + \frac{\partial P_i}{\partial q''}(3(q')^2+4qq'')\right]\epsilon_\theta, \tag{52}$$

where we used the results in equation (33) and the fact that the terms proportional to y add up to the total derivative of P_i:

$$y\left[\frac{\partial P_i}{\partial q}q' + \frac{\partial P_i}{\partial q'}q'' + \frac{\partial P_i}{\partial q''}q'''\right]\epsilon_\theta = y\frac{\mathrm{d}P_i}{\mathrm{d}x}\epsilon_\theta = 0. \tag{53}$$

Assuming that indeed $P_y = \lambda$, a somewhat tedious but straightforward calculation of δP_y according to equation (52) yields

$$\delta P_y(x, \epsilon_\theta) = \frac{(\nu - 1)|q'|^{\nu-2}}{(1+q^2)^{(3\nu+1)/2}}[(1 + 3\nu q^2)(q')^2 - \nu q(1+q^2)q'']\epsilon_\theta = P_x \epsilon_\theta. \tag{54}$$

Similarly using equation (52) one shows that $\delta P_x(x, \epsilon_\theta) = -P_y \epsilon_\theta$. From the infinitesimal transformations it is clear that $\boldsymbol{p} = (P_x, P_y)$ transforms like a vector under rotations, i.e. $\tilde{\boldsymbol{p}} = \boldsymbol{R} \cdot \boldsymbol{p}$, where $\boldsymbol{R}$ is the rotation matrix defined in equation (21). It follows that the constant λ in equation (17) can indeed be identified with the conserved momentum in the y-direction.

Using the above results it is easy to compute the change in the orbital angular momentum L under infinitesimal rotations:

$$\delta L(x, \epsilon_\theta) = x \cdot \delta P_y(x, \epsilon_\theta) - [\delta y \cdot P_x(x, \epsilon_\theta) + y \cdot \delta P_x(x, \epsilon_\theta)] = -y(q P_x - P_y)\epsilon_\theta. \tag{55}$$

Note that this is the change at a *fixed* x, so of course there is no variation with respect to x. The change in the spin S under rotations is given by

$$\begin{aligned} \delta S(x, \epsilon_\theta) &= \frac{\partial S}{\partial q}\delta q(x, \epsilon_\theta) + \frac{\partial S}{\partial q'}\delta q'(x, \epsilon_\theta) \\ &= \nu(\nu - 1)\frac{y|q'|^{\nu-2}}{(1+q^2)^{(3\nu-1)/2}}[3q(q')^2 - (1+q^2)q'']\epsilon_\theta \\ &= y(q P_x - P_y)\epsilon_\theta. \end{aligned} \tag{56}$$

It follows that the changes in L and S exactly cancel each other such that the total angular momentum J does not change under rotations, i.e.

$$\delta J(x, \epsilon_\theta) = \delta L(x, \epsilon_\theta) + \delta S(x, \epsilon_\theta) = 0. \tag{57}$$

For completeness let us also compute the changes of the three conserved charges under infinitesimal translations. Since the momenta P_x and P_y do not depend explicitly on $y(x)$ they are trivially invariant under translations in the y-direction by $\delta y(x, \epsilon_x)$, due to equation (25). Infinitesimal changes in the x-directions induce $\delta y(x, \epsilon_x)$ as given in equation (29), implying that the change of the nth derivative of $y(x)$ is given by $y^{(n)}(x, \epsilon_x) = -\epsilon_x y^{(n+1)}(x)$ (cf equation (30)). Then the changes in the momenta are proportional to their total derivative, which vanishes, i.e.

$$\delta P_i(x, \epsilon_x) = -\left(\frac{\partial P_i}{\partial q}q' + \frac{\partial P_i}{\partial q'}q'' + \frac{\partial P_i}{\partial q''}q'''\right)\epsilon_x = -\frac{\mathrm{d}P_i}{\mathrm{d}x}\epsilon_x = 0. \tag{58}$$

Finally the total angular momentum J in equation (49) does depend explicitly on $y(x)$ and consequently changes under translations in the y-direction by

$$\delta J(x, \epsilon_y) = -P_x \epsilon_y. \tag{59}$$

For the change in J due to an infinitesimal translation in the x-direction we obtain an expression similar to that in equation (58):

$$\delta J(x, \epsilon_x) = -\left(\frac{\partial J}{\partial y}q + \frac{\partial J}{\partial q}q' + \frac{\partial J}{\partial q'}q'' + \frac{\partial J}{\partial q''}q'''\right)\epsilon_x = -\left(\frac{\mathrm{d}J}{\mathrm{d}x} - \frac{\partial J}{\partial x}\right)\epsilon_x = P_y \epsilon_x, \tag{60}$$

where we used the facts that $\mathrm{d}J/\mathrm{d}x = 0$ and that J depends explicitly on x. We note that the results in equations (60) and (59) are also correct for finite translations, since the changes do not depend on J.

4. Scaling and dimensional analysis

Consider a *scaling transformation* that changes x and y by a fraction a of their original value, i.e.

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + a \begin{pmatrix} x \\ y \end{pmatrix}. \tag{61}$$

The infinitesimal change in $y(x)$ at a fixed x under such a transformation for $a = \epsilon_a \ll 1$ is given by

$$\delta y(x, \epsilon_a) = \left(y - \left. \frac{\partial y}{\partial x} \right|_x x \right) \cdot \epsilon_a. \tag{62}$$

Consequently the derivatives of $y(x)$ change by

$$\begin{aligned} &\delta q(x, \epsilon_a) = -(q' x)\, \epsilon_a, \\ &\delta q'(x, \epsilon_a) = -(q'' x + q')\, \epsilon_a, \\ &\delta q^{(n)}(x, \epsilon_a) = -(q^{(n+1)} x + n q^{(n)})\, \epsilon_a. \end{aligned} \tag{63}$$

Therefore it follows that some arbitrary function A of $y(x)$ and its derivatives changes under an infinitesimal scaling transformation by

$$\delta A(x, \epsilon_a) = -\left[\left(\frac{\mathrm{d}A}{\mathrm{d}x} - \frac{\partial A}{\partial x} \right) x - \frac{\partial A}{\partial y} y + \sum_n \frac{\partial A}{\partial q^{(n)}} n q^{(n)} \right] \epsilon_a. \tag{64}$$

Using this formula one can show that the conserved charges P_x, P_y and J transform as follows under the scaling transformation:

$$\delta P_i(x, \epsilon_a) = -\nu P_i\, \epsilon_a, \tag{65}$$

$$\delta J(x, \epsilon_a) = (1 - \nu) J\, \epsilon_a, \tag{66}$$

where $i = x, y$. Note that for each charge the infinitesimal change is proportional to its original value. This implies that the finite scaling transformations are given by $P_i \to P_i \cdot \exp(-\nu a)$ and $J \to J \cdot \exp[(1 - \nu)a]$.

The interpretation of the proportionality factor follows from the following argument. Any quantity A can be written as a product of a dimensionless number and the dimension $[A]$. Since the problem we discuss is purely geometrical we only have a fundamental length scale l_0. So $[A] = l_0^{\mu}$ can be written as some power of this scale. Now the scaling transformation in equation (61) can be viewed as a change of the length scale l_0 by some fraction $\delta l_0 = \epsilon_a l_0$. Then to first order the corresponding change in A is given by

$$\delta A = \frac{\partial A}{\partial l_0} \delta l_0 = \mu A\, \epsilon_a. \tag{67}$$

Thus the proportionality factor is nothing more than the dimension of the quantity and we find that $[P_x] = [P_y] = l_0^{-\nu}$ and $[J] = l_0^{1-\nu}$. It is reassuring that these results also follow from dimensional analysis of the definitions of the conserved charges by noting that $[x] = [y] = l_0$ and $[q^{(n)}] = l_0^{-n}$. The dimension of the action $[\mathcal{S}] = l_0^{1-\nu}$ allows us to understand *a posteriori* why the regime $0 < \nu \leqslant 1$ had to be excluded from our analysis. If $\nu = 1$ the action is invariant under scaling transformations. In particular, we can shrink or magnify sections of any possible solution and thereby transform it to any arbitrary shape without changing its action. This explains why for $\nu = 1$ the action only depends on $\Delta\alpha = \alpha_f - \alpha_i$, as mentioned after equation (7). The physics jargon is to say that the action becomes 'soft' when ν approaches unity and it is 'critical' at $\nu = 1$ [9]. Now if $0 < \nu < 1$ one can always find a 'trivial solution' which is defined as follows: just follow the rays at $\boldsymbol{r}_i$ and $\boldsymbol{r}_f$ to the point where they intersect

and bend the curve in the infinitesimal vicinity of the intersection point by $\Delta\alpha$. For this curve the action vanishes, since any length element $\mathrm{d}s(x)$ of the straight part of the curve, where $r(x) = \infty$, does not contribute to the action as long as $\nu > 0$. Moreover the (infinitesimal) part of the curve that is bent also does not affect the action, because from the scaling property of the action we know that it decreases when shrinking the unit length l_0 as long as $\nu < 1$. Consequently continuously scaling down the region where the bending takes place, we achieve a zero, and hence minimal action. For $\nu < 0$ the fact that any finite straight piece of the curve gives an infinite contribution to the action precludes this solution and for $\nu > 1$ it is not viable since, due to the inverse scaling, the action blows up if the curve is bent strongly on a section of small length.

The fact that the momenta and the action have somewhat unusual (and ν-dependent) dimensions could easily be remedied by multiplying the Lagrangian in equation (12) by $l_0^{\nu-1}$. Then the action and the angular momenta would be dimensionless and the linear momenta would have dimensions of l_0^{-1}.

Using equation (64) or just applying dimensional arguments it follows that a scaling transformation on the right-hand side of the EOM in equation (18) results in a multiplication by a factor $[1 + (1 - \nu)\epsilon_a]$. However since the left-hand side is zero it is clear that the EOM is unchanged in the new coordinate system. The important point to note is that even though the EOM is invariant under scaling, the transformation in equation (61) is *not* a symmetry transformation. The reason is that the change in the Lagrangian under an infinitesimal scaling transformation,

$$\delta\mathcal{L}(x, \epsilon_a) = -\left[\frac{\mathrm{d}}{\mathrm{d}x}(x\mathcal{L}) + \mathcal{L}(1 - \nu)\right]\epsilon_a, \tag{68}$$

cannot be written as a total derivative for $\nu \neq 1$. As a consequence there is no conserved charge related to scaling.

5. The solution

The coordinate transformations discussed in the previous section are very useful for actually solving our problem: if we manage to find the SCF in some convenient coordinate system Σ, we can use translations, rotations and scaling to fit the particular solution to any boundary conditions given in some other coordinate system $\tilde{\Sigma}$. Let us consider a particular solution for which $P_x = J = 0$. Then from equations (42) and (49) it follows that

$$0 = \frac{(\nu - 1)|q'|^{\nu}}{(1 + q^2)^{(3\nu-1)/2}} - q\,P_y, \tag{69}$$

$$0 = x\,P_y - \frac{\nu\sigma|q'|^{\nu-1}}{(1 + q^2)^{(3\nu-3)/2}}. \tag{70}$$

From equation (70) we obtain

$$|q'| = \left(\frac{\sigma x P_y}{\nu}\right)^{1/(\nu-1)} (1 + q^2)^{3/2}, \tag{71}$$

provided that $\sigma x P_y/\nu$ is non-negative. This requirement implies that when x changes sign, q' also has to change its sign σ. Plugging the result for $|q'|$ into equation (69) and solving for q^2 we find

$$q^2 = \frac{C\,|x|^{2\nu/(\nu-1)}}{1 - C\,|x|^{2\nu/(\nu-1)}}, \tag{72}$$

where $C \equiv (\nu-1)^2|\nu|^{2\nu/(1-\nu)}|P_y|^{2/(\nu-1)}$ can be set to unity by choosing $P_y = \pm|\nu|^\nu/|\nu-1|^{\nu-1}$. Because q can be positive or negative there are two solutions

$$y_\pm(x) = \pm\int_{x_0}^{x}\sqrt{\frac{(\bar{x}^2)^{\nu/(\nu-1)}}{1-(\bar{x}^2)^{\nu/(\nu-1)}}}\,\mathrm{d}\bar{x} + y_\pm(x_0), \tag{73}$$

where $y_\pm(x_0)$ are the respective integration constants. Since either $\nu \leqslant 0$ or $\nu > 1$ the solution in equation (74) is defined only in the interval $-1 < x < 1$.

In the limit where $|\nu| \to \infty$ we can solve equation (74) analytically:

$$\lim_{|\nu|\to\infty} y_\pm(x) = \pm\int_{x_0}^{x}\sqrt{\frac{\bar{x}^2}{1-\bar{x}^2}}\,\mathrm{d}\bar{x} + y_\pm(x_0) = \pm\left(\sqrt{1-x_0^2}-\sqrt{1-x^2}\right) + y_\pm(x_0). \tag{74}$$

We see that in this case the SCF describes a segment of a circle, which has a constant curvature radius and therefore presents the best solution if we only care about minimal curvature along the curve. For finite values of ν the length of the curve also plays a role. The integral can be expressed in terms of hypergeometric functions $F(a, b; c; x)$. For $0 < x < 1$ we have, up to a numerical constant,

$$y_\pm(x) = \mp x\,\mathrm{Im}\left[F\left(\frac{1-\nu}{2\nu}, \frac{1}{2}; \frac{1+\nu}{2\nu}; x^{2\nu/(1-\nu)}\right)\right]. \tag{75}$$

This representation is valid provided that $c = (1+\nu)/2\nu$ is not zero or a negative integer. We note that for the corresponding values $\nu_c \in \{-1, -1/3, -1/5, -1/7, \ldots\}$ there exist closed expressions for $y_\pm(x)$ that analytically continue the expression in equation (75) for $\nu \to \nu_c$. For example for $\nu = -1$ we have (again for $0 < x < 1$ and up to a numerical constant)

$$y_\pm(x)|_{\nu=-1} = \mp x\left\{\sqrt{x-x^2} + \frac{1}{2}\arctan\left[\sqrt{x^{-1}-1}\left(\frac{2x-1}{2x-2}\right)\right]\right\}. \tag{76}$$

We show the functions $y_+(x)$ (solid) and $y_-(x)$ (dashed) for various ν in figure 2. The two branches are monotonic and we have chosen the constants of integration such that $y_\pm(\pm 1) \equiv \lim_{|x|\to 1}[y_\pm(x)] = 0$. For this choice $y_+(x)$ and $y_-(x)$ can be obtained from each other by a reflection with respect to the y-axis. Since $q(x)$ is an even function of x it follows that $y_\pm(x)$ is antisymmetric with respect to $y = y(0) \equiv y_\pm(0)$, i.e. $y_\pm(-|x|) = 2y(0) - y_\pm(|x|)$. The curvature of the SCF changes sign at $x = 0$. For large $|\nu|$ the SCF is very close to the arc of a circle. The curves for $\nu > 1$ are below the curve for $|\nu| \to \infty$ and they become 'flatter' and thus shorter for smaller values of ν. The curves for positive ν approach $x = 0$ with a vanishing slope and their curvature radius diverges at $x = 0$. The closer ν is to unity the sooner the curve approaches the value $y_\pm(0)$ when $x \to 0$ (which can be understood from the scaling behaviour of the action, cf section 4). The curves for $\nu < 0$ all reside above the curve corresponding to $|\nu| \to \infty$. For these curves both the slope $q(x)$ and the curvature radius $r(x)$ vanish at $x = 0$.

The standard SCFs shown in figure 2 are the fundamental solutions to the boundary problem we want to solve. Any specific solution consists of a segment of a standard SCF that can be viewed as a template which may be rotated, translated and scaled in order to fit the boundary conditions. Before we continue to describe in detail how this can be done, it is useful to extend the SCFs beyond the interval they are defined on.

To this end we note that from equation (72) it follows that the curvature radius, defined in equation (5), is given by

$$r(x)_\pm = |x|^{1/(1-\nu)}\cdot\frac{\nu-1}{\nu}. \tag{77}$$

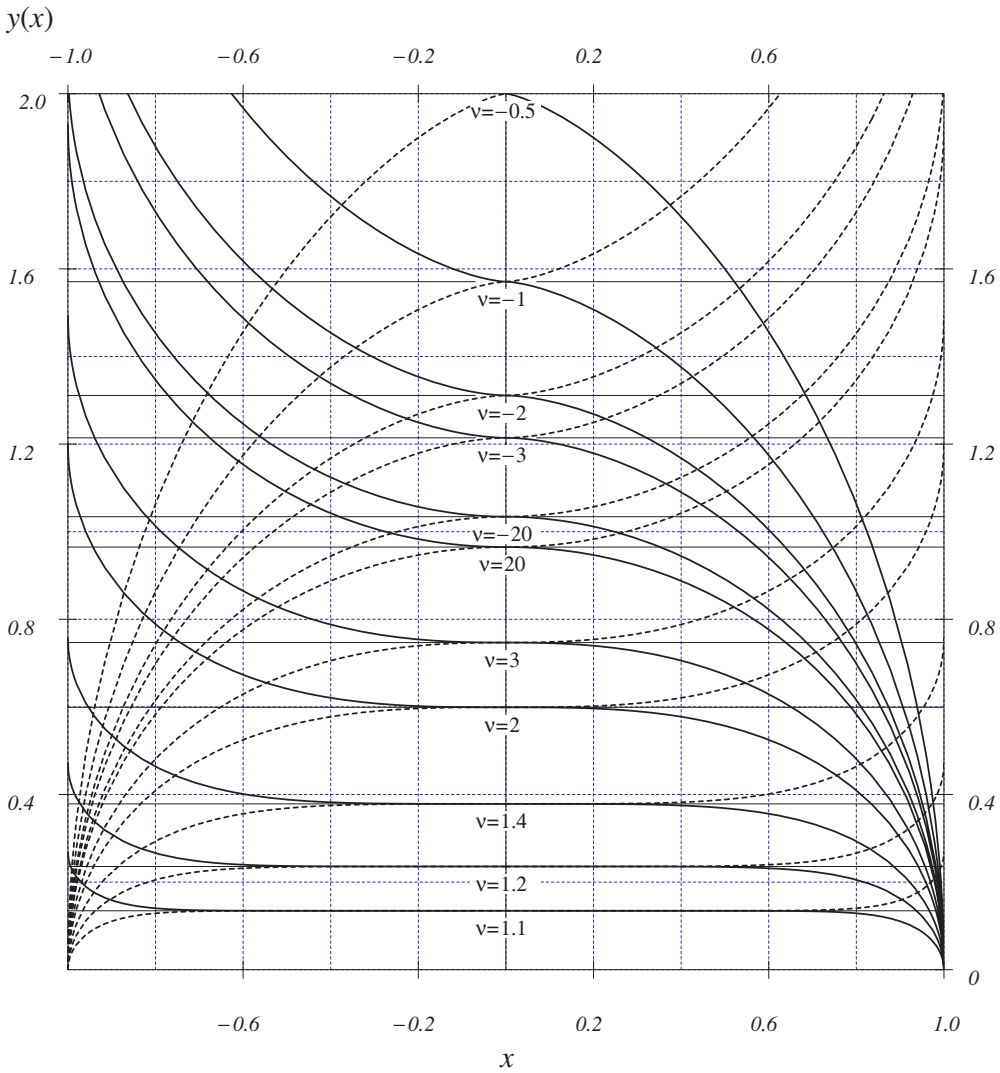

Figure 2. Standard SCFs for various ν. The two branches are monotonic and we have chosen the constants of integration such that they can be obtained from each other by a reflection with respect to the y-axis. Each SCF changes the sign of its curvature at $x = 0$. For large $|\nu|$ the SCF is very close to the arc of a circle. The curves for $\nu > 1$ are below the curve corresponding to $|\nu| \to \infty$. These curves become 'flatter' and thus shorter for smaller values of ν. They approach $x = 0$ with a vanishing slope and their curvature radius diverges at $x = 0$. The closer ν is to unity the sooner the curve approaches the value $y(0)$ when $x \to 0$. The curves for $\nu < 0$ all reside above the curve corresponding to $|\nu| \to \infty$. For these curves both the slope $q(x)$ and the curvature radius $r(x)$ vanish at $x = 0$.

It has the same value for the two branches of the SCF in equation (74). Therefore, even though $q(x)$ diverges at $|x| = 1$ the curvature radius has a well defined limit for $|x| \to 1$,

$$\lim_{|x|\to 1} r(x)_{\pm} = \frac{\nu - 1}{\nu}. \tag{78}$$

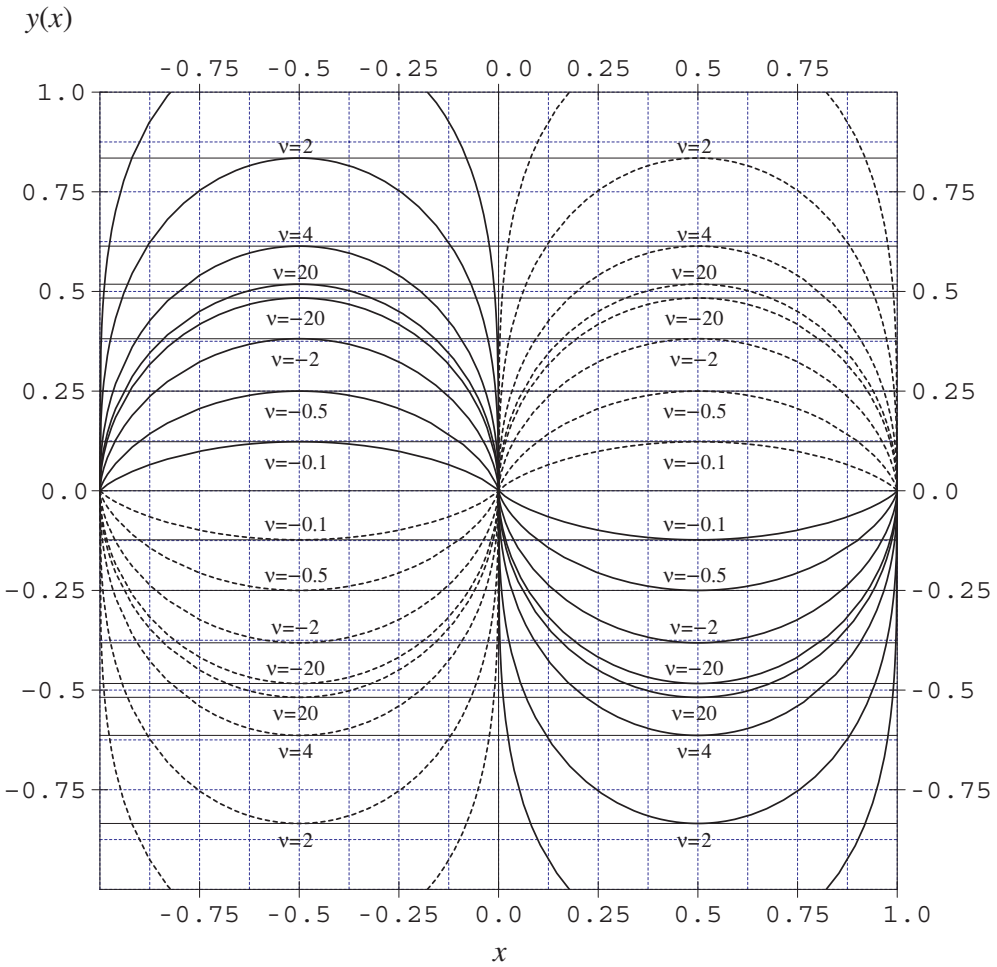

Figure 3. Extended standard SCFs for various ν. These functions are obtained from the SCFs of figure 2 by exchanging the coordinates x and y. All curves have been rescaled such that they are defined in the interval $-1 < x < 1$. The interval $-1/2 < x < 1/2$ corresponds to the SCFs of figure 2. The continuations beyond $x = \pm 1/2$ use a segment of the other branch, which is attached to the centrepiece such that the curvature radius is continuous at $x = \pm 1/2$.

Thus it is natural to connect the two solutions $y_{\pm}(x)$ to one single curve and for the following we shall fix the constant of integration such that $\lim_{|x|\to 1} y(x) = 0$ for all curves. The problem is that these curves are not single valued. In order to express them as a single function we simply exchange the coordinates x and y. The resulting curves are shown in figure 3. All curves have been rescaled such that they are defined in the interval $-1 < x < 1$. The interval $-1/2 < x < 1/2$ corresponds to the SCFs of figure 2. The continuations beyond $x = \pm 1/2$ use a segment of the other branch, which is attached to the centrepiece such that the curvature radius is continuous at $x = \pm 1/2$.

Now that we have these 'extended standard SCFs' of a given $\nu > 1$ for a particular set of values for the conserved charges, namely

$$P_x = J = 0 \qquad \text{and} \qquad P_y = \pm \frac{|\nu|^{\nu}}{|\nu - 1|^{\nu-1}}, \tag{79}$$

it is not difficult to obtain a specific SCF for *any* given boundary conditions in equations (3) and (4). The basic idea is to find first two points $\tilde{\boldsymbol{r}}_i$ and $\tilde{\boldsymbol{r}}_f$ on a 'standard SCF', where the slopes correspond to the required angles α_i and α_f, and then to apply a set of coordinate transformations to the curve in order to match the boundary conditions. The first step implies that we have to check whether it is possible to find positions $\tilde{x}_i$ and $\tilde{x}_f$ somewhere on the standard SCF such that

$$q(\tilde{x}_i) = \tan(\alpha_i + \tilde{\alpha}_0) \qquad \text{and} \qquad q(\tilde{x}_f) = \tan(\alpha_f + \tilde{\alpha}_0), \tag{80}$$

where

$$\tan\tilde{\alpha}_0 = t_0(\tilde{x}_i, \tilde{x}_f) \equiv \frac{\tilde{y}(\tilde{x}_f) - \tilde{y}(\tilde{x}_i)}{\tilde{x}_f - \tilde{x}_i}. \tag{81}$$

Using the fact that $\tan(\alpha+\beta) = (\tan\alpha+\tan\beta)/(1-\tan\alpha\tan\beta)$ we can rewrite these conditions as two coupled equations

$$\tan(\alpha_i) = t_i(\tilde{x}_i, \tilde{x}_f) \equiv \frac{q(\tilde{x}_i) - t_0(\tilde{x}_i, \tilde{x}_f)}{1 + q(\tilde{x}_i)\, t_0(\tilde{x}_i, \tilde{x}_f)}, \tag{82}$$

$$\tan(\alpha_f) = t_f(\tilde{x}_i, \tilde{x}_f) \equiv \frac{q(\tilde{x}_f) - t_0(\tilde{x}_i, \tilde{x}_f)}{1 + q(\tilde{x}_f)\, t_0(\tilde{x}_i, \tilde{x}_f)}. \tag{83}$$

This set of equations can be solved numerically[9]. For example one can apply Newton's method and use the iteration scheme

$$\begin{pmatrix} \tilde{x}_i^{(n+1)} \\ \tilde{x}_f^{(n+1)} \end{pmatrix} = \begin{pmatrix} \tilde{x}_i^{(n)} \\ \tilde{x}_f^{(n)} \end{pmatrix} - \begin{pmatrix} \partial t_i/\partial x_i & \partial t_f/\partial x_i \\ \partial t_i/\partial x_f & \partial t_f/\partial x_i \end{pmatrix}^{-1} \Bigg|_{\tilde{x}_{i,f}^{(n)}} \begin{pmatrix} t_i(\tilde{x}_{i,f}^{(n)}) - \tan(\alpha_i) \\ t_f(\tilde{x}_{i,f}^{(n)}) - \tan(\alpha_f) \end{pmatrix} \tag{84}$$

Of course such an iterative procedure will only converge provided that for a given ν one can indeed find two points on the standard SCF that satisfy equation (80). The important observation is that using the *extended* standard SCFs (as shown in figure 3) it is possible to find a suitable segment of the curves for *any* given pair of (α_i, α_f). Thus we shall use these curves in the following.

In general the endpoints $\tilde{\boldsymbol{r}}_i$ and $\tilde{\boldsymbol{r}}_f$ of the fitting segments do not satisfy the boundary conditions in equations (3) and (4). However, it is important to realize that neither translations, nor rotations, nor scaling transformations change the angles α_i and α_f. This is because they are defined relative to the line through $\tilde{\boldsymbol{r}}_i$ and $\tilde{\boldsymbol{r}}_f$. The only variable in equation (80) that does change under coordinate transformations is the angle $\tilde{\alpha}_0$ that defines the direction of this line with respect to the x-axis. Thus we can apply a rotation in order to match $\tilde{\alpha}_0$ with any given value in equation (4) for α_0. Since $\tilde{\alpha}_0$ is defined as the ratio between the differences of the coordinates of $\tilde{\boldsymbol{r}}_i$ and $\tilde{\boldsymbol{r}}_f$ (see equation (82)) it is invariant under translations and scaling transformations. This enables us also to satisfy the boundary conditions in equation (3) by first scaling the (rotated) SCF $\tilde{y}(\tilde{x})$ such that the distance between $\tilde{\boldsymbol{r}}_i$ and $\tilde{\boldsymbol{r}}_f$ coincides with the distance between $\boldsymbol{r}_i$ and $\boldsymbol{r}_f$ and then translating the resulting curve such that it connects these points. The SCF obtained like this satisfies both equations (3) and (4). In figure 4 we show a specific set of SCFs, each corresponding to a different ν, that all satisfy the same boundary conditions. Various other examples of SCFs, that all have been obtained by the numerical recipe described above, are shown in figure 5. Each plot corresponds to a particular choice of (α_i, α_f) and shows the behaviour of the solution for five different values of ν (see footnote 9).

[9] All numerical analyses in this work were performed using Mathematica 3.0. The code is available upon request from the authors.

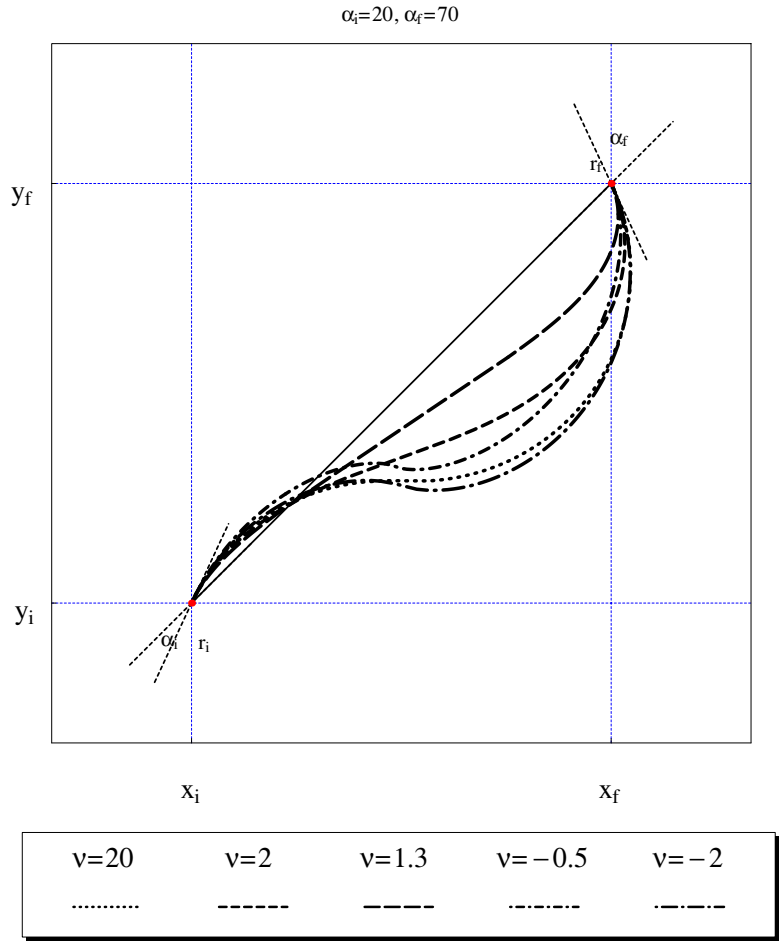

Figure 4. Example for a specific set of SCFs that all satisfy the same boundary conditions. Each curve corresponds to a different ν as indicated in the legend. All the solutions consists of a segment of an extended standard SCF that can be viewed as a template which may be rotated, translated and scaled in order to fit the boundary conditions. This example corresponds to $\alpha_i = 20^\circ$ and $\alpha_f = 70^\circ$.

6. Comparing the SCF with other curves

We would like to compare the SCF discussed in this paper with other 'conventional' interpolation functions. In particular for practical purposes it is important to know how much better the optimal path (i.e. the SCF) is with respect to some approximation. To this end it is useful to have a closed expression for the action in equation (7), which determines the 'smoothness' of any curve $y(x)$. Using equation (42) its integrand can be written in terms of the linear momenta as

$$\tilde{\mathcal{L}} = \frac{P_x + P_y\, y'(x)}{\nu - 1}. \tag{85}$$

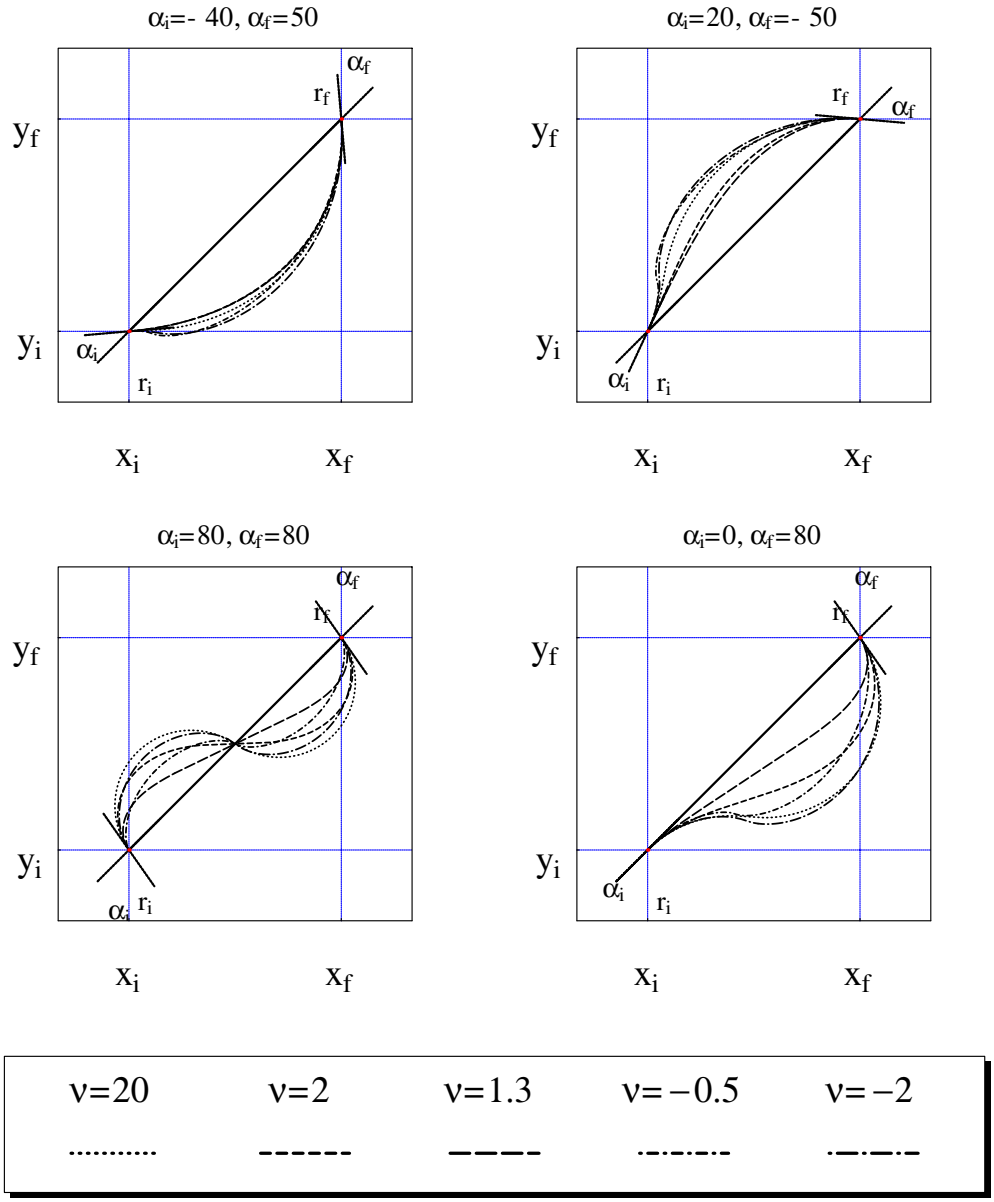

Figure 5. Examples for solving specific choices of (α_i, α_f) with the (extended) standard SCFs. Each plot corresponds to a particular choice of (α_i, α_f) and shows the behaviour of the solution for five different values of ν as indicated in the legend (in degrees).

If $y(x)$ solves the EOM, then P_x and P_y are constants and $\tilde{\mathcal{L}}$ can easily be integrated, resulting in

$$\tilde{\mathcal{S}} = \frac{P_x(x_f - x_i) + P_y(y_f - y_i)}{\nu - 1} = \frac{\boldsymbol{p} \cdot (\boldsymbol{r}_f - \boldsymbol{r}_i)}{\nu - 1}. \tag{86}$$

Note that $\tilde{\mathcal{S}}$ is manifestly invariant under rotations and translations. If $y(x)$ does not solve the EOM, the integration in equation (7) in general has to be performed numerically. Standard

interpolation curves are often given in parametric form, i.e. as $\{x(t), y(t)\}$, where t varies within a given interval $[t_i, t_f]$ along the curve. In this case the action is given by

$$\tilde{\mathcal{S}}_p = \int_{t_i}^{t_f} \frac{|\dot{x}\ddot{y} - \ddot{x}\dot{y}|^\nu}{(\dot{x}^2 + \dot{y}^2)^{(3\nu-1)/2}}\,\mathrm{d}t = \int_{t_i}^{t_f} \frac{|\dot{\boldsymbol{r}} \times \ddot{\boldsymbol{r}}|^\nu}{\dot{r}^{3\nu-1}}\,\mathrm{d}t, \tag{87}$$

where the dot denotes a derivative with respect to t. In the second expression in equation (87) the manifest rotational and translational invariance of $\tilde{\mathcal{S}}_p$ is most apparent. We note that in principle one could also minimize this action in order to determine the SCFs. However in this case the Euler–Lagrange formalism results in two coupled non-trivial differential equations, which appear to be even harder to solve than the EOM we obtained in equation (18).

A well known class of interpolations is given by the Bezier curves [5]. The shapes of these curves are related very intuitively to a set of N control points $\{\boldsymbol{c}_j\}$. $\boldsymbol{c}_1$ and $\boldsymbol{c}_N$ are the initial ($\boldsymbol{r}_i$) and final point ($\boldsymbol{r}_f$) of the curve and $\boldsymbol{c}_i = \boldsymbol{c}_2 - \boldsymbol{c}_1$ and $\boldsymbol{c}_f = \boldsymbol{c}_N - \boldsymbol{c}_{N-1}$ are tangent vectors to the curve at $\boldsymbol{r}_i$ and $\boldsymbol{r}_f$, respectively. In general the curve does not go through the intermediate control points, but it is 'dragged' towards them. The parametric form of the curve is given by the sum over the $\boldsymbol{c}_j$ times a set of polynomial basis functions. For example, the cubic Bezier curve is given by

$$\boldsymbol{b}_3(t) = (1-t)^3\boldsymbol{c}_1 + 3t(1-t)^2\boldsymbol{c}_2 + 3t^2(1-t)\boldsymbol{c}_3 + t^3\boldsymbol{c}_4, \tag{88}$$

where t goes from $t_i = 0$ to $t_f = 1$ along the curve. In this case the boundary conditions in equations (3) and (4) fix all the control parameters except the lengths of the tangent vectors $c_i = |\boldsymbol{c}_i|$ and $c_f = |\boldsymbol{c}_f|$.

An interesting question is how the weight functional behaves as a function of c_i and c_f for a given value of ν. To this end we have computed the action in equation (88) over a range of values for (c_i, c_f) for three values of ν (i.e. $\nu = \{1.1, 1.3, 2\}$ as indicated above each of the vertical panels in figure 6) (see footnote 9). We have chosen the boundary conditions such that we can compare the results for the Bezier curve with the standard SCF over its full range, i.e. $\boldsymbol{r}_{i,f} = (\mp 1, y(\mp 1))$ and $y'(\mp 1) = -\infty$. The results for the ratio between the Bezier action and the minimal action

$$\rho = \tilde{\mathcal{S}}_p(\boldsymbol{b}_3)/\tilde{\mathcal{S}} \tag{89}$$

are shown in the top panel of figure 6 as contour plots. ρ is a well behaved function of (c_i, c_f) with a global minimum $\rho_{min} > 1$. The contours are concentric around the optimal choice for (c_i, c_f), which is indicated by a dot. One can see that the plots are symmetric with respect to c_i and c_f as expected from the symmetric boundary conditions. The horizontal panel in the centre of figure 6 shows ρ as a function of $c = c_i = c_f$ (i.e. the diagonal cross-section through the contour plots). In the bottom panel of figure 6 we have plotted the optimal Bezier curves and the corresponding SCFs (dashed) for comparison. Indeed these Bezier curves present fair approximations to the SCFs. Interestingly the best result is obtained for $\nu = \nu^* \simeq 1.3$, in which case the Bezier curve has a vanishing slope at $x = 0$, just as the SCF does. For $\nu < \nu^*$ ($\nu > \nu^*$) the Bezier curves passes through $x = 0$ with a positive (negative) slope. We remark that the boundary conditions we choose present the 'worst case' scenario for the Bezier curves. In general partial segments of the standard SCF can be approximated much better by Bezier curves with ρ very close to one in most cases.

In figure 7 we show ρ_{min} as a function of ν for the above-mentioned boundary conditions (see footnote 9). One can see that $\rho_{min}(\nu)$ has a local minimum at ν^*. In the vicinity of ν^* we have $\rho_{min} \approx 1$ and the Bezier curves provide excellent approximations to the corresponding SCFs in this regime. However, for larger ν it is increasingly difficult to match the SCF by a Bezier curve and the ratio between the actions becomes significant. In fact ρ_{min} grows exponentially for $\nu \gg \nu^*$ as is apparent from the logarithmic plot. We remark that in order to

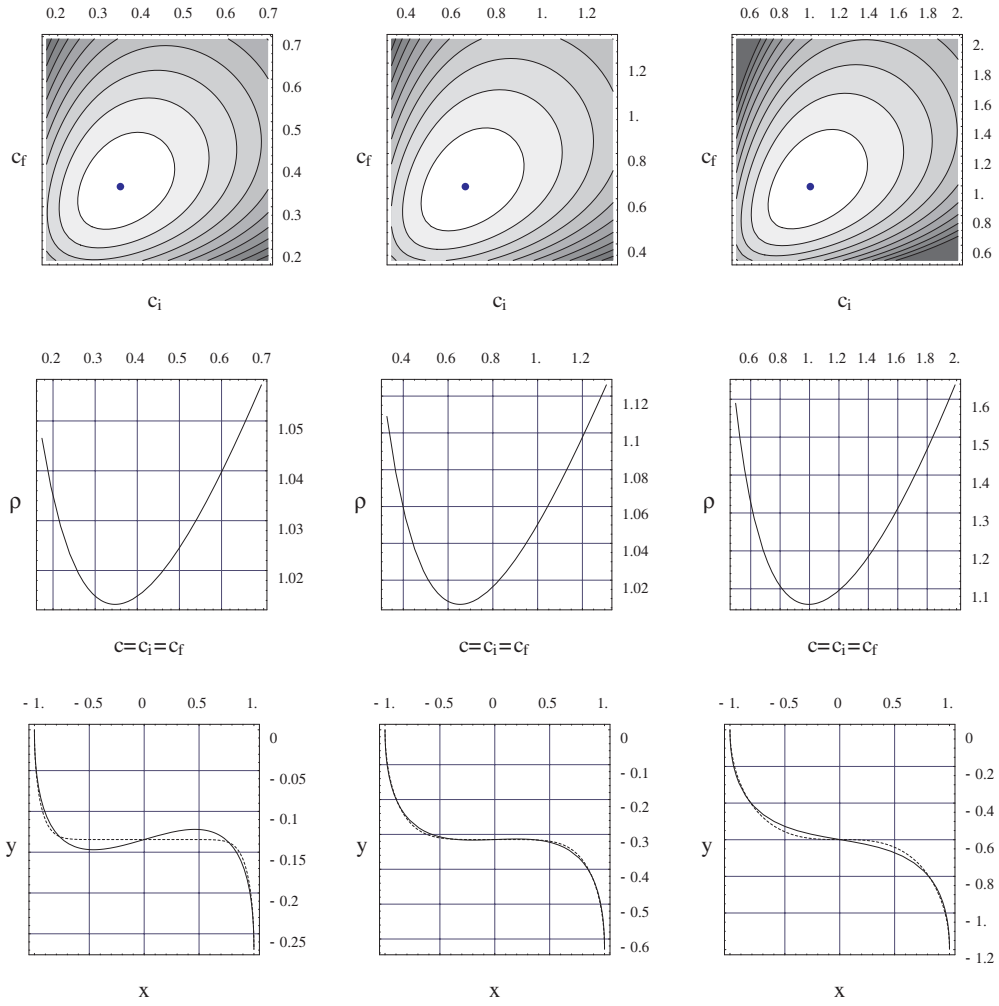

Figure 6. Bezier curves as approximations to SCFs for $\nu = \{1.1, 1.3, 2\}$ (corresponding to the left, centre and right column, respectively) as indicated above each of the vertical panels: the top panel shows the ratio ρ between the Bezier action $\tilde{\mathcal{S}}_p(\boldsymbol{b}_3)$ and the minimal action $\tilde{\mathcal{S}}$ as contour plots. ρ is a well behaved function of (c_i, c_f) with a global minimum $\rho^{min} > 1$. The contours are concentric around the optimal choice for (c_i, c_f) (indicated by a dot). The plots are symmetric with respect to c_i and c_f. The horizontal panel in the centre shows ρ as a function of $c = c_i = c_f$. In the bottom panel we have plotted the optimal Bezier curves and the corresponding SCFs (dashed).

judge the approximation in this regime one might prefer to consider the quantity $\rho^{1/\nu}$, which asymptotically approaches a constant.

The numerical results presented in figures 6 and 7 can only give a first flavour of the relation between SCFs and Bezier curves. Nevertheless they serve as an example of how to relate the fundamental variational curves we have studied in this paper to computationally more feasible curves. Clearly, in order to benefit from the accuracy of the SCFs as exact solutions to a variational problem in interactive applications efficient methods to approximate the SCFs will be needed. Detailed analyses of this kind are beyond the scope of this paper and will be pursued elsewhere.

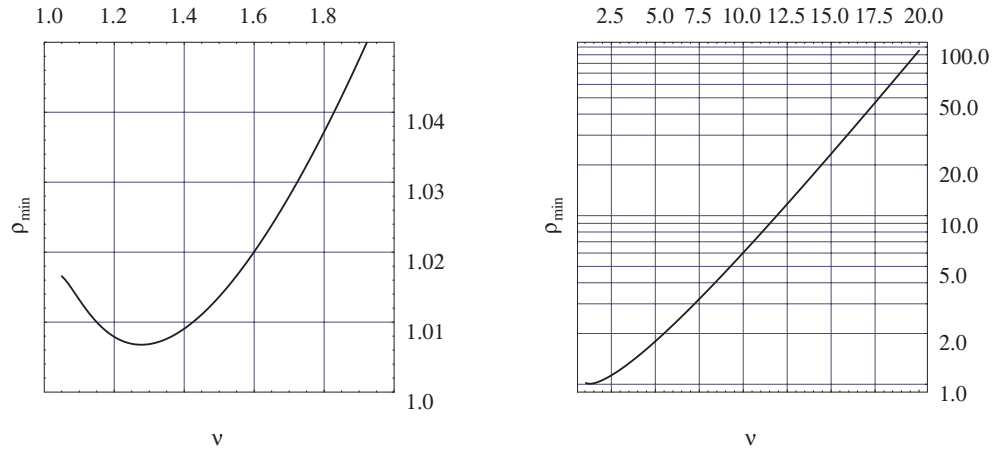

Figure 7. The ratio ρ_{min} between the optimal Bezier action $\tilde{\mathcal{S}}_p^{min}(\boldsymbol{b}_3)$ and the minimal action $\tilde{\mathcal{S}}$ is shown as a function of ν for the boundary conditions mentioned in the text. From the left-hand plot one can see that $\rho_{min}(\nu)$ has a local minimum at $\nu^* \simeq 1.3$. In the vicinity of ν^* we have $\rho_{min} \approx 1$ and the Bezier curves provide excellent approximations to the corresponding SCFs in this regime. The right-hand plot shows $\rho_{min}(\nu)$ on a log scale for a large range of ν. One can see that ρ_{min} grows exponentially for $\nu \gg \nu^*$.

7. Conclusions and outlook

We have presented a generic solution of how to connect two points in a plane by a smooth curve that goes through these points with given tangent vectors. Our approach uses extensively notions that are well known in classical mechanics. The SCF $y(x)$ has to minimize a functional that reflects the smoothness of any function by integrating the curvature to some power ν along the curve. The functional variation via the Euler–Lagrange formalism leads to a complicated non-linear third order differential equation. However the translational and rotational symmetries of the problem are of great help since they imply conserved charges, i.e. the linear and the angular momenta, that help to simplify the problem significantly. Making a specific choice for the charges it is possible to obtain the solution $y(x)$ for a given ν in terms of hypergeometric functions. Applying the appropriate coordinate transformation to this solution allows us to adjust it to arbitrary boundary conditions. A comparison of the variational curves with standard interpolation curves is of practical relevance for computational applications. We find that for a wide range of boundary conditions and not too large values of ν cubic Bezier curves can provide excellent approximations to our exact solutions.

We have worked out in detail a new formalism of how to find explicitly the variational curves that connect two points in the two-dimensional Euclidean space. Several generalizations and extensions of this basic procedure are possible. First, the number of points that define the SCF can be increased. For a single curve the solution will still be determined by four boundary conditions, but one may choose different conditions than those in equations (3) and (4). In particular one could also fix the curvature at the initial or final point. Moreover it is possible to paste together several elementary solutions in order to find interpolations between several points which cannot be achieved by a single SCF. How to do this best gives rise to a new optimization problem. Finally we note that one can also choose to apply our formalism to a different geometry. For example one may consider a time-dependent SCF $x(t)$ in Minkowski space, where the metric is defined via $r^2 = x^2 - t^2$. Also generalization of the problem to

higher-dimensional spaces is conceivable. In this case the SCF would describe some smooth manifold that connects two extended objects (such as strings). In particular an extension to three dimensions would be of practical interest (e.g. for CAGD).

Acknowledgments

We thank J Schiff for useful comments on the manuscript and the references. We are grateful to Rina and Orly for their support throughout this work.

References

[1] De Boor C, Höllig K and Sabin M 1987 High accuracy geometric Hermite interpolation *CAGD* **4** 269–78
Farin G 1993 *Curves and Surfaces for Computer Aided Geometric Design: a Practical Guide* 3rd edn (San Diego, CA: Academic)
Höllig K and Koch J 1995 Geometric Hermite interpolation *CAGD* **12** 567–80
Höllig K and Koch J 1996 Geometric Hermite interpolation with maximal order ang smoothness *CAGD* **13** 681–95
Lianghong X and Jianhong S 2001 Geometric Hermite interpolation for space curves *CAGD* **18** 817–29
As an introductory text on Hermite interpolation see for example Scheid F 1988 *Schaum's Outline of Numerical Analysis* 2nd edn (New York: McGraw-Hill) ch 10
[2] See for example
Brunnett G, Hagen H and Santarelli P 1993 Variational design of curves and surfaces *Surv. Math. Ind.* **3** 1–27
Wesselink W and Veltkamp R C 1995 Interactive design of constrained variational curves *CAGD* **12** 533–46
[3] Noether E 1918 *Nachrichten Gesell. (Wissenschaft, Göttingen)* **2** 235
[4] A comprehensive introduction to CAGD with extensive references to the literature is given by
Farin G, Hoschek J and Kim M-S (eds) *Handbook of Computer Aided Geometric Design* 2002 (Amsterdam: Elsevier Science)
(the online manuscript is currently available at http://cagd.snu.ac.kr/main.html)
[5] Bézier P 1971 Example of an existing system in motor industry: the unisurf system *Proc. R. Soc.* A **321** 207–18
Forrest A 1972 Interactive interpolation and approximation by Bézier polynomials *Comput. J.* **15** 71–9
Reprinted in Forrest A 1990 *CAD* **22** 527–37
[6] Birkhoff G and De Boor C 1965 Piecewise polynomial interpolation and approximation *Approximation of Functions* ed H L Garabedian (Amsterdam: Elsevier) pp 164–90
[7] Our favourite textbook on analytical mechanics is
Goldstein H 1980 *Classical Mechanics* 2nd edn (Reading, MA: Addison-Wesley)
[8] Pars L A 1962 *An Introduction to the Calculus of Variations* (London: Heinemann)
[9] See for example
Peskin M E and Schroeder D V 1995 *Introduction to Quantum Field Theory* (Reading, MA: Addison-Wesley) ch 8