The optimal path across barriers

1993 J. Phys. A: Math. Gen. 26 3249

(http://iopscience.iop.org/0305-4470/26/13/027)

J. Phys. A: Math. Gen. 26 (1993) 3249–3266. Printed in the UK

The optimal path across barriers

Ulf Larsen

Niels Bohr Institute, Copenhagen University, Ørsted Laboratory, Universitetsparken 5, DK 2100 Copenhagen Ø, Denmark

Received 22 April 1992

Abstract. In activated stochastic processes—such as diffusion in a condensed medium, reaction kinetics, the evolution of the genetic constitution of a population, or the motion of the index cases in a pandemic—there are barriers separating initial and final states. This work investigates the nature of the optimal path between states, such as the path followed by the quickest diffusers, and the timescale associated with it; i.e. the time required to establish communication—and the manner in which that communication takes place.

The optimal path is a property of a stochastic matrix defining transition rates in terms of the Pauli–Kolmogorov master equation. The path can therefore be computed, as shown in a numerical case study.

In the continuum limit, optimizing a timescale expressed in terms of integrals produces a new category of variational problems. This generalization introduces non-local quantities, and the equations of motion become integro-differential. A new type of mechanics emerges—with both efficient and final causes—a kinematics/dynamics with unusual implications. The motion of a fictitious 'tracer' particle, which serves to trace out the optimal path, can be formulated as a Newtonian dynamics in the same fields as influence the actual diffusers. That is, one can have Newton's second law with real efficient causes.

But the tracer's charge, regulating its coupling to the external field, varies along the path, in a manner which requires complete 'knowledge' of the full path. This teleological aspect is a consequence of the way in which the problem is defined—in terms of boundary conditions at the beginning and at the end of the path.

A computer-assisted search for optimal paths shows a complicated optimization problem with false minima. Results corresponding to a low and a high temperature indicate that:

(i) Optimal paths tend to cross barriers close to saddle points, and with negative charge.

(ii) At a lower temperature the path comes closer to the exact saddle-point.

(iii) There is a tendency to mount a slope by a steepest-descent path, and then to traverse the slope.

(iv) When traversing, the tracer chooses a level about kT below the saddle-point energy.

These findings agree with the results of the variational analysis. Some confirm general expectations, but most of the details are perhaps rather surprising.

1. Introduction

Many stochastic processes consist in diffusive motion from one location to another—in some appropriately defined 'state space'.

For instance, in reaction kinetics the system may move between states of different composition (e.g. Wolynes 1989). Or, in a condensed medium a mobile entity may diffuse, more or less freely, from one place to another (e.g. Stoneham 1989). The genetic constitution of a population may evolve stochastically in a diffusive manner (e.g. Kauffmann 1989a, b, Perelson and Kauffmann 1991).

0305-4470/93/133249+18$07.50

In *activated* processes there are *barriers* separating initial and final states. For this reason one is interested in the nature of the *optimal path*, such as the reaction coordinate, and in the *timescale* associated with it†.

I have therefore (Larsen 1987) defined this category of problem as the optimization of a certain stochastic timescale: *In a typical realization of the stochastic process, which path is followed by the quickest diffusers*?

In the presence of barriers—generally understood—timescales can become exceedingly long, and also extremely sensitive to parameter values. The most well known case being, of course, the law of thermal activation due to Arrhenius, where a slight change in temperature can alter timescales by orders of magnitude (Arrhenius 1889, 1899, Kramers 1940, Brinkman 1956, Larsen 1983a, b). Changing the temperature alters the optimal path *configuration* as well (an example of this is shown in the case-study described in section 5). Other changes of circumstance may have catalytic effects, modifying the barrier structure.

So the prospects for direct Monte Carlo style simulation are perhaps not too promising when the decisive events are, by the nature of things, highly exceptional.

In the present timescale optimization *the optimal path is a property of a stochastic matrix* defining the transition rates in the model (section 2). The path can therefore be computed—at least in principle.

1.1. Tracks and paths in diffusion

Consider a large number of repeated performances of the stochastic process, in which a 'diffuser' starts at a certain location *a* and moves according to certain transition rates, at random, between different locations, until it ends at a given location *b*.

Each realization of the process defines a *track*, from *a* to *b*. The set of locations visited (once, twice, etc...) by the track constitutes a *path*‡.

In a sufficiently large ensemble every path contains many different tracks. And, of course, every conceivable track is assigned to a path. If one imagines a diffuser 'deviating' from some path, then the new track belongs to another path which is also given consideration during the optimization (see figure 1).

Transition rates are fixed. So each path defines its own stochastic subprocess—the one which would be followed if all other transition rates, into and out of the path, were set to zero. Therefore each path can be evaluated as if it were a one-dimensional stochastic process; and the timescale for a path-process will apply to the *typical* track on that path.

The idea—to be described in more detail in section 2—is to consider transition probabilities per unit time into a variety of 'channels', some of which theoretically become closed upon path formation. The rate of proceeding to sites situated along the path remains unchanged. But the 'path-diffusers' constitute the sub-ensemble of diffusers which remain stationary at a given site, before proceeding (or backtracking) along the chosen path, whereas among all diffusers some move off into non-path directions. Within each path-ensemble no diffuser is influenced by closed channels, since the path-tracks have never used those options. In particular, their timescales depend exclusively on the transition rates of the open channels along the path.

† For example, in a pandemic it seems important to correctly forecast which path would be traced by the so-called index cases, perhaps on the basis of model calculations.

‡ This terminology agrees closely with standard English. What in physics often goes by the name 'path' is actually a 'track'. Here the distinction is necessary.

For instance, to estimate which path is most likely to include the track of the typical first-arrival at b, one must search for the path from a to b with the shortest timescale. That timescale is the typical shortest transit time†.

All paths with loops can be disregarded. For, in a Markovian stochastic process conditions revert to what they were once the diffuser resturns to the straight path. The full ensemble contains tracks similar to the looping ones, except for the missing loops. And those diffusers wasted no time looping.

Two paths are almost equally fast when they differ only in certain areas with little influence on the overall timescale. But at the col of a barrier the tracing can be extremely significant. So optimization gives a unique path, apart from symmetries that may lead to networking, i.e. alternative paths being equally fast.

1.2. Continuum limit yields new variational problem

The diffusive stochastic processes are defined in terms of the Pauli-Kolmogorov master equation, to be used in the numerical case study (section 5). The timescale analysis—under the proper circumstances—admits a continuum limit, in which the optimal path is found by optimizing a timescale quantity expressed in terms of integrals (Larsen 1983b).

This appears to produce a new category of variational problems—an analogy to Fermat's principle of least time (de Fermat 1657).

Traditional variational problems give rise to the well known Euler differential equations (e.g. Courant and Hilbert 1953). In the present case, as will be shown in section 3, one obtains integro-differential equations.

1.3. New type of mechanics emerges—with both efficient and final causes

The pragmatic background to this development was described above. In addition, it turns out that the present generalized variational problem defines a kinematics/dynamics with unusual implications. I find these interesting enough that they should be mentioned in connection with the physical aspects from which they arise. But only concepts with a precise counterpart in the mathematical formulation are being discussed here.

Consider the motion of a fictitious 'tracer' particle, which serves to trace out the optimal path. One can *formulate this motion as a Newtonian dynamics in the same fields as those which influence the actual diffusers.* That is, one can have Newton's second law (section 4), with real efficient causes.

But there are—necessarily—strange variations on this classical theme. One is that the tracer behaves in an 'intelligent' fashion: the tracer's charge, regulating its coupling to the external field, varies along the path. And it varies in a manner which requires complete 'knowledge' of the full path.

A *final cause* is normally expected only of consciously determined travellers. The present teleological aspect is a consequence of the way in which the problem is defined—in terms of boundary conditions at the beginning *and* at the end of the path‡.

† A so-called first-passage situation, in which b is absorbing (i.e. no outgoing transitions are allowed). This widely discussed topic is a special case of the present theory.

‡ In this way the situation is similar to the one behind the anthropic cosmological principle (Barrow and Tipler 1986). Earlier work on final boundary conditions includes Schrödinger (1931), Aharonov, Bergmann and Lebowitz (1964), Cocke (1967), Wheeler (1979), and the present idea (Larsen 1987).

In stochastic terms appropriate to the original problem of diffusion there is selection, whether artificial or natural. This predetermination to end up (quickly) at a certain final destination reflects back on the path-tracer's dynamics while it is travelling. The tracer's charge is changing in such a way as to serve the 'purpose' of getting the tracer to where it is supposed to turn up—whatever real fields may intervene while it is underway.

For instance, suppose a path has to cross an energy-barrier, which repels the diffusing entities in terms of a Newtonian force. The tracer negotiates the barrier by changing the sign of its charge at a certain stage. It 'decides' to mount the barrier in spite of the otherwise opposing field, 'in order to' arrive at the other side. Once over the top, the tracer's charge eventually reverts to normal, the barrier force bringing it down the opposite slope to where it is supposed to go. The charge must be very carefully adjusted all the way along such a tracing trek, in order to achieve its end. But only 'natural' forces act.

2. Background

2.1. Discrete stochastic process in continuous time—the master equation

Suppose the set $\{a\}$ labels the states of a system, and that there is a matrix of transition rates $\mathbb{L}=\{L_{ab}\}$, with

$$\sum_a L_{ab}=0 \qquad \text{for all } b. \tag{1}$$

Here L_{ab} is the probability per unit time for the jump $b \to a$. The sum-constraint in (1) means that the matrix $\mathbb{L}$ is infinitesimally stochastic; i.e. the transitions conserve the probability normalization $\Sigma_a W_a = 1$ in the associated *Pauli-Kolmogorov master equation*

$$\frac{dW_a}{dt}=\sum_b L_{ab}W_b \tag{2}$$

where $\{W_a\}$ are occupation probabilites for the states $\{a\}$. This is a Markovian process in continuous time, and—by the nature of our project—it should be assumed that $\mathbb{L}$ is time-independent.

A *path* $\{\alpha\}$ is defined as an ordered sequence of n elements drawn arbitrarily from $\{a\}$, *starting and ending at definite states a and b. With the path* $\{\alpha\}$ *one is to associate* transition rates $\{P_{\alpha\beta}\}$ drawn in this way from $\mathbb{L}$: (i) $P_{\alpha\pm1\alpha}=L_{\alpha\pm1\alpha}$ and (ii) normalize, according to (1), by choosing $P_{\alpha\alpha}=-L_{\alpha+1\alpha}-L_{\alpha-1\alpha}$; except at the ends, where $P_{11}=-P_{21}$ and $P_{nn}=-P_{n-1n}$.

For the chosen path $\{\alpha\}$ of n states one then has a tri-diagonal rate matrix $\mathbb{P}=\{P_{\alpha\beta}\}$. This path-matrix defines timescales† which pertain to the selected path 'in $\mathbb{L}$'. The physical motivation being that, in a well defined way, the optimal path characterizes the matrix $\mathbb{L}$.

Given $\mathbb{P}$ one *may* create an auxiliary master equation

$$\frac{dw_\alpha}{dt}=\sum_\beta P_{\alpha\beta}w_\beta \qquad \text{where} \sum_\alpha P_{\alpha\beta}=0 \text{ for } \beta=1,\ldots,n \tag{3}$$

† In this work I assume detailed balance, so these timescales are equal to the reciprocals of $n-1$ real positive eigenvalues of $-\mathbb{P}$; a few such standard results are reviewed in Larsen (1984a).

for a fictitious 'path-system'. But the occupation probabilities $\{w_\alpha\}$ bear no direct relation to the actual ones $\{W_a\}$. The timescale of $\mathbb{P}$ determines the approach to 'path-equilibrium'.

But it is the path timescales, *as such*, which are of interest, and how they characterize the original $\mathbb{L}$. Not global equilibrium of (2), nor the time it might take for the process (1) to establish it. The aim is to investigate the time required to establish *communication between the states a and b*—and the manner in which that specific communication takes place.

Ergodic theory, under such conditions on $\mathbb{P}$ as path connectedness, proves the existence of a path 'equilibrium state' $\{w^0_\alpha\}$ for (3). But it has nothing to say about how long it takes to reach that equilibrium (e.g. Seneta 1973). To know that duration, of course, is a prerequisite for the present optimization project.

Some time ago I found a general expression for an upper bound τ on the timescales of $\mathbb{P}$ (Larsen 1984a, b). This allows to formulate 'optimal network kinetics' for a given $\mathbb{L}$:

The optimal network, connecting selected states a and b, for a given rate matrix $\mathbb{L}$ consists in one or more paths $\{\alpha\}$ in $\{a\}$ which minimize τ, where τ is a certain path-function of $\mathbb{L}$.

Rate matrices, such as $\mathbb{L}$ or $\mathbb{P}$, have at least one zero eigenvalue (due to the sum constraint shown in (1) and (3)). That prevents standard matrix theorems (cf Marcus and Minc 1964) from giving precise information about the magnitude of the next-to-smallest eigenvalue. The dominant timescale is controlled by this eigenvalue—the largest time-quantity defined by $\mathbb{L}$ or $\mathbb{P}$. What matters is, now one has the *finite upper* bound τ on that timescale.

Information about some *average* timescale is not specific enough for the present purpose: the average may not be significantly influenced by the largest member when n is large. In the absence of an explicit *general* expression for the relevant eigenvalue, to optimize with τ—the upper bound—is a feasible approach†.

From the mathematical point of view it is the structure of τ which is interesting.

2.2. Classical diffusion in the continuum limit

In the classical limit (details in Larsen 1983b) it is assumed that $\{a\}$ becomes a d-dimensional Euclidean manifold $\{x\}$. A path $\{\alpha\}$ becomes a curve $\{x(s)\}$, with parameter s, connecting points $a = x(s_1)$ and $b = x(s_2)$. Under the influence of a force derivable from a potential energy $U(x)$, diffusive motion is assumed to be due to contact with an energy-reservoir at a temperature T‡.

The relevant timescale is then given by

$$\tau = \int_{s_1}^{s_2} \mathrm{d}s\, v \frac{H_- H_+}{HW}. \tag{4}$$

† There is a very considerable literature which handles barrier problems in terms of specific models, approximations, and potentials—far too extensive to be reviewed here. What is required for path optimization is an exact expression which holds for arbitrary potentials.

‡ But the object system of diffusing entities is *not* in equilibrium, so it has no T. The temperature parameter serves exclusively to characterize the transition probabilities in $\mathbb{L}$ which result from contact with the energy reservoir. The Boltzmann-Gibbs exponentials which occur in the following all come from $\mathbb{L}$—via the detailed balance relation (cf footnote 13 or Larsen 1984a).

Here the 'speed' is $v=(v\cdot v)^{1/2}$, $v=\mathrm{d}x(s)/\mathrm{d}s$, where s is the 'time-parameter'. Furthermore

$$W=\mathrm{e}^{-U(x)/kT}$$

$$H_-=\int_{s_1}^{s}\mathrm{d}t\,vW \tag{5}$$

$$H_+=\int_{s}^{s_2}\mathrm{d}t\,vW \qquad H=H_-+H_+=\int_{s_1}^{s_2}\mathrm{d}t\,vW.$$

The parameter t enters via x and v

The minimizing curve $x(s)$ connecting given positions $x(s_1)$ and $x(s_2)$ in $\{x\}$ is the optimal path associated with the energy $U(x)$ at temperature T. It may be said that here the function $U(x)$ represents the matrix $\mathbb{L}$. In this special case the velocity dependence is rather trivial. But I keep a general v-dependence in the variational situation suggested by this example.

3. Variational procedure

3.1. General formulation

Consider the problem to find the extremal paths $x(s)$ of the path functional

$$\tau=\int_{s_1}^{s_2}\mathrm{d}s\,f(s,x,v,p,q,r) \tag{6}$$

with endpoints fixed at $x_1=x(s_1)$ and at $x_2=x(s_2)$. Letting $g=g(t,x,v)$ be another function of the path, in (6)

$$p=\int_{s_1}^{s}\mathrm{d}t\,g \qquad q=\int_{s}^{s_2}\mathrm{d}t\,g \qquad r=p+q=\int_{s_1}^{s_2}\mathrm{d}t\,g. \tag{7}$$

It is convenient to keep r as a separate variable, and to abstain from the straightforward generalization to three different g functions.

Without the *global* quantities p, q, and r, one has a standard variational problem. Global dependences do not occur in dynamics, where all path dependence is in the local variables x and v.

This local nature of the classical variational formulations may be one reason that global optimization, as implied by variational principles, is not currently seen as a physically meaningful alternative to local dynamics. In design problems the viewpoint may be otherwise. But here one seems to be faced with a fundamentally global situation, as will become evident in the following.

Denote first derivatives (gradient vectors where x and v are concerned) as follows

$$f_x\equiv\frac{\partial f}{\partial x} \qquad f_v\equiv\frac{\partial f}{\partial v} \tag{8}$$

$$f_p\equiv\frac{\partial f}{\partial p} \qquad f_q\equiv\frac{\partial f}{\partial q} \qquad f_r\equiv\frac{\partial f}{\partial r} \tag{9}$$

and analogously for the derivatives of g.

Consider variations $x \to x + \delta x$. So

$$\delta\tau = \int_{s_1}^{s_2} \mathrm{d}s\, \delta\boldsymbol{x}\cdot\boldsymbol{Q} + \text{higher order terms.}$$

One has an extremal (the fundamental lemma) if

$$\boldsymbol{Q} = 0. \tag{10}$$

Expanding f to first order

$$\delta f = \delta\boldsymbol{x}\cdot f_x + \delta\boldsymbol{v}\cdot f_v + \delta p\, f_p + \delta q\, f_q + \delta r\, f_r. \tag{11}$$

Here, using $\delta\boldsymbol{v} = \mathrm{d}\,\delta\boldsymbol{x}/\mathrm{d}s$,

$$\delta p = \int_{s_1}^{s} \mathrm{d}t(\delta\boldsymbol{x}\cdot g_x + \delta\boldsymbol{v}\cdot g_v) = \int_{s_1}^{s} \mathrm{d}t\, \delta\boldsymbol{x}\cdot\left(g_x - \frac{\mathrm{d}g_v}{\mathrm{d}t}\right) + \delta\boldsymbol{x}\cdot g_v \tag{12}$$

$$\delta q = \int_{s}^{s_2} \mathrm{d}t(\delta\boldsymbol{x}\cdot g_x + \delta\boldsymbol{v}\cdot g_v) = \int_{s}^{s_2} \mathrm{d}t\, \delta\boldsymbol{x}\cdot\left(g_x - \frac{\mathrm{d}g_v}{\mathrm{d}t}\right) - \delta\boldsymbol{x}\cdot g_v \tag{13}$$

$$\delta r = \int_{s_1}^{s_2} \mathrm{d}t(\delta\boldsymbol{x}\cdot g_x + \delta\boldsymbol{v}\cdot g_v) = \int_{s_1}^{s_2} \mathrm{d}t\, \delta\boldsymbol{x}\cdot\left(g_x - \frac{\mathrm{d}g_v}{\mathrm{d}t}\right). \tag{14}$$

The first two terms in (11) give the conventional Euler equation. The next two terms have double integrals which we rearrange according to

$$\int_{s_1}^{s_2} \mathrm{d}s \int_{s_1}^{s} \mathrm{d}t \ldots = \int_{s_1}^{s_2} \mathrm{d}t \int_{t}^{s_2} \mathrm{d}s \ldots \qquad \text{and} \qquad \int_{s_1}^{s_2} \mathrm{d}s \int_{s}^{s_2} \mathrm{d}t \ldots = \int_{s_1}^{s_2} \mathrm{d}t \int_{s_1}^{t} \mathrm{d}s \ldots$$

thus

$$\int_{s_1}^{s_2} \mathrm{d}s\, f_p\, \delta p = \int_{s_1}^{s_2} \mathrm{d}s\, \delta\boldsymbol{x}\cdot\left[\left(g_x - \frac{\mathrm{d}g_v}{\mathrm{d}s}\right)\int_{s}^{s_2} \mathrm{d}t\, f_p + f_p g_v\right] \tag{15}$$

and likewise for the δq integral. The δr integral is straightforward.

The 'equation of motion' thus becomes

$$\boldsymbol{Q} = f_x - \frac{\mathrm{d}f_v}{\mathrm{d}s} + M\left(g_x - \frac{\mathrm{d}g_v}{\mathrm{d}s}\right) + g_v(f_p - f_q) = 0 \tag{16}$$

where

$$M = \int_{s}^{s_2} \mathrm{d}t\, f_p + \int_{s_1}^{s} \mathrm{d}t\, f_q + \int_{s_1}^{s_2} \mathrm{d}t\, f_r. \tag{17}$$

This is, of course, a necessary—but not sufficient—condition for a minimal τ. Sufficient conditions are beyond the scope of the present investigation. Inspection is probably more efficient, in connection with numerical analysis in analogous lattice models (cf section 5).

Equation (16) is the integro-differential equation announced in the introduction. The quantity M is in general expressed in terms of multiple integrals, via f_p, f_q, and f_r. The following specialization demonstrates this extension explicitly†.

3.2. Special case: the geometric path

In (4) and (5) the functions f and g depend on v only in terms of the path-geometry factor $v=(\boldsymbol{v}\cdot\boldsymbol{v})^{1/2}$, that is

$$\begin{aligned} f &= v(\boldsymbol{v})h(s, x, p, q, r) \\ g &= v(\boldsymbol{v})k(s, x). \end{aligned} \tag{18}$$

Since

$$\frac{\partial v}{\partial \boldsymbol{v}}=\frac{\boldsymbol{v}}{v}\equiv\hat{\boldsymbol{t}}$$

is the unit vector tangential to the path one gets

$$\begin{aligned} f_x &= vh_x \qquad & f_v &= h\hat{\boldsymbol{t}} \\ g_x &= vk_x \qquad & g_v &= k\hat{\boldsymbol{t}} \\ f_p &= vh_p \qquad & f_q &= vh_q \qquad & f_r &= vh_r. \end{aligned} \tag{19}$$

Define transverse components, for instance the transverse acceleration

$$\boldsymbol{a}^{\perp}\equiv\boldsymbol{a}-(\hat{\boldsymbol{t}}\cdot\boldsymbol{a})\hat{\boldsymbol{t}} \qquad \boldsymbol{a}=\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}=\frac{\mathrm{d}^2\boldsymbol{x}}{\mathrm{d}t^2}. \tag{20}$$

Then

$$\frac{\mathrm{d}\hat{\boldsymbol{t}}}{\mathrm{d}s}=\frac{\boldsymbol{a}^{\perp}}{v} \tag{21}$$

$$\frac{\mathrm{d}h}{\mathrm{d}s}=\frac{\partial h}{\partial s}+\boldsymbol{v}\cdot h_{\boldsymbol{x}}+h_p\frac{\partial p}{\partial s}+h_q\frac{\partial q}{\partial s}. \tag{22}$$

In these terms

$$\frac{\mathrm{d}f_v}{\mathrm{d}s}=\frac{\mathrm{d}h}{\mathrm{d}s}\hat{\boldsymbol{t}}+h\frac{\mathrm{d}\hat{\boldsymbol{t}}}{\mathrm{d}s}=\left[\frac{\partial h}{\partial s}+\boldsymbol{v}\cdot h_{\boldsymbol{x}}+vk(h_p-h_q)\right]\hat{\boldsymbol{t}}+h\frac{\boldsymbol{a}^{\perp}}{v} \tag{23}$$

$$\frac{\mathrm{d}g_v}{\mathrm{d}s}=\frac{\mathrm{d}k}{\mathrm{d}s}\hat{\boldsymbol{t}}+k\frac{\mathrm{d}\hat{\boldsymbol{t}}}{\mathrm{d}s}=\left[\frac{\partial k}{\partial s}+\boldsymbol{v}\cdot k_{\boldsymbol{x}}\right]\hat{\boldsymbol{t}}+k\frac{\boldsymbol{a}^{\perp}}{v}. \tag{24}$$

The last term in (16) is, by (19),

$$g_v(f_p-f_q)=kv(h_p-h_q)\hat{\boldsymbol{t}}$$

and cancels the third term in (23).

† It might be expected that a generalization of the present nature had been considered before. As far as I am aware, this is not the case. With one remarkable exception: In '*Methodus inveniendi*...'—in Chapter III—Euler considers a situation rather like the present one. The equation of motion (16) can be obtained by a generalization of his *Proposition* III (Euler 1744). It appears that no one found any use for the idea, until now, so that it has been ignored.

Now define transverse components

$$h_x^{\perp} \equiv h_x - (\hat{t}\cdot h_x)\hat{t} \qquad k_x^{\perp} \equiv k_x - (\hat{t}\cdot k_x)\hat{t}. \tag{25}$$

Then one gets

$$Q = v(h_x^{\perp} + Mk_x^{\perp}) - (h + Mk)\frac{a^{\perp}}{v} - \left(\frac{\partial h}{\partial x} + M\frac{\partial k}{\partial s}\right)\hat{t}. \tag{26}$$

Now separate transverse and longitudinal components in $Q = 0$:

$$\left(\frac{\partial h}{\partial s} + M\frac{\partial k}{\partial s}\right) = 0 \tag{27}$$

$$(h + Mk)a^{\perp} = v^2(h_x^{\perp} + Mk_x^{\perp}). \tag{28}$$

These are the 'equations of motion' under the conditions arranged in (18).

If, in addition, h and k do not depend explicitly on s—as in the case-study of the next section—then (27) is an identity. There remains no inherent time-dependence, and one is free to choose the 'natural parametrization': $v = 1$, i.e. independent of s. One gets an essentially geometric problem: to find the optimal path as a geometrically defined curve.

Yet one may imagine a 'tracer moving in time s', according to the equation of motion (28), which determines the transverse acceleration at any point. The longitudinal speed is arbitrary. In the natural parametrization the tracer's speed is constant $v = 1$, and the longitudinal acceleration is zero. Any other motion in the arbitrary 'time' s has the factor v^2 in (28) as compensation, so that the entire geometric path remains the same.

4. Case study: the optimal path in thermal diffusion

4.1. Tracer kinematics and dynamics

In this model, according to (4) and (5),

$$h = \frac{pq}{rW} \qquad k = W = \mathrm{e}^{-U/kT}. \tag{29}$$

The potential energy $U(x)$ defines the mechanical force

$$F = -U_x = -\frac{\partial U}{\partial x}. \tag{30}$$

Then

$$h_x = -\frac{h}{kT}F \qquad k_x = \frac{k}{kT}F. \tag{31}$$

Define the *charge*, ε, by

$$\varepsilon = \frac{Mk - h}{Mk + h}. \tag{32}$$

From (28) one then gets the *equation of motion*

$$ma^{\perp} = \varepsilon F^{\perp} \qquad m = \frac{kT}{v^2}. \tag{33}$$

Using $p+q=r$ one can write

$$M=\int_s^{s_2} \mathrm{d}t\, vh_p+\int_{s_1}^{s} \mathrm{d}t\, vh_q+\int_{s_1}^{s_2} \mathrm{d}t\, vh_r=\frac{1}{r^2}\left[\int_{s_1}^{s} \mathrm{d}t\, v\frac{p^2}{W}+\int_s^{s_2} \mathrm{d}t\, v\frac{q^2}{W}\right]$$
$$p=\int_{s_1}^{s} \mathrm{d}t\, vW \qquad q=\int_s^{s_2} \mathrm{d}t\, vW \qquad r=\int_{s_1}^{s_2} \mathrm{d}t\, vW. \tag{34}$$

Define

$$\chi=\tfrac{1}{2}\ln\frac{rW^2M}{pq}=\tfrac{1}{2}\ln\frac{\int_{s_1}^{s} \mathrm{d}t\, v(p^2/W)+\int_s^{s_2} \mathrm{d}t\, v(q^2/W)}{pqr}-\frac{U}{kT}. \tag{35}$$

Then

$$\varepsilon=\tanh\chi \qquad |\varepsilon|\leqslant 1. \tag{36}$$

The equation of motion (33)—together with the appropriate initial conditions—describes how the *tracer* traces the extremal path. The charge ε regulates the tracer's coupling to the external force field F, which is the same field as the one experienced by the ensemble of diffusing entities.

Consider first the option which corresponds to the natural parametrization: $v=c=$ constant. The acceleration is purely transversal: $a^{\parallel}=0$.

This dynamics is reminiscent of ultra-relativistic motion in an inertial frame at rest with respect to the static potential energy $U(x)$. But in that case m would be the 'relativistic mass' E/c^2, where E varies along the path—essentially as the kinetic energy would: $\mathrm{d}E/\mathrm{d}t=\varepsilon F\cdot v$ (Møller 1952).

In the present case $E=kT$ is fixed, no matter what ε is and no matter where the tracer is in the field F. So the transversality does not imply relativistic motion.

Yet it seems suggestive that $E=kT$, as if there was a way to exchange kinetic energy (the reservoir at T!) just so that throughout the tracer's path changes in the potential energy get compensated, making $\mathrm{d}E/\mathrm{d}t=0$†.

4.2. Exotic mechanics

Since s is an arbitrary parametrization, it is more interesting to give up the constancy of v, and the attendant transversality. Using t as a parameter, let us now *request Newton's second law*

$$\frac{\mathrm{d}p}{\mathrm{d}t}=\varepsilon F \qquad p=mv. \tag{37}$$

Here p is the tracer's momentum, and m is an appropriately variable mass of which it is *a priori* required merely that it depend on v only through the speed v.

Whereas in special relativity m as a function of v is determined by the requirement that Newton's law hold in all inertial frames-of-reference, *here we have another constraint*: the transverse equation of motion (33). It will be shown that this *implies*

$$ma^{\parallel}=-\varepsilon F^{\parallel} \qquad m=\frac{kT}{v^2} \tag{38}$$

in agreement with the second part in (33).

† For the real, diffusing entities the lack of energy conservation *is* due to exchanges with the energy reservoir—without which barriers cannot be crossed.

Consequently

$$m\boldsymbol{a} = \varepsilon(\boldsymbol{F}^{\perp} - \boldsymbol{F}^{\parallel}). \tag{39}$$

The vector in the right-hand side is the force $\varepsilon\boldsymbol{F}$ mirror inverted in the plane normal to the tangent of the orbit†.

To prove these statements:

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \hat{\boldsymbol{t}} \cdot \boldsymbol{a}$$

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = m\boldsymbol{a} + m\boldsymbol{v}\frac{\mathrm{d}\ln m}{\mathrm{d}t} = m\boldsymbol{a} - 2m\frac{\mathrm{d}v}{\mathrm{d}t}\,\hat{\boldsymbol{t}} = m\boldsymbol{a} - 2m(\hat{\boldsymbol{t}} \cdot \boldsymbol{a})\,\hat{\boldsymbol{t}}$$

$$= m(\boldsymbol{a}^{\perp} - \boldsymbol{a}^{\parallel}) = \varepsilon(\boldsymbol{F}^{\perp} + \boldsymbol{F}^{\parallel}) = \varepsilon\boldsymbol{F}.$$

The momentum decreases as v increases

$$\boldsymbol{p} = m\boldsymbol{v} = \frac{kT}{v}\,\hat{\boldsymbol{t}} \qquad \boldsymbol{p}\cdot\boldsymbol{v} = kT \tag{40}$$

due to 'mass loss'. The path *radius-of-curvature* is regulated by the charge ε:

$$\rho = \frac{v^2}{a^{\perp}} = \frac{kT}{|\varepsilon|F^{\perp}} \geqslant \frac{kT}{F} \qquad F \equiv |\boldsymbol{F}|. \tag{41}$$

From (38) one has

$$m\boldsymbol{v}\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = m\boldsymbol{a}\cdot\boldsymbol{v} = -\varepsilon\boldsymbol{F}^{\parallel}\cdot\boldsymbol{v} = -\varepsilon\boldsymbol{F}\cdot\boldsymbol{v} = \varepsilon\frac{\mathrm{d}U}{\mathrm{d}t}$$

which gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[kT\ln\frac{c}{v} + \varepsilon U\right] = U\frac{\mathrm{d}\varepsilon}{\mathrm{d}t} \tag{42}$$

where c is some reference speed. So, for constant ε one has a constant of the motion

$$kT\ln\frac{c}{v} + \varepsilon U = \text{constant}. \tag{43}$$

For instance, if $\varepsilon = 1$ the speed increases uphill (and the momentum decreases)

$$v = v_1\,\mathrm{e}^{(U-U_1)/kT} \qquad \text{for } \varepsilon = 1. \tag{44}$$

This is exotic.

4.2. The tracer adjusts its charge

The tracer of an optimal path is a determined traveller, who uses negative values of the charge ε when needed to cross barriers. The expression for ε given in (35) and (36) involves a full view of the entire path—beforehand. Besides the conventional efficient cause $\boldsymbol{F}$, one could say that, via the intrinsic quantity ε the tracer dynamics incorporates a final cause as well.

† Acceleration which is not collinear with the force is, of course, typical of variable-mass dynamics. In special relativity $E/c^2\boldsymbol{a} = \boldsymbol{F} - \boldsymbol{v}\,(\boldsymbol{F}\cdot\boldsymbol{v})/c^2$ (Møller 1952). But the present form (39) looks unique.

For example, consider a path which starts and ends at two local energy minima: for $s \simeq s_1$ and $s \simeq s_2$

$$U(x(s)) \simeq U_1 + \tfrac{1}{2}U_1''(s-s_1)^2 \qquad \text{for } U_1'' > 0$$

$$U(x(s)) \simeq U_2 + \tfrac{1}{2}U_2''(s-s_2)^2 \qquad \text{for } U_2'' > 0$$

where $U_2 > U_1$, for definiteness. At some $s \simeq s_m$ there exists—it is assumed—a single maximum in U, considered as a function of s along the path—i.e. a barrier: for $s \simeq s_m$

$$U(x(s)) \simeq U_m + \tfrac{1}{2}U_m''(s-s_m)^2 \qquad \text{for } U_m'' < 0 \text{ and } U_m > U_2.$$

For simplicity, use $v = \text{constant} = 1$. Laplace's method gives, asymptotically as $\beta = 1/kT \to \infty$

$$\begin{aligned} p(s) &\sim \tfrac{1}{2}\,\mathrm{e}^{-\beta U_1}\sqrt{2\pi/\beta U_1''} \\ q(s) &\sim \tfrac{1}{2}\,\mathrm{e}^{-\beta U_2}\sqrt{2\pi/\beta U_2''} \ll p(s). \end{aligned} \tag{45}$$

Using $r = p + q$, one gets in an analogus way

$$\begin{aligned} \tau &= \int_{s_1}^{s_2} \mathrm{d}s\, \frac{pq}{Wr} \sim \frac{p(s_m)q(s_m)}{r} \int_{s_1}^{s_2} \mathrm{d}s\, \mathrm{e}^{\beta U_m - \frac{1}{2}\beta|U_m''|(s-s_m)^2} \\ &\sim \frac{\pi kT}{\sqrt{U_2''|U_m''|}}\, \mathrm{e}^{(U_m - U_2)/kT}. \end{aligned} \tag{46}$$

The decisive quantity is the barrier height, as seen from the side where the barrier is *lowest*†. Consequently the optimal path will pass through—or at least very near to—a saddle point of the energy surface (as the path shown in figures 2-3). Reaction

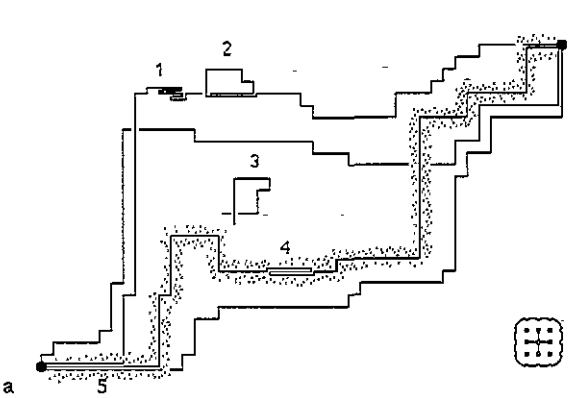

Figure 1. Some tracks and a path communicating between two states *a* and *b* on a square lattice with nearest-neighbour selection rules for transitions. Tracks belonging to a path (shaded) may backtrack and overlap, as indicated at (1) and (4). Tracks belonging to different paths may overlap partly, as at (5). Every track belongs to some path. Optimal paths, and their tracks, have no loops, as at (2) and (3), for a Markov process has no memory.

† The reason being that, in order to establish a non-directional path-equilibrium the decisive processes will be those which increase the population in the low valley by transfering entities from the high valley. One can also handle directional processes by the present methods. In that case the relevant activation-energy will be the barrier height as seen in the direction chosen. Such directed paths turn out rather similar to the present ones, but with interesting variations.

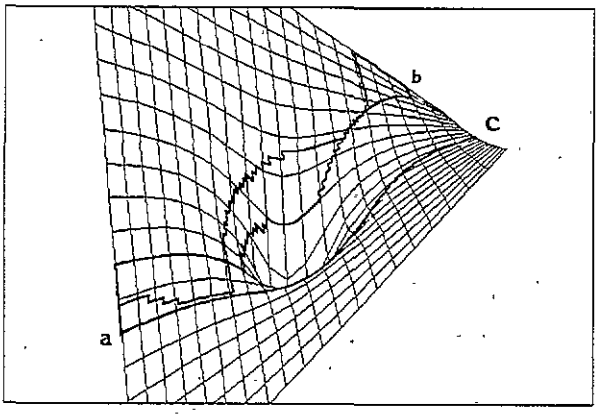

Figure 2. Optimal paths for thermal diffusion on the energy surface to be studied in section 5. The paths cross the barrier close to the saddle-points, and avoid the energy minimum.

coordinates are commonly assumed to do just this, so (46) may be said to prove a prototype *Arrhenius law* concerning the energy and temperature dependence of τ. No form of 'quasi-equilibrium transition state' was assumed to exist near the barrier saddle point. The *activation energy* is $U_m - U_2$.

To see the variation of ε, first note that $\varepsilon = 1$ at both $s = s_1$ and $s = s_2$, since either $p = 0$ or $q = 0$. At the highest point, $s \simeq s_m$:

$$r^2 M = \int_{s_1}^{s_m} \mathrm{d}t\, \mathrm{e}^{\beta U} p^2 + \int_{s_m}^{s_2} \mathrm{d}t\, \mathrm{e}^{\beta U} q^2 \sim \tfrac{1}{2}\, \mathrm{e}^{\beta U_m} \sqrt{2\pi/\beta |U_m''|}\, [p(s_m)^2 + q(s_m)^2].$$

Thus

$$\chi \sim \tfrac{1}{2}\beta(U_2 - U_m) + \tfrac{1}{4} \ln \frac{U_2''}{|U_m''|} < 0 \tag{47}$$

As $\beta \to \infty$

$$\varepsilon \to -1 \qquad \text{at barrier top.} \tag{48}$$

So there is a point on either slope where $\chi = 0$ implies $\varepsilon = 0$. In fact, as will become evident in the following section, *there are generally long stretches where the tracer moves with* $\varepsilon \simeq 0$. Accordng to (41), one expects little curvature of the path in such places. The tracer ignores the force field $\boldsymbol{F}$. To cross the barrier—once it has arrived in its vicinity—the tracer for a while turns its charge negative, using the force field to gain the necessary lift.

4.3. Example: unhindered diffusion

One can check a few things by considering $U(\boldsymbol{x}) = \text{constant}$, i.e. $\boldsymbol{F} = 0$. The optimal path is a straight line from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$. With $W = 1$ one finds ($v = 1$)

$$\varepsilon = 1 + 6s(s-1) \qquad s_1 = 0 \text{ and } s_2 = 1. \tag{49}$$

Here $\varepsilon = 0$ at $s = \frac{1}{2}(1 \pm 1/\sqrt{3})$, $\varepsilon = 1$ at $s = 0$ and 1, and $\varepsilon = -\frac{1}{2}$ at $s = \frac{1}{2}$. This relic variation of ε is consistent with the low temperature pattern deduced above.

5. Lattice model

5.1. Discrete formulation

A physically correct discrete lattice version is available from the outset—the original quantized model—so it only remains to devise a sensible expression for the charge ε. It will be modelled after τ.

But a lattice model necessarily breaks the isotropy of the continuous manifold $\{x\}$, depending on the selection rules chosen in defining the transition rate matrix $\mathbb{L}$. The present two-dimensional square lattice is 'rather anisotropic'. Yet the features of interest are perfectly evident†.

Figure 2 and 3 show the energy surface

$$U(x, y) = xy - \frac{50}{2 + x^2 + y^2} \qquad \{x\} = \{(x, y) \text{ Euclidean}\} \tag{50}$$

which has a barrier along $x = y$ with a local minimum at $(x, y) = (0, 0)$. There are two saddle-points, at $(2, -2)$ and $(-2, 2)$.

Lattice points are placed at

$$\{a\} = \{(x, y) = (\tfrac{1}{10}n_x, \tfrac{1}{10}n_y) | n_x, n_y = 0, \pm 1, \pm 2, \ldots\}.$$

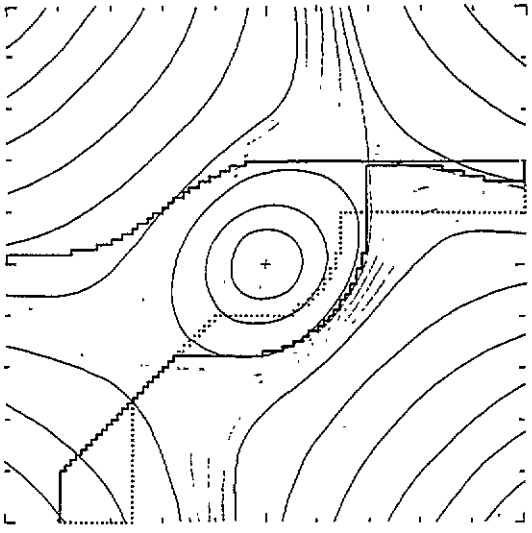

Figure 3. Optimal paths for the energy surface (50), as listed in table 1. Full lines: $kT = \frac{2}{3}$; dotted line: $kT = \frac{20}{3}$. As in figure 2, a lattice (of spacing 0.1) is superposed on the $U(x, y)$ whose contour map is shown. Note that there is no barrier between a and c, but here too traversing the slopes of the central pit is required.

† Isotropy can be improved by including transitions between next-nearest neighbours, etc. Presently, at least, such complications would obscure the essential points.

Transitions are restricted to nearest neighbours in the x- and y-directions. Rates—when not zero—are equal to standard Monte Carlo rates, with detailed balance relative to thermal equilibrium†, i.e.

$$L_{ab}=\begin{cases}1 & \text{if } U_a \leqq U_b \\ e^{(U_b-U_a)/kT} & \text{if } U_a > U_b.\end{cases} \tag{51}$$

On the path $\{\alpha\}$ define

$$w_\alpha^0 = e^{-U_\alpha/kT}$$

$$z=\sum_{\beta=1}^{n} w_\beta^0 \qquad z(1,\alpha)=\sum_{\beta=1}^{\alpha} w_\beta^0 \qquad z(\alpha+1,n)=\sum_{\beta=\alpha+1}^{n} w_\beta^0.$$

Then the timescale to minimize is

$$\tau=\frac{1}{z}\sum_{\alpha=1}^{n-1}\frac{z(1,\alpha)z(\alpha+1,n)}{\min(w_\alpha^0, w_{\alpha+1}^0)} \tag{52}$$

(Larsen 1984a).

The discrete 'charge' is defined like this

$$\varepsilon_\alpha = \tanh \chi_\alpha$$

$$e^{2\chi_\alpha}=\frac{w_\alpha^0 w_{\alpha+1}^0}{zz(1,\alpha)z(\alpha+1,n)}\left\{\sum_{\beta=1}^{\alpha}\frac{z(1,\beta)^2}{w_\beta^0}+\sum_{\beta=\alpha+1}^{n}\frac{z(\beta,n)^2}{w_\beta^0}\right\}. \tag{53}$$

Consider any pair of lattice points, say $(0,0)$ and $(n_x/10, n_y/10)$, where n_x and n_y are arbitrary positive integers. A shortest path—in terms of the number of square-lattice steps it contains—connecting these points has n_x+n_y steps. But there are a number equal to $(n_x+n_y)!/n_x!\,n_y!$ distinct versions, filling the rectangle defined by the two points. In terms of *length* in such a geometry these paths are degenerate. The lattice geometry is not Euclidean.

The degeneracy is not present in terms of τ, unless all $U(x,y)=0$. Nevertheless, there is a certain 'chain-like' pliability of the square-lattice paths. This permits a path—in varying degrees—to slide down the $U(x,y)$ surface. Thereby such a path can gain an advantage that would be offset by the disadvantage of a longer length in the continuum. The 'taxicab-length' is the same (cf Krause 1975).

For instance, when crossing the barrier in (50) the path tends to slide down towards the saddle-points a bit more than it would in the continuum (see figures 2 and 3). Generally, the lattice paths tend to be more straight, and to turn more sharply where the situation allows it. One has to be aware of this when interpreting the following results, but the distortions relative to the continuum ideal are not too disturbing.

5.2. General observations on optimal diffusion paths

A reasonably elaborate computer-assisted search for minimal τ resulted in the paths shown in figure 3. In all likelihood one has to deal with a complicated optimization problem—where an automatic search using a greedy algorithm gets stuck in false minima of τ.

† Recall that *detailed balance* implies relations $L_{ab}W_b^0 = L_{ba}W_a^0$ for all pairs a, b and some set $\{W_\alpha^0\}$ of non-negative numbers, which *become* equilibrium occupations *provided* one waits long enough (which may be a very long time indeed). The rates of Metropolis *et al* (1953) can be shown to be the most efficient ones aiming towards this goal (cf Larsen 1987). A thermal situation corresponds to such a rate matrix $\mathbb{L}$ that detailed balance is satisfied with canonical Boltzmann–Gibbs factors: $W_\alpha^0 \propto \exp(-U_\alpha/kT)$.

The results are shown in table 1, corresponding to a low and a high temperature (i.e. $kT = \frac{20}{3}$ is higher than the barrier at seen from a). Here are some essential features shared by lattice and continuum models:

(i) *Minimal paths tend to cross barriers close to saddle-points, and with negative charge* ε.

(ii) *At a lower temperature the path comes closer to the exact saddle-point.*

In thermal diffusion the energy kT is readily available. If there is an advantage in it, the tracer can afford to pass the barrier about kT higher in energy than the highest saddle-point. This accounts for (ii).

Another way to understand the effect (ii) is in terms of the local field: to obtain a suitably small path radius-of-curvature ρ (cf (41)) the tracer must pass where $|F^{\perp}|$ is sufficiently large compared to kT ($\varepsilon \approx -1$ here). That is, further away from the exact saddle-point for the larger kT. The tracer can only pass more or less directly through a saddle-point if there is to be essentially no curvature of the path there. As in a–c at $(2, -2)$, or in a–b at $(-2, 2)$ and $kT = \frac{2}{3}$.

(iii) *There is a tendency to mount a slope by a steepest-descent path, and then to traverse the slope.*

This is especially evident in the a-to-$(2, -2)$ section, where a–b ascends to the ≈ -9 energy level directly, and then traverses despite the length penalty involved.

More intriguing is the bend close to c in the a–c. Since $\varepsilon \approx 1$ near c, this must be the only way to achieve the correct overall direction towards a†.

(iv) *The steep slopes of the mid-barrier minimum are traversed.*

This observation seems important. The tracer cannot venture down into the energy minimum, because the exit barrier—i.e. $(0, 0)$-to-$(-2, 2)$—is much higher than any other barrier. It would cause an enormous delay, compared to what the optimal τ of the traversing path is.

If entities start at a, the first ones to arrive at b come *directly from* a—not from the population in the deep minimum (whose members also come from a). Despite the fact that, in an eventual equilibrium situation, the minimum at $(0, 0)$ may become the most heavily populated area.

A traversing track which circumvents the minimum is nearly as quickly at b as the time it takes for another track from a to end up at the bottom of the $(0, 0)$-minimum. Therefore the first entity to turn up at b will do so surprisingly rapidly. Compared to the time it would take for an entity of the emerging minimum-population around $(0, 0)$ to get out and to reach b. The first entity to reach b will have been travelling a highly exceptional path—one which traverses a very sparsely populated area.

There is a path alternative to a–b traversing the far side of the pit. Its τ is only marginally larger than the present τ. In a symmetric configuration, say to go between $(3, -3)$ and $(-4, 4)$, the path would bifurcate inside the energy minimum region, causing optimal network formation.

Table 1. Optimal path parameters.

path	kT	τ
a–b	$\frac{2}{3}$	16 988.906
a–c	$\frac{2}{3}$	1212.445
a–b	$\frac{20}{3}$	5063.830

† The differences in values of τ between paths which differ in these latter respect are significant, but not major.

(v) *When traversing, the tracer chooses a level about kT below the saddle-point energy.* The high-temperature path a–b traverses at a deeper level inside the pit than does the low-temperature one.

(vi) *When traversing, the charge ε is nearly zero.* The transverse force field $|F^{\perp}|$ is large. So to get a modest curvature one needs a small ε.

These features of ε are illustrated in the record of $\{\varepsilon_\alpha\}$ shown in figure 4. This figure shows $\{\varepsilon_\alpha\}$ for the path a–b of $kT=\frac{2}{3}$. At this lower temperature there is better relation between the lattice path shape and the quantities $\{\varepsilon_\alpha\}$, as compared to the exact relation between a continuous path and its charge ε. Despite the peculiarities of the lattice paths, $\{\varepsilon_\alpha\}$ varies with good general resemblance to what ε would be. Smoothing the unavoidable fluctuations demonstrates how the charge is very small during both traverses (figure 4).

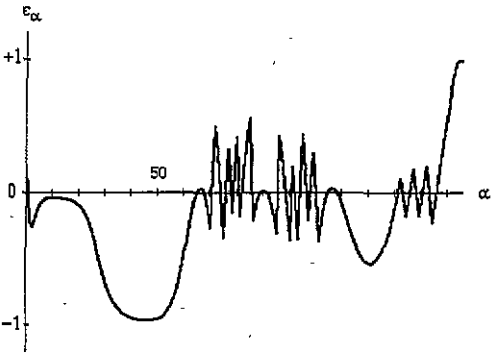

Figure 4. Charge variation along the path a–b at the temperature $\frac{2}{3}$. The charge turns negative at the saddle-points, and is essentially zero during the traverses (fluctuations being a lattice effect).

6. Discussion

As is well known, the *conventional* Euler-Lagrange variational calculus—when applied to problems in dynamics—leads to equations of motion which represent Newton's law. According to Hamilton's principle of stationary action, a path which is extremal for the action integral $S=\int \mathrm{d}t\, L(t, \boldsymbol{x}, \boldsymbol{v})$, where L is the Lagrangian, is traced by a motion which reacts to the *local* force field as prescribed in $\mathrm{d}\boldsymbol{p}/\mathrm{d}t = \varepsilon \boldsymbol{F}$. The charge, ε, remains constant.

For this reason—the locality—one tends towards not associating any basic physical significance with the global quantity S. It is sufficient to account for the local influence, the efficient cause, in terms of the field $\boldsymbol{F}$. What is more, at least in $d=1$, any ordinary second-order differential equation, $a=\phi(t, x, v)$, is the Euler equation for some Lagrangian $L(t, x, v)$ (Darboux 1894, Bolza 1909).

When the Lagrangian L depends on t, x, and $\boldsymbol{v}$, the equations of motion are differential equations. The present generalization introduces the non-local quantities p, q, and r, and the equations of motion become integro-differential.

At least in the special cases where the dependence on $\boldsymbol{v}$ is trivial, the equations of motion turn out to agree in form with Newton's $\mathrm{d}\boldsymbol{p}/\mathrm{d}t = \varepsilon \boldsymbol{F}$, but now the charge ε is a varying quantity. Its instantaneous value incorporates a global aspect: the entire optimal path must be known in order to compute the value of ε.

As demonstrated, ε regulates the tracer's reaction to the physically real local field, F. Both positive and negative values of ε are required in order to trace a path across a barrier. At certain stages the tracer has to climb against a field that would repel a particle with the physical charge $\varepsilon = 1$ of the real diffusing entities.

This situation—unusual for particle dynamics—seems analogous to what may be called 'intelligent behaviour'. The knowledge expressed in the variation of ε is precisely the knowledge acquired by *calculating* what the optimal path is. With a map and a computer one is capable of planning the sequence of ε-values required to perform the motion that traces the optimal diffusion path. Besides the efficient cause F, there is thus a *final cause* inherent in ε.

References

Aharonov Y, Bergmann P G, Lebowitz J L 1964 Time symmetry in the quantum process of measurement *Phys. Rev.* **134B** 1410

Arrhenius S 1989 Über die reaktionsgeschwindigkeit bei der inversion von rohrzucker durch säuren *Z. Phys. Chem. (Leipzig)* **4** 226

Arrhenius S 1899 Zur theorie der chemischen reaktionsgeschwindigkeit *Z. Phys. Chem. (Leipzig)* **28** 317

Barrow J D and Tipler F J 1986 *The Anthropic Cosmological Principle* (Clarendon Press, Oxford)

Bolza O 1909 *Variationsrechnung* (Berlin, Teubner)

Brinkman H C 1956 Brownian motion in a field of force and the diffusion theory of chemical reactions II *Physica* **22** 149

Cocke W J 1967 Statistical time symmetry and two-time boundary conditions in physics and cosmology *Phys. Rev.* **160** 1165

Courant R and Hilbert D 1953 *Methods of Mathematical Physics* vols I–III (Wiley-Interscience, New York)

Darboux G 1894 *Théorie Générale des Surfaces* Part 3 (Paris, Gauthier-Villars)

de Fermat P 1657 Letters LXXXVI (1657) and CXII (1662) to de la Chambre in *Oevres* vol 2: *Correspondances* eds P Tannery and C Henry (Gauthier-Villars, Paris, 1894)

Euler L 1744 *Methodus Inveniendi Lineas Curvas Maximi Minimive Proprietate Gaudentes* = Opera (Ser I) vol 24 ed C Carathéodory (Bern, 1952)

Kauffmann S A 1989a Adaptation on Rugged Fitness Landscapes p 527 in Stein (1989)

—— 1989b Principles of Adaptation in Complex Systems p 619 in Stein (1989)

Kramers H A 1940 Brownian motion in a field of force and the diffusion model of chemical reactions *Physica* **7** 284

Krause E F 1975 *Taxicab Geometry* (Dover, New York)

Larsen U 1983a The timescale of Monte Carlo experiments *Phys. Lett.* **97A** 147

—— 1983b The timescale of classical diffusion with arbitrary energy barriers *Phys. Lett.* **99A** 295

—— 1984a Bounds on the logest timescale in master equations *J. Phys. A: Math. Gen.* **17** 267

—— 1984b New exact approach to models of diffusion in one-dimensional random systems and optimal network kinetics in spin glass models *Physica Lett.* **105A** 307

—— 1987 Optimal network kinetics: A new significance to the principle of least time *Phys. Lett.* **121A** 371

Marcus M and Minc H 1964 *Matrix Theory and Matrix Inequalities* (Allyn and Bacon, Boston)

Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 Equation of state calculations by fast computing machines *J. Chem. Phys.* **21** 1087

Møller C 1952 *The Theory of Relativity* (Clarendon Press, Oxford)

Perelson A S, Kauffmann S A 1991 *Molecular Evolution on Rugged Landscapes: Proteins, RNA and the Immune System* Proceedings Vol IX Santa Fe Institute, Studies in the sciences of complexity (Addison-Wesley, Redwood City, California)

Schrödinger E 1931 Über die umkehrung der naturgesetze *Sitz. Ber. Preuss. Akad. Wiss., Phys.-Math. Kl.*, p 144

Seneta E 1973 *Non-negative Matrices* (Allen and Unwin, London)

Stein D L (ed) 1989 *Lectures in the Sciences of Complexity* Lectures vol I Santa Fe Institute, Studies in the sciences of complexity (Addison-Wesley, Redwood City, California)

Stoneham A M 1989 Defect migration in solids: Microscopic calculation of diffusion rates *Physica Scripta* **T25** 17

Wheeler J A 1979 Frontiers of Time in Problems in the Foundations of Physics eds N Toraldo di Francia and B van Fraassen (North-Holland, Amsterdam)

Wolynes P G 1989 Chemical Reaction Dynamics in Complex Molecular Systems p 355 in Stein (1989)