OPTIMALITY IN MULTI-STAGE OPERATIONS WITH ASYMPTOTICALLY VANISHING COST

JAMES D. NULTON AND PETER SALAMON

Department of Mathematical and Computer Sciences
San Diego, CA 92182-7720

Abstract. This paper sets out a framework for discussing operations whose cost can be made to approach zero by subdividing the operation into an increasing number (K) of stages. Examples of such processes include what thermodynamics books call quasistatic processes with wasted work as the cost. More generically, they include processes taking place with near-equilibrium conditions between the participants. For any fixed value of K, there are always many ways to carry out the subdivision. This paper addresses some questions related to the asymptotic (large K) behavior of the minimum total cost. In particular, we show that corresponding points in optimal subdivisions for objectives with the same quadratic part differ by $O(1/K)$, while the values of the objective functions differ by $O(1/K^2)$. Our main result is the construction of a geometrically motivated near-optimal partition scheme whose total cost (for each K) differs from the true minimum by $O(1/K^3)$. The research is motivated by recent efforts in the analysis of entropy production minimization for thermodynamic processes. In that context, our result shows that the equal thermodynamic distance subdivision will come within $O(1/K^3)$ of the true minimum entropy production.

1. Introduction

Generally speaking the cost of an operation decreases with its size or scale. In the present article, our concern is with processes for which, in the limit of small size, the cost increases quadratically with size. This gives an advantage to running a process in many small steps rather than one total step. To see how this advantage arises in a simple example, suppose that when the size of a certain operation is cut in half, the cost is scaled down by a factor of four, i.e., the cost is proportional to the square of the size or scale. Now suppose the operation (originally carried out in one stage) is subdivided into K stages, each $1/K$ times the scale of the total operation. If C denotes the cost of the operation before subdivision into stages, the cost of each stage will now be C/K^2, and the total cost of the operation after subdivision will be $K \cdot C/K^2 = C/K$. Clearly, for this particular cost/scale relation, the more stages, the better.

As a general class of physical examples we have in mind operations that take a thermodynamic system from one equilibrium state to another at the cost of producing entropy. The subdivision into stages corresponds to the appropriate selection of intermediate equilibrium states through which the system is led. Such a scenario often gives rise to an entropy production with an asymptotic dependence on K very much like the simple example cited above.

A prototypical example is the following. Suppose a fixed quantity of some material (heat capacity C_p) with temperature T equilibrates (isobarically) to a bath with temperature $T+\epsilon$. We measure the *size* of the operation by ϵ, the temperature change of the material, and the *cost* of the operation by the total entropy change of the material and bath, i.e. the entropy produced, ΔS^u. A direct calculation shows

$$\Delta S^u = \frac{C_p}{2T^2}\epsilon^2 + \cdots , \tag{1}$$

where the ellipsis denotes higher terms in ϵ. We see that this provides a simple example of the quadratic cost/size relation discussed above when the size, ϵ, is sufficiently small to make the higher order contributions negligible.

Another important physical example is a distillation column, where the overall operation is the separation into components of a binary, two-phase mixture. Two neighboring trays in the column represent a single stage in the process, each tray representing an incrementally improved separation. The interface between two such trays is a site of entropy production or exergy loss associated with an exchange of material across a gap in temperature and chemical potential. If the temperature profile along the column can be thermostatically controlled, this particular source of internal dissipation can be driven closer and closer to zero by increasing the number of stages.

The motivation behind our presentation comes primarily from the two thermodynamic examples above [1, 2]. The present treatment puts on a rigorous mathematical basis the approach of neglecting higher order terms in the type of optimizations pursued in that field for large values of K. One important reason for the present paper is to put forth a rigorous argument for the asymptotic optimality of the equal thermodynamic distance solution to the optimal partitions problem. But the motivation also extends beyond physical examples. The asymptotically quadratic nature of the cost arises many other places. For example, it shows up in the near-equilibrium behavior of forgone benefit in theories of systems whose equilibrium is characterized by optimal match to the environment. This includes economic theories where the cost is foregone welfare gains [3], coding theory where the cost is the expected number of extra bits transmitted [4], and 2-person game theory where the cost is in lost payoff. In these areas also, the issue of stepwise control to adjust to a changing environment leads to a problem of the sort discussed below.

In Section 2 we define the notions of a *locally 2nd-order cost function*, φ, and its *quadratic part*, φ_*, and establish some of their basic properties. Here the reader can think of φ as the available (useful) energy lost (or entropy produced) when a thermodynamic system equilibrates to a bath at a new temperature. As explained above, this lost availability, for a small temperature step, varies as the square of the temperature change, and, in this context, φ_* represents the (temperature-dependent) proportionality factor. We also introduce the notion of an *optimal partition scheme*. If we imagine the system to be brought from one temperature to another in a series of K equilibrations, the optimal partition scheme specifies for each K the staging of the intermediate temperatures that minimizes the total availability lost.

Section 3 proves some technical results on the "uniform" character of optimal partition schemes in general. Included is the fact (Theorem 2) that consecutive intervals within an optimal partition differ in length by $O(K^{-2})$. This result, while not surprising, has the consequence (Proposition 6, to be used in Section 4) that the minimum values of cost functions with the same quadratic part also differ by no more than $O(K^{-2})$.

The proofs in Sections 2 and 3 (with the possible exception of Theorem 2) are quite straightforward and can be skipped by the savvy reader. They are included for completeness.

In Section 4 we introduce a near-optimal scheme based on a special metric defined on the parameter space. This *equal distance partition scheme* has the virtue that it is directly computable in terms of the quadratic part of the cost function, and yet its total cost differs from the true minimum by no more than $O(K^{-3})$. This result (Theorem 6) is quite surprising and underlines the significance of the asymptotic questions posed here and the importance of their careful analysis.

We now lay out the mathematical framework for our discussion.

2. Staged Operations with Continuous Progress Parameter

Let us suppose that the progress of a certain operation is measured by a suitable continuous parameter, t, that varies over an interval $I = [a, b]$ from start to completion[1]. Further suppose that for any two intermediate levels of completion r and s with $a \le r < s \le b$, we are able to formulate the cost, $\varphi(r, s)$, of taking the operation from r to s. If we regard this transition as a potential stage in the subdivision of the full operation, then the quantity $s - r$ is a measure of the scale or size of that stage. As we noted in our simple example, the issue of whether there is any benefit to increasing the number of stages hinges on how the cost $\varphi(r, s)$ of a stage varies with its scale $s - r$. We choose a smooth class of functions for which the total cost of the operation can be made to vanish in the limit $K \to \infty$. For this class, a comprehensive theory can be developed for studying how the minimum cost depends asymptotically on K, the number of stages into which the operation is divided.

The local cost functions φ that we will study are positive valued and have the property that, for s close to r, the cost is essentially quadratic in the difference $(s - r)$. More precisely, a C^∞ function $\varphi(r, s)$ defined on $I \times I$ is called a *locally 2nd-order cost function* iff

$$\varphi(r, s) > 0, \quad \text{whenever } s \neq r, \tag{2}$$

and the following limit exists *and is positive* for all $t \in I$, defining a C^∞ function φ_* called the *quadratic part* of φ:

$$0 < \varphi_*(t) = \lim_{(r,s)\to(t,t)} \frac{\varphi(r, s)}{(s - r)^2}, \tag{3}$$

where, of course, $r = s$ is avoided in the limit.

Suppose the operation is divided into K stages, where t_j represents the level of completion after the jth stage. Such a division amounts to a K-fold partition, P_K, of the interval I:

$$a = t_0 \le t_1 \le \cdots \le t_K = b, \tag{4}$$

wherein the system progresses during the jth step from t_{j-1} to t_j.

[1]In the thermodynamics problems of interest, I represents the interval over which some convenient parameter varies along a given path in (many dimensional) state space.

We take the *total cost* for the process to be cumulative over the K steps:

$$\Phi[P_K] = \sum_{j=1}^{K} \varphi(t_{j-1}, t_j). \tag{5}$$

We are particularly interested in partitions that minimize the total cost. That such partitions exist for each K can be seen by the following argument. The set of $(K+1)$-tuples $(t_0, \cdots, t_K)$ corresponding to partitions (4) form a compact subset of R^{K+1}. The total cost (5) as a function defined on that set is continuous, and therefore assumes a minimum value. A sequence $\{P_K\}$ of such minimizing partitions, one for each K, will be called an *optimal partition scheme* for the cost function φ.

Such schemes will be studied in Section 3, but first we clarify the relation between a locally 2nd-order cost function φ and its quadratic part φ_*. We will use subscripts to denote partial derivatives of $\varphi(r, s)$, e.g., $\varphi_{11} = \partial^2\varphi/\partial r^2$, $\varphi_{12} = \partial^2\varphi/\partial r\partial s$.

Proposition 1. *Let $\varphi(r, s)$ be a locally 2nd-order cost function with quadratic part $\varphi_*(t)$. Then the following two equivalent statements hold.*

The 2nd-order Taylor polynomial for $\varphi(r, s)$ about the point (t, t) is simply

$$\varphi_*(t)(s - r)^2. \tag{6}$$

The following relations hold among the first few derivatives of $\varphi(r, s)$ on the "diagonal":

$$\varphi(t, t) = \varphi_1(t, t) = \varphi_2(t, t) = 0, \tag{7}$$

$$\varphi_*(t) = \frac{1}{2}\varphi_{11}(t, t) = \frac{1}{2}\varphi_{22}(t, t) = -\frac{1}{2}\varphi_{12}(t, t). \tag{8}$$

We prove the second statement of the propostion first. To establish $\varphi(t, t) = 0$, we observe that the limit in the numerator of (3) is $\varphi(t, t)$. If this were not zero, the limit of the quotient would fail to exist, since the limit in the denominator is clearly zero.

To verify $\varphi_1(t, t) = 0$ and $\varphi_*(t) = \frac{1}{2}\varphi_{11}(t, t)$, we make use of L'hopital's rule in definition (3) along the path $(r, s) = (t + h, t)$:

$$\varphi_*(t) = \lim_{h\to 0} \frac{\varphi(t + h, t)}{h^2} = \lim_{h\to 0} \frac{\varphi_1(t + h, t)}{2h} \tag{9}$$

As in the above argument, if, in the second limit, the limit of the numerator were not zero, the limit of the quotient would fail to exist. Thus we must have $\varphi_1(t, t) = 0$. With this we may apply L'hopital's rule once again to obtain

$$\lim_{h\to 0} \frac{\varphi_1(t + h, t)}{2h} = \frac{1}{2}\varphi_{11}(t, t). \tag{10}$$

Thus equations (9) and (10) together establish $\varphi_1(t, t) = 0$ and $\varphi_*(t) = \frac{1}{2}\varphi_{11}(t, t)$. We establish $\varphi_2(t, t) = 0$ and $\varphi_*(t) = \frac{1}{2}\varphi_{22}(t, t)$ in exactly the same way by applying L'hopital's rule in definition (3) along the path $(r, s) = (t, t + h)$. We have now established (7) and the first two equations in (8).

The third equation in (8) can be obtained in a similar manner by considering the path $(r, s) = (t + h, t - h)$. Using (7) we can apply L'hopital's rule twice in succession to obtain

$$\varphi_*(t) = \lim_{h\to 0} \frac{\varphi(t + h, t - h)}{4h^2} = \frac{1}{8}\left(\varphi_{11}(t, t) - 2\varphi_{12}(t, t) + \varphi_{22}(t, t)\right). \tag{11}$$

The third equation of (8) now follows from its first two equations. The second statement of the propostition is now established.

The first statement now follows from the identity

$$(s-r)^2 = (s-t)^2 - 2(s-t)(r-t) + (r-t)^2, \tag{12}$$

and the proposition is proved. □

One succinct consequence of Proposition 1 is that we may write

$$\varphi(r, r+\epsilon) = \varphi_*(r)\epsilon^2 + \cdots, \tag{13}$$

where the remainder involves ϵ to order at least 3. The propostion below shows that (13) is also sufficient to imply that the function φ is locally 2nd-order.

Proposition 2. *A function φ defined on $I \times I$ is locally 2nd-order iff equation (13) holds for all $(r, r+\epsilon)$ in $I \times I$.*

The implication was argued above. For the converse, define $s = r + \epsilon$ and note that equation (13) rearranges directly to

$$\frac{\varphi(r,s)}{(s-r)^2} = \varphi_*(r) + \cdots, \tag{14}$$

where the remainder involves $(s-r)$ to order at least 1 and hence approaches zero as $(r,s) \to (t,t)$. □

Equation (13) is the form in our physical example above regarding the isobaric heating of some material. Proposition 2 shows that Definition (3) is merely a more symmetric form of this same assumption. Equation (13) is, however, physically the more palatable form[2].

Proposition 3. *Let φ and φ_* be as in Proposition 1. There are constants $A, B > 0$ so that, for any $r, s, t \in I$,*

$$A(s-r)^2 \le \varphi(r,s) \le B(s-r)^2, \tag{15}$$

$$A \le \varphi_*(t) \le B. \tag{16}$$

Eq. (3) allows the function $\hat{\varphi}(r,s) = \varphi(r,s)/(s-r)^2$ to be extended to a positive-valued continuous function on all of $I \times I$. It therefore has a maximum B and a minimum A distinct from zero, and the proposition follows. □

Much of the development below concerns assertions regarding the degree of sub-optimality obtained in using an optimal partition for a modified objective function that has the same quadratic part as our original cost. Before proceeding, we will need one further technical result on the difference in stage cost for two functions with the same quadratic part.

Proposition 4. *Let $\varphi^{(1)}$ and $\varphi^{(2)}$ be locally 2nd-order cost functions with the same quadratic part, $\varphi_*^{(1)} = \varphi_*^{(2)}$. Then there exists a constant Q such that*

$$|\varphi^{(1)}(r,s) - \varphi^{(2)}(r,s)| \le Q|s-r|^3, \tag{17}$$

for all $r, s \in I$.

[2]To see that entropy production indeed is locally 2nd-order, we need to interpret $\varphi(r,s)$ as the entropy production of a system initially at state r reaching equilibrium at s with an environment (reservoir) whose constant intensities guarantee that at equilibrium the system will be in state s.

In order to prove this proposition and later results, we will find it useful to define the *norm* of a continuous function $f(r,s)$ for $r,s \in I$:

$$||f|| = \max\{f(r,s) \,:\, r,s \in I\}. \tag{18}$$

Apply Taylor's theorem to third order to both $\varphi^{(1)}(r,s)$ and $\varphi^{(2)}(r,s)$ about the point (r,r). We obtain, for each s, points s_1 and s_2 lying between r and s for which we can write

$$\varphi^{(m)}(r,s) = \varphi_*(r)(s-r)^2 + \frac{1}{6}\varphi^{(m)}_{222}(r,s_m)(s-r)^3, \tag{19}$$

for $m = 1, 2$. Subtracting the two functions gives

$$\varphi^{(1)}(r,s) - \varphi^{(2)}(r,s) = \frac{1}{6}\{\varphi^{(1)}_{222}(r,s_1) - \varphi^{(2)}_{222}(r,s_2)\}(s-r)^3, \tag{20}$$

from which it follows that

$$|\varphi^{(1)}(r,s) - \varphi^{(2)}(r,s)| \le \frac{1}{6}\{||\varphi^{(1)}_{222}|| + ||\varphi^{(2)}_{222}||\}|s-r|^3. \tag{21}$$

This proves Proposition 4. □

3. Optimal Partition Schemes

In this section we prove some properties of optimal partition schemes. The main result is that, for such a scheme, the length of the maximum interval varies like $1/K$. For this and subsequent discussion, it will be useful to make some definitions and introduce some notational conventions.

First, when discussing a partition, we generally use "Δ" to represent the length of a subinterval. For example, in the partition $P = \{t_j\}$, we will write $\Delta_j = |t_j - t_{j-1}|$. Similarly, the lengths in a second partition $\bar{P}_K = \{\bar{t}_j\}$ will be denoted by $\bar{\Delta}_j$. Secondly, we make use of the "O" notation. By $F(K) = O(K^{-m})$ we will mean that $K^m \cdot F(K)$ is bounded, i.e., there exists a constant M for which

$$|F(K)| \le \frac{M}{K^m}. \tag{22}$$

This bound will always be independent of K; what it does depend on should be clear from the context.

The following preliminary result will be used in the theorems, but (23) below will be improved in Proposition 8.

Proposition 5. *Let $\{P_K\}$ be an optimal partition scheme for a locally 2nd-order cost function φ for the interval I. Then*

$$\Phi[P_K] = O(K^{-1}), \tag{23}$$

$$\sum_{j=1}^{K} (\Delta_j^{(K)})^2 = O(K^{-1}). \tag{24}$$

Consider the strictly uniform partition scheme $\bar{P}_K = \{\bar{t}_j^{(K)}\}$, defined by $\bar{t}_j^{(K)} = a + (b-a)(j/K)$ and $\bar{\Delta}_j^{(K)} = (b-a)/K$. Employing the constant B from (15), we have

$$\Phi[\bar{P}_K] = \sum_{j=1}^{K} \varphi(\bar{t}_{j-1}, \bar{t}_j) \le \sum_{j=1}^{K} B\,(\bar{\Delta}_j^{(K)})^2 = \frac{B(b-a)^2}{K}. \tag{25}$$

Since P_K is optimal, we must have $\Phi[P_K] \le \Phi[\bar{P}_K]$. This proves (23). To prove (24) we employ the constant A in (15) to write

$$A \sum_{j=1}^{K} (\Delta_j^{(K)})^2 \le \Phi[P_K]. \tag{26}$$

Now (24) follows from (23), and the proposition is proved. □

The next two theorems are probably the most that we can say in general about the "uniformity" of an optimal partition scheme. They will be used in subsequent sections.

Theorem 1. *Let $\{P_K\}$ be an optimal partition scheme for a locally 2nd-order cost function φ. Then*

$$\max_{j=1}^{K} \Delta_j^{(K)} = O(K^{-1}). \tag{27}$$

Theorem 2. *Let $\{P_K\}$ be an optimal partition scheme for a locally 2nd-order cost function φ. Then*

$$\max_{j=1}^{K-1} |\Delta_{j+1}^{(K)} - \Delta_j^{(K)}| = O(K^{-2}). \tag{28}$$

For the proof of Theorem 1, choose a particular K and let $P_K = \{t_j\}$. For simplicity of notation we temporarily supress the mark of K-dependence on t_j and Δ_j. For any value of j, the total cost in (5) can be written as

$$\Phi[P_K] = \varphi(t_{j-1}, t_j) + \varphi(t_j, t_{j+1}) + \cdots , \tag{29}$$

where all the terms in the ellipsis are independent of t_j. A necessary condition that P_K be optimal is that the derivative of (29) with respect to t_j vanish, i.e.,

$$\varphi_2(t_{j-1}, t_j) + \varphi_1(t_j, t_{j+1}) = 0, \tag{30}$$

for all $1 \le j \le K-1$.

Apply Taylor's theorem to second order to both terms in (30) at the point (t_j, t_j). Then there exist $t_j^* \in [t_{j-1}, t_j]$ and $t_j^{**} \in [t_j, t_{j+1}]$ for which

$$\varphi_2(t_{j-1}, t_j) = \varphi_2(t_j, t_j) - \varphi_{21}(t_j, t_j)\Delta_j + \frac{1}{2}\varphi_{211}(t_j^*, t_j)\Delta_j^2, \tag{31}$$

$$\varphi_1(t_j, t_{j+1}) = \varphi_1(t_j, t_j) + \varphi_{12}(t_j, t_j)\Delta_{j+1} + \frac{1}{2}\varphi_{122}(t_j, t_j^{**})\Delta_{j+1}^2. \tag{32}$$

Using Proposition 1 these can be simplified to

$$\varphi_2(t_{j-1}, t_j) = 2\varphi_*(t_j)\Delta_j + \frac{1}{2}\varphi_{211}(t_j^*, t_j)\Delta_j^2, \tag{33}$$

$$\varphi_1(t_j, t_{j+1}) = -2\varphi_*(t_j)\Delta_{j+1} + \frac{1}{2}\varphi_{122}(t_j, t_j^{**})\Delta_{j+1}^2. \tag{34}$$

Adding these two equations and invoking the optimality condition (30), we have

$$\Delta_{j+1} - \Delta_j = \frac{1}{4\,\varphi_*(t_j)} \left(\varphi_{211}(t_j^*, t_j)\Delta_j^2 + \varphi_{122}(t_j, t_j^{**})\Delta_{j+1}^2\right). \tag{35}$$

Let A be the constant in (16) and let $M = \max\{||\varphi_{211}||, ||\varphi_{122}||\}$. Then we have

$$|\Delta_{j+1} - \Delta_j| \le \frac{M}{4A}(\Delta_j^2 + \Delta_{j+1}^2). \tag{36}$$

Now for any pair l, m with $1 \le l < m \le K$,we have

$$|\Delta_m - \Delta_l| \le \sum_{j=l}^{m-1} |\Delta_{j+1} - \Delta_j|. \tag{37}$$

Combining (36) and (37) gives

$$|\Delta_m - \Delta_l| \le \frac{M}{4A} \sum_{j=l}^{m-1} (\Delta_{j+1}^2 + \Delta_j^2) \le \frac{M}{2A} \sum_{j=1}^{K} \Delta_j^2. \tag{38}$$

In particular,

$$\Delta_{max} \le \Delta_{min} + \frac{M}{2A} \sum_{j=1}^{K} \Delta_j^2. \tag{39}$$

We certainly have $\Delta_{min} \le (b-a)/K$. Theorem 1 now follows from (24) of Proposition 5. Now if we apply Theorem 1 to (36), we obtain Theorem 2. □

Proposition 6. *Let $\varphi^{(1)}$ and $\varphi^{(2)}$ be locally 2nd-order cost functions with the same quadratic part. Let $\{P_K\}$ be an optimal partition scheme for a third cost function φ (not necessarily distinct from $\varphi^{(1)}$ and $\varphi^{(2)}$). Then*

$$\Phi^{(1)}[P_K] - \Phi^{(2)}[P_K] = O(K^{-2}). \tag{40}$$

This proposition is a corollary of Theorem 1 and Proposition 4. Let $P_K = \{t_j\}$. First write

$$|\,\Phi^{(1)}[P_K] - \Phi^{(2)}[P_K]\,| \le \sum_{j=1}^{K} |\,\varphi^{(1)}(t_{j-1}, t_j) - \varphi^{(2)}(t_{j-1}, t_j)|. \tag{41}$$

Now invoke Proposition 4 to obtain

$$|\,\Phi^{(1)}[P_K] - \Phi^{(2)}[P_K]\,| \le Q \sum_{j=1}^{K} \Delta_j^3 \tag{42}$$

The proposition now follows by applying Theorem 1. Note that Theorem 1 required the optimality of P_K for some locally 2nd-order cost function. □

Proposition 7. *Let $\varphi^{(1)}$ and $\varphi^{(2)}$ be locally 2nd-order cost functions with the same quadratic part. Let $\{P_K^{(1)}\}$ and $\{P_K^{(2)}\}$ be optimal partition schemes for $\varphi^{(1)}$ and $\varphi^{(2)}$, respectively. Then all four of the numbers*

$$\Phi^{(1)}[P_K^{(1)}], \Phi^{(1)}[P_K^{(2)}], \Phi^{(2)}[P_K^{(2)}], \Phi^{(2)}[P_K^{(1)}], \tag{43}$$

are within $O(K^{-2})$ of each other.

By Proposition 6 the statement is true for the two outside terms of (43) as well as the two inside terms. The rest follow from optimality

$$\Phi^{(1)}[P_K^{(1)}] \le \Phi^{(1)}[P_K^{(2)}], \tag{44}$$

$$\Phi^{(2)}[P_K^{(2)}] \le \Phi^{(2)}[P_K^{(1)}], \tag{45}$$

and the triangle inequality. □

This proposition will be very useful in the next section, where, for a given cost function φ, we will construct a related cost function $\hat{\varphi}$ with the same quadratic part and an optimal partition scheme $\hat{P}_K$ for which $\hat{\Phi}[\hat{P}_K]$ can be calculated explicitly.

4. Distance in Parameter Space

The progress parameter $t \in I = [a, b]$ is simply a device for coordinatizing the various stages of completion of the operation under study. In this section we explore the effect of a smooth "change of coordinates". Let F be a diffeomorphism from the interval I to some other interval J that preserves the relation "<". This is simply an increasing one-to-one map of I onto J in which both F and F^{-1} are differentiable. The map F induces a related map of cost functions. That is, a locally 2nd-order cost function φ on I is associated with a locally 2nd-order cost function ψ on J by the relation

$$\psi(u, u') = \varphi(F^{-1}(u), F^{-1}(u')) \tag{46}$$

In fact, it is an easy calculation to show that the quadratic parts are related by

$$\psi_*(u) = \frac{\varphi_*(t)}{F'(t)^2}, \tag{47}$$

where $t = F^{-1}(u)$. This new cost function ψ can be regarded as the same cost expressed in terms of a new progress parameter (coordinate) u related to the old coordinate t by $u = F(t)$.

A partition $P = \{t_j\}$ of I is carried over to a partition $Q = F(P) = \{F(t_j)\}$ of J. The total costs are identical on corresponding partitions:

$$\Psi[Q] = \Phi[P]. \tag{48}$$

It follows that Q is optimal for ψ if and only if P is optimal for φ.

¿From (47) we see that if F satisfies the differential equation

$$F'(t)^2 = \varphi_*(t), \tag{49}$$

for $t \in [a, b]$, i.e.,

$$F(t) = \int_a^t \sqrt{\varphi_*(r)}\, \mathrm{d}r, \tag{50}$$

then we have

$$\psi_*(u) = 1. \tag{51}$$

This makes (50) a natural coordinate system for the cost function φ. The original parameter t may be appropriate for the *formulation* of the problem, e.g., temperature in the example in Section 1, but $u = F(t)$ will be more appropriate for the *analysis* of the problem.

¿From the viewpoint of the original parameter space $I = [a, b]$, F provides a natural measure of "distance" between two parameter values:

$$\mathrm{dist}(t, t') = F(t') - F(t) = \int_t^{t'} \sqrt{\varphi_*(r)}\, \mathrm{d}r. \tag{52}$$

For that reason, in connection with the example in Section 1,

$$\int_{T_a}^{T_b} \frac{\sqrt{C_p(T)}}{T}\, \mathrm{d}T \tag{53}$$

has been called the *thermodynamic distance* between the two temperatures[3].

Let L denote the "length" of I, i.e.,

$$L = \int_a^b \sqrt{\varphi_*(t)}\,\mathrm{d}t, \tag{54}$$

so that L is literally the length of $J = [0, L]$.

There is a special cost function on the space $[0, L]$, namely,

$$\hat{\psi}(u, u') = (u' - u)^2. \tag{55}$$

This function has three important features. First, evidently $\hat{\psi}_*(u) = \psi_*(u) = 1$, and, for that reason, we expect, in view of Propositions 5 and 6 of the previous section, that the optimality theory of $\hat{\psi}$ will be intimately related to that of ψ. Second, the optimal partition scheme for $\hat{\psi}$ is unique and can be written down explicitly. Third, $\hat{\psi}$ "pulls back" to a cost function $\hat{\varphi}$ in the space $I = [a, b]$ in such a way that the connection between $\hat{\psi}$ and ψ is reflected in $\hat{\varphi}$'s relation to φ, the original function under study. We now develop these connections in detail.

First of all, an optimal partition of $J = [0, L]$ for $\hat{\psi}$ minimizes the sum of the squares of the subintervals. It is intuitively clear (and an easy exercise to show) that this requires the subintervals to have equal length. Let $\hat{Q}_K = \{\hat{u}_j^{(K)}\}$ denote the special *uniform partition scheme* for the interval $J = [0, L]$ defined by

$$\hat{u}_j^{(K)} = j \cdot \frac{L}{K}, \tag{56}$$

for $j = 1$ to K. This is the optimal partition scheme for $\hat{\psi}$.

The inverse image $\hat{P}_K = \{\hat{t}_j^{(K)}\} = \{F^{-1}(\hat{u}_j^{(K)})\}$ will then be an optimal partition scheme for the "pull-back" function $\hat{\varphi}(t, t') = \hat{\psi}(F(t), F(t'))$. We write this out explicitly using (55) and (50):

$$\hat{\varphi}(t, t') = \left(\int_t^{t'} \sqrt{\varphi_*(r)}\,\mathrm{d}r \right)^2. \tag{57}$$

¿From this, one can readily verify that

$$\hat{\varphi}_*(t) = \varphi_*(t). \tag{58}$$

The scheme $\hat{P}_K$ can be computed by using (52). It is the partition $\hat{t}_j^{(K)}$ of $I = [a, b]$ defined by

$$\mathrm{dist}(\hat{t}_{j-1}^{(K)}, \hat{t}_j^{(K)}) = \int_{\hat{t}_{j-1}^{(K)}}^{\hat{t}_j^{(K)}} \sqrt{\varphi_*(t)}\,\mathrm{d}t = \frac{L}{K}. \tag{59}$$

We call $\hat{P}_K$ the *equal distance partition scheme*. It is the optimal partition scheme for $\hat{\varphi}$, in view of the remarks surrounding (48). In fact, as in (48), the total cost

[3]In the thermodynamic context this length actually corresponds (up to a factor of $\sqrt{2}$) to the length of the path in the manifold of states equipped with a semi-Riemannian structure. The Riemannian metric is given by the second derivative matrix of the entropy of the system with respect to the extensive variables. For a recent review, see [5].

functions for $\hat{\psi}$ and $\hat{\varphi}$ have the same minimum value for each K:

$$\hat{\Phi}[\hat{P}_K] = \hat{\Psi}[\hat{Q}_K] = \sum_{j=1}^{K} \left(\frac{L}{K}\right)^2 = \frac{L^2}{K}. \tag{60}$$

Using (54) we can write this as

$$\hat{\Phi}[\hat{P}_K] = \frac{1}{K}\left(\int_a^b \sqrt{\varphi_*(t)}\,\mathrm{d}t\right)^2. \tag{61}$$

Now let $\{P_K\}$ be an optimal partition scheme for φ. In view of (58) we can apply Proposition 7 to $\hat{\varphi}$ and φ to obtain

$$\Phi[P_K] = \hat{\Phi}[\hat{P}_K] + O(K^{-2}). \tag{62}$$

Combining this with (61) we have the following corollary of Proposition 7.

Proposition 8. *Let φ be a locally 2nd-order cost function on $I = [a, b]$ with quadratic part φ_*. Let $\{P_K\}$ be an optimal partition scheme for φ. Then*

$$\Phi[P_K] = \frac{1}{K}\left(\int_a^b \sqrt{\varphi_*(t)}\,\mathrm{d}t\right)^2 + O(K^{-2}). \tag{63}$$

This can be regarded as a strengthening of (23) in Proposition 5.

We are now ready to describe more precisely the close connection between $\hat{\psi}$, ψ, $\hat{Q}_K$ and Q_K (an optimal partition for ψ). We do that in the next two theorems. The proofs are rather technical and are deferred to Section 5. However, their consequences for $\hat{\varphi}$, φ, $\hat{P}_K$ and P_K are almost immediate and give the content of Theorems 5 and 6.

Theorem 3. *Let ψ be a locally 2nd-order cost function on J with quadratic part $\psi_* = 1$. Let $Q_K = \{u_j^{(K)}\}$ be an optimal partition scheme for ψ. Let $\hat{Q}_K = \{\hat{u}_j^{(K)}\}$ be the uniform partition scheme (optimal for $\hat{\psi}$). Then*

$$\max_{j=1}^{K} |\Delta_j^{(K)} - \hat{\Delta}_j^{(K)}| = O(K^{-2}), \tag{64}$$

$$\max_{j=1}^{K} |u_j^{(K)} - \hat{u}_j^{(K)}| = O(K^{-1}), \tag{65}$$

(Note: If $J = [0, L]$, $\hat{\Delta}_j^{(K)} = L/K$ and $\hat{u}_j^{(K)} = jL/K$.)

Eq. (64) of Theorem 3 will be used in the proof of Theorem 4.

Theorem 4. *Let ψ be a locally 2nd-order cost function on J with quadratic part $\psi_* = 1$. Let $Q_K = \{u_j^{(K)}\}$ be an optimal partition scheme for ψ. Let $\hat{Q}_K = \{\hat{u}_j^{(K)}\}$ be the uniform partition scheme (optimal for $\hat{\psi}$). Then*

$$\Psi[\hat{Q}_K] - \Psi[Q_K] = O(K^{-3}). \tag{66}$$

These two theorems show that if $\psi_*(u) = 1$, then the *uniform* partition scheme $\hat{Q}_K$ is close to any optimal partition scheme, and the ("unhatted!") total cost Ψ for $\hat{Q}_K$ is *very* close to the minimum. The next two theorems state that the *equal distance* partition scheme $\hat{P}_K$ plays the same role for φ and Φ, i.e. translate Theorems 3 and 4 back to the t domain.

Theorem 5. *Let φ be a locally 2nd-order cost function on $I = [a, b]$. Let $P_K = \{t_j^{(K)}\}$ be an optimal partition scheme for φ. Let $\hat{P}_K = \{\hat{t}_j^{(K)}\}$ be the equal distance scheme defined above. Then*

$$\max_{j=1}^{K} |\hat{t}_j^{(K)} - t_j^{(K)}| = O(K^{-1}). \tag{67}$$

Theorem 6. *Let φ be a locally 2nd-order cost function on $I = [a, b]$. Let $P_K = \{t_j^{(K)}\}$ be an optimal partition scheme for φ. Let $\hat{P}_K = \{\hat{t}_j^{(K)}\}$ be the equal distance scheme defined above. Then*

$$\Phi[\hat{P}_K] - \Phi[P_K] = O(K^{-3}). \tag{68}$$

Theorem 6 is immediate, since the mapping F preserves total costs. Theorem 5 requires a few remarks.

Apply the inequality (16) to φ_* in (52) to obtain the relation

$$\sqrt{A}\,|\,t' - t\,| \le |\,u' - u\,| \le \sqrt{B}\,|\,t' - t\,|, \tag{69}$$

for any numbers related by $u = F(t)$ and $u' = F(t')$. This shows that (65) and (67) are equivalent. □

5. The Case $\psi_*(u) = 1$

Proof of Theorem 3: For notational simplicity we suppress the mark of K-dependence on the partition points, i.e., we write $Q_K = \{u_j\}$; we also assume $J = [0, L]$.

The necessary condition for optimality (30) adapted for ψ is

$$\psi_2(u_{j-1}, u_j) + \psi_1(u_j, u_{j+1}) = 0, \tag{70}$$

for all $1 \le j \le K - 1$. Apply Taylor's theorem to third order to each term in (70) at the point (u_j, u_j). There exist $u_j^* \in [u_{j-1}, u_j]$ and $u_j^{**} \in [u_j, u_{j+1}]$ for which

$$\psi_2(u_{j-1}, u_j) = -2\Delta_j + \frac{1}{2}\psi_{211}(u_j, u_j)\Delta_j^2 - \frac{1}{6}\psi_{2111}(u_j^*, u_j)\Delta_j^3, \tag{71}$$

$$\psi_1(u_j, u_{j+1}) = 2\Delta_{j+1} + \frac{1}{2}\psi_{122}(u_j, u_j)\Delta_{j+1}^2 + \frac{1}{6}\psi_{1222}(u_j, u_j^{**})\Delta_{j+1}^3. \tag{72}$$

Here we have used Proposition 1 adapted to ψ and the fact that $\psi_* = 1$ to simplify the first and second derivative terms at the point (u_j, u_j). Also note that, by Theorem 1, the fourth derivative terms can be replaced by $O(K^{-3})$. With that, we add these two equations and invoke the optimality condition (70) to obtain

$$\Delta_{j+1} - \Delta_j = \frac{1}{4}\psi_{122}(u_j, u_j)\Delta_{j+1}^2 + \frac{1}{4}\psi_{211}(u_j, u_j)\Delta_j^2 + O(K^{-3}). \tag{73}$$

The third derivative terms have to be handled more delicately. This is where the hypothesis $\psi_*(u) = \textit{constant}$ becomes indispensible. From $2 = 2\psi_*(u) = -\psi_{12}(u, u) = -\psi_{21}(u, u)$ we obtain

$$0 = 2\psi_*'(u_j) = -\psi_{211}(u_j, u_j) - \psi_{122}(u_j, u_j). \tag{74}$$

Now (73) can be rewritten

$$\Delta_{j+1} - \Delta_j = \frac{1}{4}\psi_{122}(u_j, u_j)(\Delta_{j+1}^2 - \Delta_j^2) + O(K^{-3}). \tag{75}$$

Factor $(\Delta_{j+1}^2 - \Delta_j^2) = (\Delta_{j+1} + \Delta_j)(\Delta_{j+1} - \Delta_j)$. By Theorem 1, $(\Delta_{j+1} + \Delta_j) = O(K^{-1})$. By Theorem 2, $(\Delta_{j+1} - \Delta_j) = O(K^{-2})$. Therefore $(\Delta_{j+1}^2 - \Delta_j^2) = O(K^{-3})$. Finally (75) can be written

$$\Delta_{j+1} - \Delta_j = O(K^{-3}). \tag{76}$$

For any m and l we have

$$\Delta_m - \Delta_l = \sum_{j=l}^{m} (\Delta_{j+1} - \Delta_j). \tag{77}$$

(If $l > m$ the sum must be written a little differently.) Furthermore,

$$\Delta_m - \hat{\Delta}_m = \Delta_m - L/K = \Delta_m - \frac{1}{K}\sum_{l=1}^{K} \Delta_l = \frac{1}{K}\sum_{l=1}^{K}(\Delta_m - \Delta_l). \tag{78}$$

Now combining (76), (77) and (78), and noting that the sum in (77) has no more than K terms, we have

$$\Delta_m - \hat{\Delta}_m = O(K^{-2}). \tag{79}$$

This is formula (64) of Theorem 3.

To obtain the second part of the theorem, write

$$u_j - \hat{u}_j = \sum_{m=1}^{j} (\Delta_m - \hat{\Delta}_m). \tag{80}$$

Since there are no more than K terms in this sum, (65) follows from (79). □

The proof of Theorem 4 requires a

Lemma 1. *Let I be a finite interval and let f be a C^∞-function on $I \times I$. Then there is a constant $M > 0$ so that for any choice of numbers $x_1 \le y_1 \le z_1$ and $x_2 \le y_2 \le z_2$ all belonging to I we have*

$$|f(x_2, y_2) - f(x_1, y_1)| \le M(|x_2 - x_1| + |z_2 - z_1| + |z_1 - x_1| + |z_2 - x_2|). \tag{81}$$

Apply Taylor's theorem to $f(x_2, y_2)$ to first order about (x_1, y_1) to obtain

$$f(x_2, y_2) = f(x_1, y_1) + f_1(x^*, y^*)(x_2 - x_1) + f_2(x^*, y^*)(y_2 - y_1), \tag{82}$$

where (x^*, y^*) is on the line segment joining the point (x_1, y_1) and (x_2, y_2). Choose $M = \max\{||f_1||, ||f_2||\}$, then we have

$$|f(x_2, y_2) - f(x_1, y_1)| \le M(|x_2 - x_1| + |y_2 - y_1|). \tag{83}$$

Now if $y_1 = \mu x_1 + (1 - \mu)z_1$ and $y_2 = \lambda x_2 + (1 - \lambda)z_2$ we can write

$$y_2 - y_1 = z_2 - z_1 + \mu(z_1 - x_1) - \lambda(z_2 - x_2), \tag{84}$$

from which follows

$$|y_2 - y_1| \le |z_2 - z_1| + |z_1 - x_1| + |z_2 - x_2|. \tag{85}$$

Combining (83) and (85), the lemma is proved. □

We are now ready for the

Proof of Theorem 4: Apply Taylor's theorem to third order to $\psi(u_{j-1}, u_j)$ about (u_{j-1}, u_{j-1}) and to $\psi(\hat{u}_{j-1}, \hat{u}_j)$ about $(\hat{u}_{j-1}, \hat{u}_{j-1})$. There exist points $u_j^* \in [u_{j-1}, u_j]$ and $\hat{u}_j^* \in [\hat{u}_{j-1}, \hat{u}_j]$ for which

$$\psi(u_{j-1}, u_j) = \Delta_j^2 + f(u_{j-1}, u_j^*)\Delta_j^3, \tag{86}$$

$$\psi(\hat{u}_{j-1}, \hat{u}_j) = \hat{\Delta}_j^2 + f(\hat{u}_{j-1}, \hat{u}_j^*)\hat{\Delta}_j^3, \tag{87}$$

For notational convenience we have introduced $f(u, u') = \psi_{111}(u, u')/6$. Subtract (86) from (87) and rearrange terms as follows:

$$\begin{aligned} \psi(\hat{u}_{j-1}, \hat{u}_j) - \psi(u_{j-1}, u_j) &= 2\hat{\Delta}_j(\hat{\Delta}_j - \Delta_j) && (88)\\ &- (\hat{\Delta}_j - \Delta_j)^2 && (89)\\ &+ f(u_{j-1}, u_j^*)(\hat{\Delta}_j^3 - \Delta_j^3) && (90)\\ &+ (f(\hat{u}_{j-1}, \hat{u}_j^*) - f(u_{j-1}, u_j^*))\hat{\Delta}_j^3. && (91) \end{aligned}$$

We will now show that each of (89), (90) and (91) is $O(K^{-4})$. First, (89) $= O(K^{-4})$ follows immediately from (64) of Theorem 3. Second, factor $(\hat{\Delta}_j^3 - \Delta_j^3) = (\hat{\Delta}_j - \Delta_j)(\hat{\Delta}_j^2 + \hat{\Delta}_j\Delta_j + \Delta_j^2)$. The first factor is $O(K^{-2})$ by Theorem 3 and the second factor is $O(K^{-2})$ by Theorem 1. Thus (90) $= O(K^{-4})$. The term (91) is the trickiest. Apply Lemma 1 to obtain

$$|\, f(\hat{u}_{j-1}, \hat{u}_j^*) - f(u_{j-1}, u_j^*)\,| \leq M(|\,\hat{u}_{j-1} - u_{j-1}| + |\hat{u}_j - u_j| + \Delta_j + \hat{\Delta}_j). \tag{92}$$

The first two terms on the right are $O(K^{-1})$ by (65) of Theorem 3. The second two terms are $O(K^{-1})$ by Theorem 1. Now $\hat{\Delta}_j^3 = L^3K^{-3}$. Thus, we have (91) is $O(K^{-4})$ and (88) can finally be written

$$\psi(\hat{u}_{j-1}, \hat{u}_j) - \psi(u_{j-1}, u_j) = 2\hat{\Delta}_j(\hat{\Delta}_j - \Delta_j) + O(K^{-4}). \tag{93}$$

Now sum these equations from $j = 1$ to K:

$$\Psi[\hat{Q}_K] - \Psi[Q_K] = 2\sum_{j=1}^{K} \hat{\Delta}_j(\hat{\Delta}_j - \Delta_j) + O(K^{-3}). \tag{94}$$

A quick check shows the sum is identically zero, since $\hat{\Delta}_j = L/K$, a constant, and both $\hat{\Delta}_j$ and Δ_j sum to L. This completes the proof of Theorem 4. □

6. Conclusions

In the present paper we have considered the optimal partition schemes of locally 2nd-order cost functions. Under an optimal partition scheme, such cost functions tend to zero as $O(K^{-1})$. The mesh of the optimal partition also tends to zero as $O(K^{-1})$.

Our main result concerns a special sequence of partitions that are equidistant, where distance is defined using the square root of the second derivative of the cost function [5]. The main theorem is Theorem 6, which states that the difference between the values of the cost on the sequence of optimal partitions and on the sequence of equidistant partitions is order K^{-3}. This is highly surprising; we approximated the cost function locally to lowest order and find that added up over the K terms for the partition the sum agrees not only to lowest order but to an extra order which we had no right to expect. A first order agreement between the equidistant partition and the optimal partition is rather straightforward and has appeared long ago [1]. The second order agreement is much more subtle. In this regard, we mention that other suggested solutions for an asymptotically optimal sequence of partitions (ones based on an isoforce principle [6, 7]) do not agree even to first order. Our results further show that the j^{th} points in the equal distance partition and in the optimal partition match with an error of no more than $O(K^{-1})$.

Finding an optimal solution for a given finite K is more difficult than might first be imagined. It is easy to write down necessary conditions for optimality. They are even recursive in character, suggesting that a simple shooting method can insure consistency with the boundary requirement $t_K = b$ in a numerical solution. Unfortunately, even for smooth functions φ, this problem for fixed finite K can give rise to many local minima. The near-optimal (equidistant) partition circumvents all of the technical difficulties of finding the exact solution, and yet matches the optimal objective to $O(K^{-3})$.

The implications of these facts for optimally staged thermodynamic processes are many and will be pursued elsewhere. In particular, they explain the surprisingly good agreement between the minimal entropy production and the equal thermodynamic distance entropy production found in diabatic distillation columns [8].

References

[1] Nulton J, Salamon P, Andresen B, Anmin Q. Quasistatic Processes as Step Equilibrations. *Journal of Chemical Physics*, 83:334-338, 1985.

[2] Salamon P, Nulton J. The Geometry of Separation Processes: The Horse-Carrot Theorem for Steady Flow Systems. *Europhysics Letters*, 42:571-576, 1998.

[3] Salamon P, Komlos J, Andresen B, Nulton J. A Geometric View of Consumer Surplus with Non-Instantaneous Adjustment. *Mathematical Social Sciences*, 13:2, 1986.

[4] Flick JD, Salamon P, Andresen B. Metric Bounds on Losses in Adaptive Coding. *Information Sciences*, 42:239, 1987.

[5] George Ruppeiner. Riemannian geometry in thermodynamic fluctuation theory. *Reviews of Modern Physics*, 67:605–659, 1995.

[6] Sauar, E, Rivero, R, Kjelstrup, S, Lien, KM. Diabatic column optimization compared to isoforce columns. *Energy Conversion and Management*, 38:1777-1783, 1997.

[7] de Koeijer, GM, Kjelstrup, S, van der Kooi, HJ, Groß, B, Knoche, KF, Andersen, TR. Positioning Heat Exchangers in Binary Tray Distillation Using Isoforce Operation. Proceedings of ECOS99, ed. M. Ishida, Tokyo ISBN4-9980762-0-5 (1999) 471-476.

[8] Schaller M, Hoffmann K.H, Siragusa G, Salamon P, and Andresen B. Optimal Performance in Diabatic Distillation Columns submitted to Journal of Non-Equilibrium Thermodynamics.