1.2.1 Optimum calibration

An immediate application of the notion of statistical distinguishability occurs when we calibrate a scale of a measuring apparatus. We shall assume that the precision Δx of the measuring apparatus is known for all measured values x. Let x_0 be the first point of calibration. Where shall we put the next mark x_1? Obviously, the reasonable choice is such that x_0 and x_1 be distinguishable: $x_1 = x_0 + \Delta x_0 + \Delta x_1$. The subsequent calibration marks satisfy the distinguishability condition (1.1):

$$x_{\nu+1} = x_\nu + \Delta x_\nu + \Delta x_{\nu+1}, \quad (\nu = 0, 1, \ldots). \tag{1.3}$$

Hence, the resolutions of measurements and calibration marks will match.

These calibration marks are not equidistantly distributed on the scale x. We can, nevertheless, re-parameterize the scale by a new variable $\tilde{x}$ such that its standard deviation is constant:

$$\Delta \tilde{x} \equiv 1. \tag{1.4}$$

Then the calibration marks are thus located at the equidistant steps $\tilde{x}_\nu = \nu$ for $\nu = 0, 1, 2, \ldots$. The natural scale $\tilde{x}$ defines the so-called *statistical length* [4]. By construction, the statistical lengths of the confidence intervals equal 1, according to eq.(1.4). The statistical length of a bigger interval (x_i, x_f) is equal to the maximum number of non-overlapping confidence intervals between x_i and x_f which turns out to be

$$\ell_{if} = |\tilde{x}_f - \tilde{x}_i|. \tag{1.5}$$

We can formulate the principle of optimum calibration in such a way: the neighbouring marks should be separated by unit statistical lengths from each other.

Let us derive the natural scale $\tilde{x}$ of the relative frequency x. From the eqs.(1.2,1.4) and from the (asymptotic) relation $\Delta \tilde{x} = \frac{d\tilde{x}}{dx} \Delta x$ we obtain:

$$\tilde{x} = 2\sqrt{N}c. \tag{1.6}$$

where $c = \sqrt{x}$. The $\nu' th$ optimum calibration point is $\tilde{x}_0 + \nu$. This will corresponds to $x_\nu = (x_0 + \nu/\sqrt{4N})^2$ on the scale of relative frequencies. Eq.(1.5) yields the statistical length of an interval (x_i, x_f):

$$\ell_{if} = 2\sqrt{N}|c_i - c_f|. \tag{1.7}$$

Fig.2. The natural scale $\tilde{x}$ replaces x in such a way that the statistical lengths of the confidence intervals become equal to 1. The optimum calibration marks are separated by 1 on the new scale $\tilde{x}$.

1.2.2 Naive Optimum Control

We turn to another application of the concept of statistical length in the field of optimum control in a noisy environment. A system with a single continuous parameter x is to be driven from an initial state x_i into a given final state x_f. The standard deviation Δx of the parameter may depend on the current value x. Initially, the minimum significant change is just Δx_i. So, a cautious strategy might consist of a sequence of minimum, yet significant, steps from x_i until x_f: each step is equal to the local value of Δx. Thus the steps must correspond to the optimum calibration marks (1.3) starting from $x_i = x_0$.

On the natural scale $\tilde{x}$ each step will have unit length, according to eq.(1.4). If, furthermore, we perform a constant number v of steps per unit time then

$$\frac{\Delta \tilde{x}}{\Delta t} \equiv v. \tag{1.8}$$

This is the principle of *constant statistical speed* which seems to be a kind of cautious control in noisy environments.

Imagine, for instance, the inflation rate x in an economy whose financial policy is to decrease x from a higher value x_i to the lower one x_f. To minimize the risk of uncontrolled changes, one can exert minimum significant perturbations to the current economical system. For instance, the inflation rate can be decreased by Δx each year, where Δx is the variance of the given annual inflation x. This is a strategy at unit statistical speed. If necessary, the process can be made more intensive if we choose a higher constant velocity $|v| > 1$, i.e. we perform $|v|$ such steps per year.

The *ad hoc* principle of constant statistical speed and similar "cautious control schedules" will be formulated analytically in Section 1.5. We will see that such controls yield powerful tool in various tasks of process optimization.

1.2.3 More Parameters

We can easily extend the above statistical concepts to the case with more continuous parameters $\mathbf{x} = (x_1, x_2, \ldots, x_k, \ldots)$ characterize the given system. The simultaneous standard deviations Δx_k are correlated and the role of confidence intervals are played by multi-dimensional ellipsoidal confidence volumes. Two vectors $\mathbf{x}$ and $\mathbf{x}'$ are, by our convention, statistically *distinguishable* if their confidence ellipsoids contact each other.

Fig 3. Two vectors $\mathbf{x}$ and $\mathbf{x}'$ are considered statistically distinguishable if their confidence ellipsoids contact each other.

The optimum calibration of a curve in the multi-dimensional parameter space invokes considerations similar to the single parameter scale. Neighbouring calibration marks should be distinguishable. Their confidence volumes, centered along the curve, shall contact each other. The statistical length of the curve will be equal to the number of such confidence volumes. This length may completely depart from any apparent length of the curve. The statistical length depends largely on the variability of the ellipticities and orientations of the confidence volumes along the curve. The statistical length of a curve connecting two points $\mathbf{x}_i$ and $\mathbf{x}_f$ will depend on the curve itself. The minimum length can, as usual in geometry, be called the *statistical distance* between $\mathbf{x}_i$ and $\mathbf{x}_f$. The minimizing curve will correspond to a geodesic curve in the analytic theory. For the time being, however, we are discussing the empirical concepts. Of course, the naive optimum control at constant statistical speed, suggested for single variable, applies easily to the simultaneous control of parameters $\mathbf{x} = (x_1, x_2, \ldots)$ at constant statistical speed along a given curve. A new feature of the optimization is the choice of the shortest path (geodesic) to connect the initial and final parameters $\mathbf{x}_i, \mathbf{x}_f$.

We recall the re-scaling procedure which led to uniform confidence intervals (1.4) for a single variable x. In perfect analogy, one would attempt to use new variables $\tilde{\mathbf{x}} = (\tilde{x}_1, \tilde{x}_2, \ldots)$ such that their confidence volumes be hyper-spheres of unit radii. Hence, in these natural variables $\tilde{\mathbf{x}}$, the statistical lengths of curves are equal to the ordinary lengths. Consequently, statistical distances coincide with the corresponding Euclidean distances[2]:

$$\ell_{if} = \|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_f\|. \tag{1.9}$$

Minimum lengths curves, i.e. geodesics, are straight lines in $\tilde{\mathbf{x}}$.

The existence of the above natural parameterization is a delicate problem. Even at an empirical level we can see that sometimes the natural parameters $\tilde{\mathbf{x}}$ may *not* exist. The variation of the confidence ellipsoids from site to site may paralyze our attempts at constructing the natural parameterization. This problem, with all its complexity, is part of a paradigm. If one goes beyond the empirical considerations, it turns out [10, 11] that the parameter space $\mathbf{x}$ of statistical ensembles constitutes a manifold of Riemannian-geometry. If, in particular, this geometry is Euclidean then, and only then, will the natural parameters $\tilde{\mathbf{x}}$ exist.

Previously we considered, as an example, the relative frequency x of a single event in a large sample N. Now we are going to generalize the example for the relative frequencies $x_1, x_2, \ldots$ of a number of mutually exclusive events from the same large sample N. When the frequencies are

---

[2]The squared Euclidean norm of a vector $\mathbf{x}$ is defined as $\|\mathbf{x}\|^2 = \sum_k x_k^2$.

small, their variances Δx_k are independent and given by the same eq.(1.2), respectively for each x_k. The $k'th$ axis of the confidence ellipsoid is parallel to the coordinate axis x_k and its half-length is equal to

$$\Delta x_k = 2\sqrt{N}c_k, \tag{1.10}$$

with the notation $c_k = \sqrt{x_k}$. Introducing the natural parameters $\tilde{\mathbf{x}}$ by eq.(1.6), we achieve that all $\Delta\tilde{x}_k$ are of unit length: the confidence volumes become unit spheres in the parameter space $\tilde{\mathbf{x}}$. Consequently, one can apply eq.(1.9) to write down the statistical distance between arbitrary two parameters $\mathbf{x}_i$ and $\mathbf{x}_f$:

$$\ell_{if} = 2\sqrt{N}\|\mathbf{c}_f - \mathbf{c}_i\|. \tag{1.11}$$

1.3   Theory of Statistical Distance

The concept of statistical distinguishability, considered on intuitive grounds in the previous section, implies a natural geometry on the space of statistical ensembles. In the present section we give an insight into the general structure of this geometry for classical as well as quantum ensembles.

1.3.1   Classical Statistics

We consider discrete classical statistical ensembles. They are parameterized by normalized probability distributions $\mathbf{p} = (p_1, p_2, \ldots, p_k, \ldots)$ corresponding to a complete set of mutually exclusive events. The parameter space is a hyper-plane:

$$\sum_k p_k = 1, \quad p_k \geq 0. \tag{1.12}$$

We can introduce an alternative parameterization $\mathbf{c} = (c_1, c_2, \ldots)$ where the components of the vector $\mathbf{c}$ are the *square roots of the probabilities*: $c_k = \sqrt{p_k}$ for $k = 1, 2, \ldots$. Then the parameter space (1.12) becomes spherical:

$$\|\mathbf{c}\| = 1, \quad c_k \geq 0. \tag{1.13}$$

i.e. a sector on the surface of the unit hyper-sphere [11].

We note that in practice the theoretical probabilities p_k appear as relative frequencies x_k of the corresponding events in a given sample. Regarding the statistical distinguishability of our ensembles, the probabilities (1.12) will be treated as relative frequencies in a large sample of size N. Then we can apply the equations and considerations of the previous section. We

know from eq.(1.6) that in natural parameterization $\tilde{\mathbf{x}} = 2\sqrt{N}\mathbf{c}$ the confidence volumes are just the unit spheres. In this parameterization the statistical distance is identical to the Euclidean distance (1.11). One would conclude that the geometry of the parameter space of discrete ensembles is Euclidean. This is not exactly the case! Nevertheless, the true geometry is amazing as we see below.

There has been a loophole in our derivation of eq.(1.11). We assumed that the events and the corresponding relative frequencies $x_1, x_2, \ldots$ were independent and this holds indeed for small values of the relative frequencies. If, however, we identify them by the normalized probabilities (1.12) then the very constraint of normalization puts constraint upon them. The Euclidean geometry (1.11), when extended deliberately[3] for all possible values of $x_k = p_k = (c_k)^2$, induces a non-Euclidean (spherical) geometry on the parameter space (1.13):

$$\ell_{if} = 2\sqrt{N}\arccos(\mathbf{c}_i\mathbf{c}_f) \tag{1.14}$$

which is just $2\sqrt{N}$ times the angle[4] between the two unit vectors $\mathbf{c}_i$ and $\mathbf{c}_f$ [11, 6].

This result is remarkable! From purely classical considerations we have found something very similar to the quantum mechanical formalism. We found that the natural parameters of statistical ensembles are unit *vectors* $\mathbf{c} = (c_1, c_2, \ldots)$. The ensemble's usual parameters, i.e. the normalized probabilities $p_1, p_2, \ldots$ are equal to the *square* of the vector's components:

$$p_1 = (c_1)^2, \quad p_2 = (c_2)^2, \quad \ldots . \tag{1.15}$$

The resemblance to quantum mechanics is puzzling! Note, however, that the analogy is not perfect: our state vectors $\mathbf{c}$ are always real [6].

We close this section with an interesting relation between the statistical distance and the entropy function $s(\mathbf{p}) = -\sum p_k \log p_k$. Eqs. (1.14,1.15) yields the following result for the squared statistical distance $d\ell$ between two infinetisimally close distributions $\mathbf{p}$ and $\mathbf{p} + d\mathbf{p}$:

$$d\ell^2 = N\sum_k (dc_k)^2 = N\sum_k \frac{(dp_k)^2}{p_k}. \tag{1.16}$$

Anticipating the formalism of Riemannian geometry (Sec.1.4), we mention that the metric tensor, yielding this infinitesimal statistical length, turns

---

[3]Our derivation is illustrative. Rigorous calculations [12] confirm the distance (1.14).
[4]The angle γ between two vectors $\mathbf{a}, \mathbf{b}$ is defined via their scalar product $\mathbf{ab} = \sum_k a_k b_k = \cos\gamma\|\mathbf{a}\|\|\mathbf{b}\|$.

out to be the second derivative of the entropy function:

$$g^{ik}(\mathbf{p}) = N\delta_{ik}\frac{1}{p_k} = -N\frac{\partial^2 s(\mathbf{p})}{\partial p_i \partial p_k}. \tag{1.17}$$

1.3.2 Quantum Statistics

The concept of distinguishability can be interpreted for quantum ensembles as well. The same considerations that we have applied to classical ensembles extends to them. The result is a unique notion of statistical distance. In the present development we do not going to present develop the details of specific problems which make, as usual, the quantum case more subtle than the classical one. We present only the resulting equations.

The so-called *pure* quantum ensembles are characterized by normalized *complex* state vectors $\mathbf{c}$. Wootters [6] pointed out first that the statistical distance between pure quantum ensembles

$$\ell_{if} = 2\sqrt{N}\arccos|\mathbf{c}_i^\star\mathbf{c}_f| \tag{1.18}$$

coincides with the statistical distance (1.14) between classical ensembles, apart from the fact that the state vector components are now complex numbers.

General quantum ensembles are described by hermitian positive definite density matrices $\hat{\rho} \equiv \{\rho_{kl}\}$. The statistical squared distance between two ensembles $\hat{\rho}_i, \hat{\rho}_f$ has the following form:

$$\ell_{if}^2 = 8N\left(1 - tr\sqrt{\hat{\rho}_i^{1/2}\hat{\rho}_f\hat{\rho}_i^{1/2}}\right) \tag{1.19}$$

which was, with a different pre-factor, suggested by Bures [13]. (It can be proved that the distance remains the same if we interchange the two density matrices.)

It is instructive to study the distance between two ensembles whose density matrices commute with each other. In this case they have a common canonical basis where they both are diagonal. The nested square roots largely simplify:

$$\ell_{if} = 2\sqrt{2N}\sqrt{1 - tr\sqrt{\hat{\rho}_i\hat{\rho}_f}}. \tag{1.20}$$

Recall that the diagonals of the density matrices serve as classical probability distributions $\mathbf{p}_i$ and $\mathbf{p}_f$, respectively. This means that we can formally apply the classical distance (1.14) to our commuting density matrices, yielding[5]

$$\ell_{if}^{(cl)} = 2\sqrt{N}arccos\big(tr\sqrt{\hat{\rho}_i\hat{\rho}_f}\big). \tag{1.21}$$

---

[5] We apply the identity $\mathbf{p}_i\mathbf{p}_f = tr(\hat{\rho}_i\hat{\rho}_f)$.

Let's compare this to the quantum distance (1.20):

$$\frac{\ell_{if}}{4\sqrt{N}} = \sin\left(\frac{\ell_{if}^{(cl)}}{4\sqrt{N}}\right). \tag{1.22}$$

The quantum and classical distances tend to coincide for neighbouring ensembles. For distant ensembles the quantum distance is always smaller. This is not surprising geometrically. Quantum states are usually harder to distinguish from each other due to their overlap in Hilbert space. Furthermore, the shortest path between the two endpoint goes through commuting density matrices in case of the classical distance while it may find a shorter way through non-commuting matrices in the quantum case.

One can calculate the distance between infinitesimally close density matrices $\hat{\rho}$ and $\hat{\rho} + d\hat{\rho}$. For convenience, we use the canonical basis where $\rho_{kl} = \delta_{kl}p_k$. It is important to note that, while ρ_{kl} is now diagonal, its increment $d\rho_{kl}$ may be nondiagonal. The squared statistical distance (1.19) yields the following:

$$d\ell^2 = 2N \sum_{k,l} \frac{|d\rho_{kl}|^2}{p_k + p_l}. \tag{1.23}$$

The proof that this distance is in fact a measure of the Fisher statistical distinguishability (Sect.1.2) has been carried out by Braunstein and Caves [5]. Their proof shows that the Bures distance (1.19) is really the Fisher statistical distance applied this time to quantum statistical ensembles.

1.4 Riemannian Geometry

We have shown in the previous section that the natural geometry, reflecting statistical distinguishability of ensembles (probability distributions) is non-Euclidean. In various fields of applications, we are dealing with certain subclasses of probability distributions. Here we restrict ourselves to the case where the probability distributions are parameterized by a finite number of parameters. Gibbs distributions in statistical physics are particularly relevant: they underly the statistical geometry of thermodynamic parameter space.

1.4.1 Parameterized Statistics

Consider, for simplicity, discrete statistical ensembles whose probability distributions $\mathbf{p}$ are parameterized by a finite number of parameters[6] $\mathbf{y} =$

---

[6] We use coordinates with superscripts following the traditional notation in Riemannian geometry and tensor analysis.

$(y^1, y^2, \ldots, y^n)$. These probability distributions constitute an $n-$dimensional submanifold of the spherical hypersurface (1.13) and this hypersurface inherits a Riemannian geometry from the enveloping space. The resulting curvature and metric will, in general, change with $\mathbf{y}$.

We can no longer write down the statistical distance between two distant elements $\mathbf{p}(\mathbf{y}_i)$ and $\mathbf{p}(\mathbf{y}_f)$ in simple angular form as in eq.(1.14) since now the distance must be measured along a path staying entirely within the submanifold of distributions described by our parameters $\mathbf{y}$. Fortunately, the distance $d\ell$ between infinitesimally close elements of parameter values $\mathbf{y}$ and $\mathbf{y} + d\mathbf{y}$ remains the same. In particular, eq.(1.11) holds:

$$d\ell = 2\sqrt{N}\|d\mathbf{c}\|, \tag{1.24}$$

where

$$d\mathbf{c} = \sum_{k=1}^{n} \frac{\partial \mathbf{c}}{\partial y^k} dy^k. \tag{1.25}$$

If we introduce the following *Riemann metric* $g_{ik}(\mathbf{y})$ on the parameter space $\mathbf{y}$ [10]:

$$g_{ik} = 4N \sum_{r=1}^{n} \frac{\partial c_r}{\partial y^i} \frac{\partial c_r}{\partial y^k} \tag{1.26}$$

then we can write the infinitesimal statistical distance (1.24) in the standard Riemannian form

$$d\ell = \sqrt{\sum_{i=1}^{n} \sum_{k=1}^{n} g_{ik}(\mathbf{y}) dy^i dy^k}. \tag{1.27}$$

These equations define a Riemannian geometry on the manifold of parametrized discrete distributions $\mathbf{p}(\mathbf{y})$. Common Riemannian expressions will yield the statistical lengths/distances between arbitrary two distributions of respective parameter values $\mathbf{y}_i$ and $\mathbf{y}_f$.

It is straightforward to extend the above considerations beyond discrete probability distributions. Invoking the relation (1.15), we write the statistical metric (1.26) as follows:

$$\begin{aligned} g_{ik}(\mathbf{y}) &= 4N \sum_{r} \frac{\partial c_r}{\partial y^i} \frac{\partial c_r}{\partial y^k} \\ &= N \sum_{r} p_r \frac{\partial lnp_r}{\partial y^i} \frac{\partial lnp_r}{\partial y^k}. \end{aligned} \tag{1.28}$$

This has the following compound forms:

$$g_{ik}(\mathbf{y}) = N\left\langle \frac{\partial lnp(\mathbf{y})}{\partial y^i} \frac{\partial lnp(\mathbf{y})}{\partial y^k} \right\rangle = -N\left\langle \frac{\partial^2 lnp(\mathbf{y})}{\partial y^i \partial y^k} \right\rangle \tag{1.29}$$

where $\langle \ldots \rangle$ denotes expectation value calculated with $\mathbf{p}(\mathbf{y})$. These equations apply to continuous distributions as well. Let $p(\mathbf{\Gamma}; \mathbf{y})$ be the probability distribution of the continuous random variables $\mathbf{\Gamma}$, depending on the continuous parameters $\mathbf{y}$. The probability distributions are normalized:

$$\int p(\mathbf{\Gamma}; \mathbf{y})d\mathbf{\Gamma} = 1. \tag{1.30}$$

Eq.(1.29) takes the following form:

$$g_{ik}(\mathbf{y}) = -N \int \frac{\partial^2 lnp(\mathbf{\Gamma}; \mathbf{y})}{\partial y^i \partial y^k} p(\mathbf{\Gamma}; \mathbf{y})d\mathbf{\Gamma}. \tag{1.31}$$

This metric tensor is, apart from the scale factor N, identical to Fisher's *information matrix* [14]. It plays a particular role in mathematical statistics, namely in the theory of parameter estimation. This metric was also used by Amari [15] as the metric for his *information geometry* (see also Chentzov [16]).

1.4.2   From Gibbs Statistics to Thermodynamics

Phenomenological thermodynamics can, as is well known, be derived from the statistical physics of dynamical systems. Let us start with a large dynamical system consisting of M moles of molecules. If the system is in equilibrium then, according to Gibbs, its phase point $\mathbf{\Gamma}$ follows the probability distribution

$$p(\mathbf{\Gamma}; \mathbf{y}) = \exp\big(-\phi(\mathbf{y}) - \mathbf{y}\mathbf{F}(\mathbf{\Gamma})\big), \tag{1.32}$$

where $\mathbf{y} = (y^1, y^2, \ldots, y^n)$ are the entropic intensive parameters of the equilibrium state, $\mathbf{F}(\mathbf{\Gamma}) = \big(F_1(\mathbf{\Gamma}), F_2(\mathbf{\Gamma}), \ldots, F_n(\mathbf{\Gamma})\big)$ are the conjugated conserved dynamic quantities. The function $\phi(\mathbf{y})$ assures the normalization (1.30). When the size M of the dynamic system goes to infinity, the ratio $\phi(\mathbf{y})/M$ converges to the phenomenological *thermodynamic potential* (per moles) $\varphi(\mathbf{y})$ of the system:

$$\frac{\phi(\mathbf{y})}{M} \to \varphi(\mathbf{y}). \tag{1.33}$$

This *thermodynamic limit* is at the heart of the Gibbs theory.

Let us calculate the statistical metric of the Gibbs ensembles (1.32)! On substituting eq.(1.32) into the rightmost expression in eq.(1.29), we get

$$g_{ik}(\mathbf{y}) = N\frac{\partial^2 \phi(\mathbf{y})}{\partial y^i \partial y^k}. \tag{1.34}$$

Assuming that the system is large enough to take the thermodynamic limit (1.33), we can replace the function ϕ in eq.(1.34) by $M\varphi$. Now, this would left us with a factor NM on the r.h.s. of the above equation. Recall that, in the theory of statistical distance, N was the sample size. In the present case it would mean the number of $M-$mole thermodynamic systems in the *same* equilibrium state. That makes no difference if we unify them into a single $NM-$mole system. Finally, we simply absorb the factor M into the number N characterizing the overall size (in moles) of the thermodynamic system. So we obtain:

$$g_{ik}(\mathbf{y}) = N\frac{\partial^2\varphi(\mathbf{y})}{\partial y^i\partial y^k}. \tag{1.35}$$

This metric, generating the statistical distance between Gibbs ensembles at various thermodynamic parameters $\mathbf{y}$, can be expressed by the second derivative of the thermodynamic potential. Note an important aspect of this result: the concept of statistical distance has induced a notion of distance between thermodynamic states. This is the *thermodynamic distance* anticipated by Weinhold [1] and turned by Ruppeiner [2] into the corresponding Riemannian geometry on the thermodynamic state space.

It is rather instructive to derive the Riemannian metric if we use extensive parameters[7] $x_k = -\partial\varphi/\partial y^k$ for $k = 1, 2, \ldots, n$, instead of the intensives. The result reads:

$$g^{ik}(\mathbf{x}) = -N\frac{\partial^2 s(\mathbf{x})}{\partial x_i\partial x_k}. \tag{1.36}$$

Here $s(\mathbf{x})$ is the specific entropy function, related to the specific thermodynamic potential $\varphi(\mathbf{y})$ by the Legendre transformation $s = \mathbf{x}\mathbf{y} - \varphi$.

Note that the forms of our metric in eq. (1.36) can be termed macroscopic since now our metric matrix is just the matrix of second partial derivatives of the macroscopic entropy with respect to the extensive variables of the macroscopic system. It is interesting to note that the metric matrix is the second derivative in both eqs. (1.35) and (1.36). We note that this holds only for the entropy and its complete Legendre transform and not for any partial Legendre transforms [17, 18].

Another surprise is the fact that the metric matrix is the second derivative of the entropy both in a macroscopic description (1.36) and in a microscopic description (1.17). This is surprising in light of the fact that metric matrices and second derivative matrices transform differently under a change of coordinates [19].

---

[7]In thermodynamics we alter the traditional covariant notation of Riemannian geometry in a particular way. By our convention, the intensive parameters y^k bear upper labels but the extensive parameters x_k bear lower ones. Consequently, the covariant metric tensor will have lower labels in intesive and upper labels in extensive coordinates.

For our discussion of multiphase systems of variable composition, it is convenient to have an expression for our metric in terms of the macroscopic extensive variables $X_k = Nx_k, \quad k = 1 \ldots, n$ and $X_{n+1} = N$ instead of the *specific* extensive parameters[8] $s, x_1, \ldots, x_n$. This transforms our thermodynamic metric (1.36) into the following form[9]:

$$g^{ik}(\mathbf{X}) = -\frac{\partial^2 S(\mathbf{X})}{\partial X_i \partial X_k}, \tag{1.37}$$

where $S(\mathbf{X}) = Ns(\mathbf{x})$ is the extensive entropy function of the system. Since

$$\frac{\partial^2 S}{\partial X_{n+1}^2} = \frac{\partial^2 S}{\partial N^2} = 0, \tag{1.38}$$

it follows that $g^{ik}(\mathbf{X})$ is degenerate, i.e. there exist directions along which the metric measures a zero distance. Such directions always correspond to scaling one phase of the system. If the scale of the system is not fixed, the structure is only semi-Riemannian. Null directions result from the linear growth of the entropy as we scale any one phase. This linearity makes the second derivative, i.e., the components of the metric tensor, vanish along such directions. As remarked in Weinhold's original papers [20], forming an appropriate combination of such null directions by simultaneously scaling two phases of a pure substance can represent a phase transition. We will return to a discussion of the use of such null directions in the following sections.

1.5 Relevance of Riemannian Geometry in Macroscopic Thermodynamics

For macroscopic thermodynamics, the use of a closely related metric

$$G^{ik}(\mathbf{Z}) = \frac{\partial^2 U(\mathbf{Z})}{\partial Z_i \partial Z_k} \tag{1.39}$$

introduced by Weinhold [1] preceded the use of the metric described above. Here U is the internal energy and $\mathbf{Z} = (S, V, N_1, N_2, ...) \in \mathbb{R}^{n+1}$ is the vector of extensive variables of the system in the energy representation for

---

[8]We refer to an extensive quantity divided by the mole number as specific extensive. A related terminology is the extensive density which is the extensive quantity divided by the volume. For simple equilibrium systems the number of independent extensive quantities exceeds the number of specific extensive quantities by 1.

[9]We use the same letter g and adopt the functional notation $g(\mathbf{X})$ to indicate the fact that the actual matrix will depend on our choice of parameters.

which $U = U(\mathbf{Z})$ constitutes complete information [21]. Weinhold's papers used the geometry only locally to conveniently express relationships among differential changes in our state variables. He also suggested its possible use as a Riemannian metric. Soon thereafter, Gait et al. [22] recognized that distances computed for the ideal gas using the metric G represented known expressions for the changes in the kinetic energy of the molecules in a gas resulting from the passing of a shock wave. This led to the recognition of the connection between geometry and dissipation as developed below.

At about the same time, Ruppeiner introduced the use of the metric g in eq. (1.36) for extending the scales accessible to thermodynamic fluctuation theory. While the metrics g in eq. (1.36) and G in eq. (1.39) are conformally equivalent[10] [18], the metric g turns out to be more fundamental. Similar to the role of the energy representation elsewhere in thermodynamics, G serves mainly as a convenient device for calculating $g(\mathbf{X})$, with $\mathbf{X} = (U, V, N_1, N_2, ...) \in I\!\!R^{n+1}$ the vector of extensive variables in the entropy representation for which $S = S(\mathbf{X})$ constitutes complete information [21].

We begin our treatment of macroscopic applications of the Riemannian structure with a discussion of fluctuations since this follows most closely from the arguments in the previous section. We will then turn our attention to dissipation and a discussion of several horse-carrot theorems which relate the dissipation associated with coaxing a system to traverse a given sequence of states. The applications depend on the second order expansion of the entropy and thus on the identity between the metric and the second derivative of entropy.

1.5.1 A Covariant Fluctuation Theory

Consider the traditional expression for a fluctuation in a system of size[11] V inside a system of infinite size $V_0 = \infty$ with densities $\mathbf{x}_0$. The likelihood of the subsystem having densities $\mathbf{x}$ is given by the Einstein-Smoluchowski theory as

$$P(\mathbf{x}, V|\mathbf{x}_0, \infty)d^n\mathbf{x} = Cexp(S(\mathbf{x}, \mathbf{x}_0)/k_B)d^n\mathbf{x}, \qquad (1.40)$$

where C is a normalization constant, k_B is Boltzmann's constant, and $S(\mathbf{x}, \mathbf{x}_0)$ is the total entropy of the reservoir at extensive densities $\mathbf{x}_0$ containing the subsystem of finite volume V at state $\mathbf{x}$. This expression is valid for small fluctuations (i.e. at large volumes V) but it turns out that,

---

[10]Two metrics are conformally equivalent iff the squares of the length elements differ by a (possibly position dependent) scale factor.

[11]Here we follow Ruppeiner [23, 24] in using the volume to set our scale.

in a subtle way, it also contains the statistics of larger fluctuations (i.e. at smaller volumes V).

At infinite volume there are no fluctuations at all:

$$P(\mathbf{x}, \infty | \mathbf{x}_0, \infty) d^n \mathbf{x} = \delta(\mathbf{x} - \mathbf{x_0}). \qquad (1.41)$$

For large finite volumes the eq. (1.40) leads to the Gaussian approximation:

$$
\begin{aligned}
P(\mathbf{x}, V | \mathbf{x}_0, \infty) &= \\
&= Cexp\Big(\frac{V}{2k_B} \sum_{i,k} \frac{\partial^2 s(\mathbf{x}_0)}{\partial x_i \partial x_k}(x - x_0)_i (x - x_0)_k\Big) = \\
&\equiv Cexp\Big(-\frac{1}{2k_B} \sum_{i,k} g^{ik}(\mathbf{x}_0)(x - x_0)_i (x - x_0)_k\Big) (1.42)
\end{aligned}
$$

where the exponent becomes proportional to the square of the statistical (or themodynamic) distance measured from the equilibrium value. Recall that this length element is the natural scale for measuring the size of fluctuations.

Ruppeiner's important observation [2] was that although eq. (1.40) depends on which parameters $\mathbf{x}$ we use to define our state, by way of the volume form $d^n \mathbf{x}$, its Gaussian approximation (1.42) is invariant under reparameterization[12] and thus avoids this unphysical dependence. Hence, we must restore this invariance when we extend its validity for smaller volumes V.

Ruppeiner [23, 24] and Diósi and Lukács [25] used the Gaussian fluctuation theory as a starting point for an improved covariant theory of fluctuations, valid also for smaller volumes.

The physical intuition leading to the improved theory comes by considering a nested sequence of systems. We begin with an equilibrium system in the thermodynamic limit $V = \infty$ at the state $\mathbf{x}_0$. As the volume of the subsystem we consider gets smaller, its fluctuations depend on the state of its immediate surroundings. In this way we get a Markov process for fluctuations inside fluctuations inside fluctuations .... It is worth noting that the Gaussian approximation (1.42) provides the *exact* transition rates. The corresponding Chapman-Kolmogorov equation describes how a particular fluctuation in a system of size V depends on the state of the system at slightly larger size $V' = V + dV$.

Note that the role of time in the Chapman-Kolmogorov equation is played by $1/V$ which starts at 0, where the distribution is the delta-function (1.41), and then takes on small values, where the Gaussian distribution

---

[12]In this approximation, the Jacobian matrix of a coordinate transformation can be taken as constant for $\mathbf{x}$ sufficiently near $\mathbf{x}_0$.

(1.42) still holds, and then tends to infinity as the system size becomes small. This is the newly explored regime where a covariant description of fluctuations emerges.

The Chapman-Kolmogorov equation takes the form of a covariant Fokker-Planck equation. Its ultimate form, assuring all conservation laws, was derived by Diósi and Lukács [25]. Here, to abandon using covariant differential calculus, we present the equation in extensive (density) parameters:

$$\frac{\partial}{\partial V^{-1}} P(\mathbf{x}, V|\mathbf{x}_0, \infty) = \frac{1}{k_B} \sum_{i,k} \left(\frac{\partial^2}{\partial x_i \partial x_k} g(\mathbf{x})^{ik} P(\mathbf{x}, V|\mathbf{x}_0, \infty) \right). \quad (1.43)$$

¿From the initial distribution (1.41) at $V^{-1} = 0$, this Fokker-Planck equation evolves the distribution function of thermodynamic fluctuations at all finite volumes V. By construction, the equation provides initiallly the Gaussian distribution (1.42), and preserves normalization and the mean values of the extensive variables $\mathbf{x}$. This latter assures the fulfillment of the conservation laws.

For small fluctuations, Gaussian fluctuation theory does well. It seems that it also yields an overall covariant formalism which improves the match with experiment for fluctuations of moderate size [7]. This is especially useful for understanding system behavior near the critical point where fluctuations become large.

Requiring covariance of such partial differential equations along with some hypotheses connecting the Riemannian curvature and the free energy yields euqtions of state connecting the critical exponents [7].

1.5.2 Entropy Production

At the macroscopic level, the Riemannian structure introduced above is intimately connected with entropy production. Let $\mathbf{X} = (U, V, N_1, N_2, ...) \in I\!\!R^{n+1}$ and $\tilde{\mathbf{X}} = (\tilde{U}, \tilde{V}, \tilde{N}_1, \tilde{N}_2, ...) \in I\!\!R^{n+1}$ be the vectors of extensive variables of systems A and $\tilde{A}$ respecitvely. In an interaction between these two systems in which the infinitesimal vector of flows $d\mathbf{X}$ moves from $\tilde{A}$ to A, the entropy production is

$$dS_u = dS_A + dS_{\tilde{A}} = \sum_{i=1}^{n} (Y_i - \tilde{Y}_i) dX_i \quad (1.44)$$

where $\mathbf{Y} = \partial S/\partial \mathbf{X} = (1/T, p/T, \mu_1/T, \mu_2/T...)$, and we have made use of the conservation laws $dX_i = -d\tilde{X}_i$, $i = 1, ...n$. Eq. (1.44) is the familiar flow-times-force expression for entropy production and, as we will see, bears

a close resemblance to the length element $d\ell^2$ in our Riemannian geometry. To emphasize this similarity, we rewrite eq. (1.44) in the form

$$dS_u = -\sum_{i=1}^{n} \Delta Y_i dX_i = -\Delta \mathbf{Y} \cdot d\mathbf{X}. \tag{1.45}$$

where the $\Delta \mathbf{Y} = \tilde{\mathbf{Y}} - \mathbf{Y}$. Note that this sum must be positive by the second law.

1.5.3 The Metric as a Symmetric Product

We now express our length element

$$d\ell^2 = \sum_{j=1}^{n}\sum_{i=1}^{n} g(\mathbf{X})^{ij} dX_i dX_j \tag{1.46}$$

as a symmetric product. Since

$$dY_j \;=\; \sum_{i=1}^{n} \frac{\partial Y_j}{\partial X_i} dX_i \tag{1.47}$$

$$=\; \sum_{i=1}^{n} \frac{\partial^2 S}{\partial X_i \partial X_j} dX_i \tag{1.48}$$

$$=\; -\sum_{i=1}^{n} g(\mathbf{X})^{ij} dX_i, \tag{1.49}$$

we can write the length element as

$$d\ell^2 = -d\mathbf{Y} d\mathbf{X} \tag{1.50}$$

Note the similarity between this expression and eq. (1.45). Note also that while $d\mathbf{Y}$ and $\Delta \mathbf{Y}$ look similar, they represent very different quantities: $d\mathbf{Y}$ is an infinitesimal change in the state of system A, while $\Delta \mathbf{Y}$ is the difference $\tilde{\mathbf{Y}} - \mathbf{Y}$.

¿From the symmetric product $-d\mathbf{Y} d\mathbf{X}$ in eq. (1.50), it is easy to obtain the conformal equivalence between the metrics G and $g(\mathbf{X})$. On substituting

$$dX_1 \;=\; dU \;=\; TdS + \sum_{i=2}^{n} W_i dZ_i, \tag{1.51}$$

$$W_i \;=\; \frac{\partial U}{\partial Z_i} \;=\; -TY_i, \quad i = 2, .., n \tag{1.52}$$

$$Z_i \;=\; X_i, \qquad\qquad i = 2, ..., n. \tag{1.53}$$

into eq. (1.50) we obtain on rearrangement

$$d\ell^2 \quad = \quad -d\mathbf{Y}d\mathbf{X} \quad = \quad \frac{d\mathbf{W}d\mathbf{Z}}{T} \tag{1.54}$$

$$= \quad \frac{G^{ij}dZ_idZ_j}{T} \quad = \quad \frac{dL^2}{T}. \tag{1.55}$$

Eqs. (1.54) and (1.55) are the infinitesimal form of the Gouy-Stodola theorem [26] expressing the well known relationship between loss of availability at a temperature T and the associated entropy production. This fact will become more apparent after our discussion of the discrete horse-carrot theorem.

1.5.4 The group of transformations

The conformal equivalence of the geometries defined by the second derivatives of U, S, and ϕ leads naturally to the question of what other potential functions $\wp$ one might consider with the property that the metric $g(\mathbf{X})$ is a multiple of the second derivative matrix of $\wp$ with respect to $\wp$'s natural variables. The question is elegantly posed using the formalism introduced by Hermann [27]. Define an n degree of freedom thermodynamic system as a maximal integral submanifold of a contact form[13] ω on a space of dimension $2n+1$. In usual coordinates this takes the form $\omega = dS - \sum_{i=1}^{n} Y_i dX_i$. Asking for a maximal integral submanifold is asking for an n dimensional surface on which the differential expression of the first law, $\omega = 0$, holds. The additional structure implied by the symmetric two-form $\eta = d\mathbf{Y}d\mathbf{X}$ gives an interesting class of manifolds [28, 29, 30]. It turns out that there are very many potentials $\wp$. The set of such potentials can be characterized by considering the group of coordinate transformations which preserve ω and η up to scale factors. The group turns out to equal the semidirect product of the integers modulo 2, Z_2, the multiplicative group of non-zero real numbers, $\mathbb{R}^*$, the general linear group, $Gl(n)$, and the Heisenberg group, $H(n)$. If we ask that η be preserved without scaling, the group shrinks to just the semidirect product $Gl(n) \otimes_\alpha H(n)$ but which excludes the transformation from S to either U or ϕ.

While this shows there are many possible such potentials, they have not been extensively applied. Diosi and Lukacs showed that the action of the renormalization operator belongs to this group [25]. Mrugala [28] has applied these contact transformations to find laws of corresponding states. More recently, Eu [31] has used them to study Pfaffian formulations of uncompensated heat ??.

---

[13]A nowhere vanishing differential form of maximal rank.

1.5.5   Dissipation in a small equilibration

There is an important special case for which there is a close relationship between $\Delta \mathbf{Y}$ and $d\mathbf{Y}$ in eqs. (1.45) and (1.50). This is the case of a small equilibration with a bath. Let us take $\tilde{A}$ sufficiently large that any changes in its intensive variables $\tilde{\mathbf{Y}}$ can be neglected. Furthermore we allow A to equilibrate to $\tilde{A}$ so the final values of $\mathbf{Y}$ equal $\tilde{\mathbf{Y}}$. Integrating to equilibrium gives

$$\Delta S_u = \int dS_u = \int -\Delta \mathbf{Y} d\mathbf{X} \tag{1.56}$$

which in light of eq. (1.49) becomes

$$\Delta S_u = \int \Delta \mathbf{X}^t g(\mathbf{X}) d\mathbf{X} \tag{1.57}$$

to first order in $\Delta \mathbf{X} = \mathbf{X}_0 - \mathbf{X}$ where $\mathbf{X}_0$ is the vector of extensive variables of system A after equilibration with the bath $\tilde{A}$. To this order we may take the metric matrix $g(\mathbf{X})$ to be constant in which case eq. (1.57) can be integrated to give

$$\Delta S_u = \frac{1}{2} \ \Delta \mathbf{X}^t g(\mathbf{X}) \Delta \mathbf{X} \tag{1.58}$$

$$= \frac{1}{2} \ \Delta \ell^2. \tag{1.59}$$

1.5.6   The discrete horse carrot theorem

The discrete horse-carrot theorem follows at once from the general expression (1.59) expressing the relationship between the length of a small equilibration and the corresponding entropy production. Consider a path in the state space of system A and the process whereby we select the states of k baths to match the system's intensive variables $\mathbf{Y}$ at k points along this path. The discrete horse-carrot theorem answers the question: How should the states be chosen so as to minimize the total entropy produced in the k successive equilibrations which bring the system to equilibrium with the k successive baths along the path? For large k the answer is simply that one should place the baths equidistant in the geometry given by $g(\mathbf{X})$. This follows by noting that the total entropy production is given by

$$\Delta S_u = \sum_{l=1}^{k} \Delta_l S_u = \frac{1}{2} \ \sum_{l=1}^{k} \Delta_l \ell^2 \tag{1.60}$$

which is to be minimized while fixing the total length

$$\ell = \sum_{l=1}^{k} \Delta_l \ell. \tag{1.61}$$

This minimization is easily handled using Lagrange multipliers to give

$$\Delta_l \ell = constant = \frac{\ell}{k}. \tag{1.62}$$

Substituting this back into eq. (1.60) gives the horse carrot inequality

$$\Delta S_u \geq \Delta S_u^{min} = \frac{\ell^2}{2k}. \tag{1.63}$$

A more thorough analysis including the dynamics of incomplete relaxation is possible. [9]. For large times the analysis tells us to allocate the same number of relaxation times to each equilibration. Thus the minimum entropy producing way to bring the system in a finite time along a given path using a fixed number k of intermediate equilibrations is to make the steps equidistant with a constant number of relaxation times alloted for each step. This is the origin of the idea of constant thermodynamic speed $v = d\ell/d\xi$, where $d\xi = dt/\epsilon$, t is time, and ϵ is the relaxation time of the system. The optimality of constant thermodynamic speed for a k-step process is hereby established. The optimality of this control in other contexts has led to some confusion as we discuss further below.

Some comments on loss of availability

An expression entirely analogous to our eq. (1.59) can be derived for the loss of availability ΔA_u in a small equilibration.

$$\Delta A_u = \frac{1}{2} \Delta\mathbf{Z}^t G \Delta\mathbf{Z} \tag{1.64}$$

$$= \frac{1}{2} \Delta L^2. \tag{1.65}$$

We can now see more clearly why we refered to the conformal equivalence of the two metrics given by the second derivatives of S and U expressed in eq. (1.55) as the differential form of the Gouy-Stodola theorem

$$\Delta A_u = -T_a \Delta S_u \tag{1.66}$$

where T_a is the temperature of the atmosphere. For the infinitesimal process in eq. (1.55), the role of the large bath is played by $\tilde{A}$ rather than

the atmosphere. The ambiguity of where the heat equivalent of the lost availability ends up, severely limits the usefulness of the analogous horse carrot inequality for ΔA_u,

$$\Delta A_u \geq \frac{L^2}{2k}. \tag{1.67}$$

For isothermal processes the two inequalities (1.63) and (1.67) are equivalent. For non-isothermal processes the lengths given by ℓ and L are not simply related. Although the lengths given by L can be of interest for system control in certain circumstances [32], dL^2 is more often useful as a device for calculating $d\ell^2$ using eq. (1.55) as we illustrate below.

1.5.7 The continuous horse carrot theorem

In this section we treat the continuum version of the discrete control considered above. The problem is now as follows: Given that system A traversed the path $\mathbf{X}(t)$, $t \in [0, \tau]$ how much entropy production had to occur? For sufficiently large τ, this question has a very similar answer to what we found for the discrete process although the optimal control turns out to be constant entropy production rate rather than constant thermodynamic speed. We assume that we can control system $\tilde{A}$ reversibly and that system A is affected only indirectly through its contact with $\tilde{A}$. Our argument proceeds from the integral form of eq. (1.45) for the total entropy production

$$\Delta S_u = \int_0^\tau dS_u = -\int_0^\tau \Delta \mathbf{Y} d\mathbf{X}. \tag{1.68}$$

Some basic expressions connecting dissipation and geometry

There are a number of interesting rearrangements of eq. (1.68) for the total dissipation which reveal connections between this dissipation and our geometry. Define $\mathbf{X}_e$ by the formula

$$-(\tilde{\mathbf{Y}} - \mathbf{Y}) = g(\mathbf{X})(\mathbf{X}_e - \mathbf{X}). \tag{1.69}$$

Since $g(\mathbf{X})$ is not necessarily invertible, the scale of different homogeneous phases in $\mathbf{X}_e$ must be set separately by specifying how these scales of A evolve. Then for $\mathbf{X}_e$ close to $\mathbf{X}$, we can interpret $\mathbf{X}_e$ as the state of A which would minimize the entropy production rate in contact with the current state of $\tilde{A}$ subject to the constraint of keeping A's state on the line through $\mathbf{X}$ in the direction $d\mathbf{X}$. On substituting (1.69) into (1.68) and replacing $d\mathbf{X}$ by $\frac{d\mathbf{X}}{d\ell} d\ell$ we have

$$\Delta S_u = \int_0^\tau (\mathbf{X}_e - \mathbf{X})^t g(\mathbf{X}) \frac{d\mathbf{X}}{d\ell} d\ell. \tag{1.70}$$

Recall that a metric on a vector space defines a dot product [33]. The integrand in eq. (1.70) is the dot product using the metric $g(\mathbf{X})$ of the deviation $\mathbf{X}_e - \mathbf{X}$ and the unit tangent vector $\frac{d\mathbf{X}}{d\ell}$. Thus we may interpret this integrand as the distance between our current state and the state the system is trying to reach, projected onto the direction of $d\mathbf{X}$. We call this projected lag distance $D = dS_u/d\ell$ and see by the mean value theorem that we can write the total dissipation as

$$\Delta S_u = \int_0^\ell D d\ell = \bar{D}\ell. \tag{1.71}$$

This gives an expression for the total entropy production as the product of the mean distance to equilibrium and the total distance traversed.

A second interesting relation can be found by considering the quantity

$$\epsilon = \frac{dS_u/dt}{(d\ell/dt)^2}. \tag{1.72}$$

Note that ϵ has the units of time. In fact, for a sufficiently slow process with separable time scales, ϵ is just the relaxation time. We can see this by writing

$$\frac{d\mathbf{X}}{dt} = (\mathbf{X}_e - \mathbf{X})/\epsilon. \tag{1.73}$$

Note that by our definition of $\mathbf{X}_e$, $d\mathbf{X}/dt$ and $\mathbf{X}_e - \mathbf{X}$ must be in the same direction and hence must be proportional. If our dynamics is sufficiently slow and the time scales are separable, then all but the slowest mode of our system must equilibrate essentially instantaneously and thus $\mathbf{X}_0 - \mathbf{X}$ must be proportional to $d\mathbf{X}/dt$, i.e. $\mathbf{X}_e = \mathbf{X}_0$. With or without our assumptions of slow process and separable time scales, the definition in eq. (1.72) allows us to express the total entropy production as

$$\Delta S_u = \int_0^\tau \epsilon \frac{d\mathbf{X}}{dt}^t g(\mathbf{X}) \frac{d\mathbf{X}}{dt} dt = \int_0^\tau \epsilon \left(\frac{d\ell}{dt}\right)^2 dt. \tag{1.74}$$

We can again apply the mean value theorem, to give the alternative form

$$\Delta S_u = \bar{\epsilon} \int_0^\tau \frac{d\mathbf{X}}{dt}^t g(\mathbf{X}) \frac{d\mathbf{X}}{dt} dt = \bar{\epsilon} \int_0^\tau \left(\frac{d\ell}{dt}\right)^2 dt \tag{1.75}$$

A third interesting expression for ΔS_u results if we change parameters along the path $X(t)$ and express our dissipation integral in terms of the number of relaxations ξ. Recall that $d\xi = dt/\epsilon$ and so our integral becomes

$$\Delta S_u = \int_0^\Xi \frac{d\mathbf{X}}{d\xi}^t g(\mathbf{X}) \frac{d\mathbf{X}}{d\xi} d\xi = \int_0^\Xi \left(\frac{d\ell}{d\xi}\right)^2 d\xi. \tag{1.76}$$

where

$$\Xi = \int_0^\tau d\xi = \int_0^\tau \frac{dt}{\epsilon} \tag{1.77}$$

All our expressions in this section are valid generally. Nevertheless, without our hypotheses of a sufficiently slow process with separable time scales in which A and $\tilde{A}$ are near equilibrium with each other at each instant, the physical meaning of D and ϵ are merely formal [34].

For our fourth and final version of the integrated dissipation ΔS_u, we start from the Onsager-Prigogine type of linearized flux-force relationship

$$\frac{d\mathbf{X}}{dt} = \gamma \Delta \mathbf{Y} \tag{1.78}$$

where γ is the matrix of kinetic coefficients [35]. Note that γ must be symmetric and positive definite. If we solve this for $\Delta \mathbf{Y}$ and substitute into eq. (1.68) for the entropy production, we get

$$\Delta S_u = \int_0^\tau \frac{d\mathbf{X}}{dt}^t \gamma^{-1} \frac{d\mathbf{X}}{dt} dt = \int_0^\tau \left(\frac{d\lambda}{dt} \right)^2 dt. \tag{1.79}$$

which can again be interpreted as the integral of a speed squared. This time, the lengths λ are given by yet another metric (γ^{-1}). There is a fundamental difference between this metric and the ones which we have so far considered. The coefficients in γ are *kinetic* as opposed to *equilibrium* quantities. Stated another way, the metric coefficients in $g(\mathbf{X})$ are covariances while the coefficients in γ are time correlations.

A simple lemma from optimization

We now pause our development for a simple result which will show us how to minimize the entropy production and how to obtain a number of inequalities corresponding to the various expressions for ΔS_u derived in the previous subsection. While these inequalities can be obtained from the Cauchy-Shwartz inequality [8, 9], we use a variational argument here to emphasize their connection to optimal process control.

To minimize an integral of the form

$$\int_0^\tau f(x) \left(\frac{dx}{dt} \right)^2 dt \tag{1.80}$$

with given values of $x(0)$ and $x(\tau)$, the first order necessary conditions of Euler-Lagrange for our autonomous Lagrangian, K,

$$K = f(x) \left(\frac{dx}{dt} \right)^2 \tag{1.81}$$

give

$$K - \frac{dx}{dt}\frac{\partial K}{\partial \frac{dx}{dt}} = constant. \tag{1.82}$$

Substituting our expression for K in (1.81) into (1.82), we find that for optimality, the Lagrangian K should be constant.

An immediate and useful corrolary follows for the special case of eq. (1.80) with $f \equiv 1$. For this case

$$K = \left(\frac{dx}{dt}\right)^2 = constant \tag{1.83}$$

implies

$$\frac{dx}{dt} = constant = \Delta x/\tau. \tag{1.84}$$

The minimum value of the integral then simplifies to the right hand side of

$$\int_0^\tau \left(\frac{dx}{dt}\right)^2 dt \geq \Delta x^2/\tau. \tag{1.85}$$

This special case takes on particular importance when x is the arc length with respect to some metric M. In this case, letting $\mathbf{X}$ represent the coordinates in this space, the inequality becomes

$$\int_0^\tau \frac{d\mathbf{X}}{dt}^t M \frac{d\mathbf{X}}{dt} dt \geq \left(\int_0^\tau \sqrt{\frac{d\mathbf{X}}{dt}^t M \frac{d\mathbf{X}}{dt}} dt\right)^2 /\tau. \tag{1.86}$$

This general inequality leads directly to the fact that extremal curves for the speed squared coincide with extremal curves for the length (geodesics), a fact which is the starting point for Morse theory [36].

Applications of the lemma

Applying the lemma to eqs. (1.74) or (1.79) tells us that to minimize entropy production, we should proceed at a *constant entropy production rate*.

$$\dot{S}_u = \epsilon \left(\frac{d\ell}{dt}\right)^2 = \left(\frac{d\lambda}{dt}\right)^2 = constant \tag{1.87}$$

Applying the corollary for squared speed to eqs. (1.75) and (1.76) leads to the continuous versions of the horse-carrot inequality

$$\Delta S_u \quad \geq \quad \bar{\epsilon}\ell^2/\tau, \tag{1.88}$$

$$\Delta S_u \quad \geq \quad \ell^2/\Xi. \tag{1.89}$$

These inequalities bound the dissipation by the squared thermodynamic distance ℓ^2 divided by the number of relaxations. As such, they bear a strong resemblance to our discrete horse carrot inequality (1.63). Despite some confused claims in the literature, these generally valid inequalities do not say anything useful about how to minimize the entropy production in a given time. The averaging process that goes into $\bar{\epsilon}$ and Ξ for fixed total time[14] hides a dependence. For the process with a fixed number Ξ of relaxations, inequality (1.89) becomes sharp and the minimum entropy production strategy is to drive the process with constant thermodynamic speed $d\ell/d\xi$. Thus constant thermodynamic speed is optimal *for a given number of relaxations* in both the discrete and the continuous case. This fact was only recently elucidated independently by [37] and [38].

The factor of 1/2, present in the discrete but not in the continuous case, is real and comes from the fact that in the discrete case we repeatedly relax (almost all the way) to equilibrium while in the continuous case we maintain an approximately fixed distance. It follows that in the discrete case we are on the average about half as far from equilibrium and thus by eq. (1.71) should expect about half the dissipation.

Finally we note that our corollary about the integral of the squared speed applies also to eq. (1.79) and gives an alternative route to the last part of our minimum entropy production condition (1.87) for fixed time. It also gives an associated inequality

$$\Delta S_u \geq \lambda^2/\tau \tag{1.90}$$

The implications of the geometry of time correlations given by γ^{-1} is left for another paper. For preliminary results in this direction, the interested reader is refered to [39, 37, 38].

1.5.8   Cooling rates for simulated annealing

The above formalism has been applied to the control of the temperature in simulated annealing – an algorithm for solving global optimization problems. Here the idea is to associate a (usually fictitious) physical system with the optimization problem by identifying the objective function with the energy of such a system. We then simulate relaxations to equilibrium at a decreasing sequence of temperatures by a random walk over states using the Metropolis algorithm [40]. Much has been written about the ideal cooling rate [41, 42, 43]. Several authors have advocated a constant statistical velocity cooling schedule which keeps $d\ell/dt$ constant based on arguments

---

[14]Note that by eq. (1.77) we can consider Ξ as a time weighted harmonic mean ϵ.

along the lines of section 1.2.2. In fact, this schedule has been incorporated in the popular simulated annealing package known as Timberwolf [44, 45].

Although no direct connection has been established between entropy production and performance of the algorithm, such conjectures are tantalizing [46, 47]. Motivated by these conjectures, both constant thermodynamic speed and constant entropy production rate schedules have been tried. Empirically, it seems that the constant thermodynamic speed schedule outperforms others [41, 48, 49] but the difference for most systems is small.

The argument in favor of constant thermodynamic speed for these problems runs along the same lines as the argument already presented in section 1.2.2 in favor of constant statistical velocity: We cool as fast as possible consistent with the constraint of never being too far out of equilibrium [50]. The distance to equilibrium is measured by the metric $g(\mathbf{X})$ and this argument leads to keeping the distance D between the system and the bath constant. The role of the bath is played here by the parameter T used in the Metropolis algorithm. For this (thermodynamically one degree of freedom) system, the metric, $g(\mathbf{X}) = -d^2S/dE^2$, leads to the length element

$$
\begin{aligned}
d\ell &= \sqrt{-\frac{d^2S}{dE^2}dE^2} & (1.91) \\[2mm]
&= \sqrt{-\frac{d}{dE}(1/T)dE} & (1.92) \\[2mm]
&= \sqrt{\frac{1}{\sigma_E^2}dE} & (1.93) \\[2mm]
&= \frac{dE}{\sigma_E} & (1.94)
\end{aligned}
$$

where σ_E is the standard deviation of the energy and we have used the fact that the heat capacity dE/dT is equal to σ_E^2/T^2 [51]. Letting E_0 stand for the equilibrium energy of the system at the current temperature T, in the Metropolis algorithm, we get that $D = (E - E_0)/\sigma_E$ gives our distance measure of disequilibrium.

Keeping D constant keeps the system moving with its own time scale. If the cooling is sufficiently slow and the time scales are separable, the system moves with its own relaxation time. This gives one popular way to implement what has been called constant thermodynamic speed schedules. Alas, time scales are not separable in typical problems of simulated annealing interest; physically these systems act like glasses. Accordingly, keeping the lag distance constant is not equivalent to constant thermodynamic speed. A constant D schedule does share an attractive feature with constant ther-

Figure 1.1: A schematic distillation column with flows: feed F, distillate D and bottoms B. The close up shows two adjacent trays including overflow tubes for downward flow of liquid L and bubble caps for upward flow of vapor V.

modynamic speed annealing: both schedules measure energy and time on natural scales of the system. Their "optimality" thus follows from an old meta-theorem of applied mathematics: the more one exploits the structure of the problem the better.

There exist several other means of implementing constant thermodynamic speed. One popular technique for well studied problems which gets around the difficulty associated with adaptive algorithms is to model or fit the constant speed schedule obtained by laborious adaptive analysis and then use a rescaled version of this schedule for other similar problems [52].

1.6 Staged steady flow processes

Most recently [53], the connection between dissipation and geometry has been extended to treat a staged steady-flow process of considerable industrial interest: fractional distillation. The example involves some surprises which hint at the existance of other applications of horse-carrot type analyses. The first surprise is that the scale of the process is set by the flow rates rather than the states along the process. The second surprise is that the null directions for our semi-Riemannian metric turn out to be useful.

1.6.1 Dissipation in a Distillation Column

Fractional distillation is a process for separating a mixture of compounds based on the differences in the boiling points of the components. Fractional distillation is performed within a vertical column divided into trays which constitute the k stages for the process. The mixture to be separated is introduced near the middle of the column at the feed tray and the separated components are removed at the top as distillate D and at the bottom as bottoms B. (see figure 1). Boiling occurs on each tray resulting in the formation of vapor which is then bubbled through the liquid at the next higher tray. Similarly, each tray is equipped with an overflow tube which returns excess liquid to the next lower tray.

We treat a binary mixture at constant pressure[15]. For steady state

---

[15]More components and a pressure differential along the column can be handled similarly, albeit at a significant cost in complexity

operation, the net difference between the upward flow of vapor and the downward flow of liquid must equal D at each tray-tray interface above the feed and B at each interface below the feed. Formally, numbering the trays from the top of the column (see figure ???) and letting V_m and L_m stand for the number of moles of vapor and liquid leaving tray m, the mass balance equations at the interface between trays m and $m+1$ are

$$V_{m+1} - L_m = \begin{cases} D & \text{above feed} \\ -B & \text{below feed} \end{cases} \tag{1.95}$$

$$y_{m+1}V_{m+1} - x_m L_m = \begin{cases} x_D D & \text{above feed} \\ -x_B B & \text{below feed} \end{cases} \tag{1.96}$$

where x and y are the mole fractions of the first component in the gaseous and liquid phases respectively. In the limit of an infinite number of trays this becomes

$$V - L = \begin{cases} D & \text{above feed} \\ -B & \text{below feed} \end{cases} \tag{1.97}$$

$$yV - xL = \begin{cases} x_D D & \text{above feed} \\ -x_B B & \text{below feed} \end{cases} \tag{1.98}$$

with x, y, V and L now smooth functions of T except at the feed plate where we switch between the appropriate balance conditions. The whole process becomes a continuous, piecewise smooth path in the state space of the two-phase binary system by including a rescaling branch at the feed plate which contributes length zero. This path is known in the literature [54] as the minimum reflux values of V and L at each T. This is also the path we will dissect into k equal length pieces for a discrete horse-carrot process.

In the conventional operation of the column, a heat source is connected at the bottom tray and a heat sink is connected at the top tray creating a temperature gradient along the column. This results in the net upward motion of low boiling component and downward motion of high boiling component. We depart from the conventional design and use additional heat sources (sinks) along the column to adjust the temperature at each plate. We then ask for the sequence of temperatures which minimizes the

total dissipation inside the column. We take the transport of heat and matter between the column and its surroundings as reversible.

Since we assume that each stage is in equilibrium, the losses occur as the upward flow of vapor and downward flow of liquid equilibrate at the next trays. For concreteness, consider the bubble of vapor going up – the analysis for the downward flow of liquid proceeds similarly. The losses can be counted by a conceptual rearrangement of what occurs. We consider the bubble of vapor to be isolated except for the exchange of heat and pdV work with the two phase fluid in the tray above. Accordingly, this fluid acts as a bath with a certain temperature and pressure. In this manner, the bubble is brought to equilibrium at the temperature and pressure of the next tray by a horse-carrot process whose entropy production is given by the distance squared, $\Delta \ell^2$. This squared distance is an extensive quantity; the scale is set by the number of moles of material moving per unit time. In the final state of each bubble, some of the vapor has condensed to liquid, but each phase is exactly at the composition in the next tray and we can reversibly mix the bubble and its surroundings. Our conceptual rearrangement of events is justified since in either case the net effect is the complete equilibration between the bubble of vapor and the equilibrium system in the next tray.

Since we assume constant pressure, the form of the metric in eq. (1.55) is the most convenient since only one term in the sum is non-zero.

$$\Delta S_u = \frac{1}{2}(\Delta \ell)^2 = \frac{1}{2}\frac{\Delta T \Delta U}{T} = \frac{1}{2}\frac{C_\sigma (\Delta T)^2}{T} \qquad (1.99)$$

where C_σ is the constant pressure saturation heat capacity of the two phase mixture in equilibrium [55]. We get the same expression for the liquid, although ΔT has the opposite sign. Since the dissipation only depends on $(\Delta T)^2$, we would get the same entropy production if the liquid flow were reversed and also went up the column.

We have hereby established that the dissipation of small relaxation steps along this path equals the squared length of the corresponding displacement. Therefore, the discrete horse-carrot theorem applies and we can conclude that, to minimize total entropy production in the column, the tray temperatures should be adjusted to equalize the thermodynamic distance between trays. To find the optimal temperature profile, we need to find temperatures T_j such that

$$\int_{T_j}^{T_{j+1}} \sqrt{\frac{C_\sigma}{T}} dT = \frac{1}{k} \int_{T_0}^{T_k} \sqrt{\frac{C_\sigma}{T}} dT, \qquad j = 0, ..., k-1. \qquad (1.100)$$

This derivation shows an application of the discrete horse-carrot theorem to the steady state operation of a separation process. The results

express the dissipation in terms of the length of a path in the equilibrium state space of the mixture and show how to optimally control the temperatures of the stages along such a separation. The procedure is readily adapted to any staged steady flow process in the limit of many stages. We start from the flow vectors along the process. Since these flows equilibrate at the next stage, the entropy produced by such small relaxations is the square of a length element. For the purpose of counting dissipation, all flows can be taken unidirectional and summed exactly as for distillation. The corresponding path consists of the flows for the process in the limit of infinitely many stages. This should have implications for the control of many real processes.

1.7 Conclusions

This chapter presented a review of the geometry of distinguishability in all its guises ranging from the quantum to the macroscopic. We tried to present a thorough overview of the results and applications along with the connections to related geometries.

Bibliography

[1] F. Weinhold: "Metric geometry of thermodynamics I-IV", J. Chem. Phys. 63 (1975) 2479, 2484, 2488, 2496.

[2] G.Ruppeiner: "Thermodynamics: a Riemannian geometric model", Phys. Rev. A 20 (1979) 1608-1613.

[3] L.Diósi, G.Forgács, B.Lukács and H.L.Frisch: "Metricization of thermodynamic state space and the renormalization group", Phys. Rev. A 29 (1984) 3343-3345.

[4] R.A. Fisher: "????", Proc. R. Soc. Edinburgh 42 (1922) 321.

[5] S.L. Braunstein and C.M. Caves: "Statistical Distance and the geometry of quantum states", Phys. Rev. Lett. 72 (1994) 3439-3443.

[6] W.K.Wootters: "Statistical distance and Hilbert space", Phys. Rev. D 23 (1981) 357-362.

[7] G.Ruppeiner: "Riemannian geometry in thermodynamic fluctuation theory", Rev. Mod. Phys. 67 (1995) 605-659.

[8] P. Salamon and R.S. Berry: "Thermodynamic Length and Dissipated Availability", Phys. Rev. Lett. 51 (1983) 1127-1130.

[9] J. Nulton, P. Salamon, B. Andresen, and Qi Anmin: "Quasistatic Processes as Step Equilibrations", J. Chem. Phys. 83 (1985) 334.

[10] C.R.Rao: "Information and the accuracy attainable in the estimation of statistical parameters", Bull. Calcutta Math. Soc. 37 (1945) 81-91.

[11] A.Batthachariya: "On a measure of divergence between two multinomial populations", The Indian J. Stat. 7 (1946) 401-406.

[12] C.Atkinson and A.F.S.Mitchell: "Rao's distance measure", Indian J. Stat. 43 (1981) 345-365.

[13] D.J.C.Bures: "????", Trans. Am. Math. Soc. 135 (1969) 199.

[14] R.A. Fisher: "On the Mathematical Foundations of Theoretical Statistics" Phil. Trans. Roy. Soc. A 222 (1921) 309-368.

[15] S.Amari: "Differential-Geometric Methods in Statistics", Lecture Notes in Statistics Vol.28, Springer (1985).

[16] N.N.Chentzov: "Categories of mathematical statistics", (in Russian) Dokl. Akad. Nauk. SSSR 164 (1965) 3.

[17] F. Schlögl: "Thermodynamic Metric and Stochastic Measures", Z. Phys. B 59 (1985) 449.

[18] P. Salamon, J. Nulton, and E. Ihrig: "On the Relation between Energy and Entropy Versions of Thermodynamic Length", J. Chem. Phys. 80 (1984) 436.

[19] P. Salamon, J.D. Nulton, and R.S. Berry: "Length in Statistical Thermodynamics", J. Chem. Phys. 82 (1985) 2433-2436.

[20] F. Weinhold: "Metric geometry of equilibrium thermodynamics V.", J. Chem. Phys. 65 (1976) 559. See in particular figure 1.

[21] H. Callen: Thermodynamics, J. Wiley & Sons, New York (1985).

[22] P. Salamon, B. Andresen, P. Gait, and R.S. Berry: "Interpretation of Weinhold's Metric", J. Chem. Phys. 73 (1980) 1001.

[23] G.Ruppeiner: "Thermodynamic critical fluctuation theory?", Phys. Rev. Lett. 50 (1983) 287.

[24] G.Ruppeiner: "New thermodynamic fluctuation theory using path integrals", Phys. Rev. A 27 (1983) 1116.

[25] L.Diósi and B.Lukács: "Covariant evolution equation for the thermodynamic fluctuations", Phys. Rev. A 31 (1985) 3415.

[26] A. Bejan: Entropy generation through heat and fluid flow, J. Wiley, New York (1982).

[27] R. Hermann: Geometry, Physics and Systems, Marcel Dekker, New York (1981).

[28] R. Mrugala, J. D. Nulton, J. C. Schn, and P. Salamon: "Contact Transformations in Thermodynamics" Reports on Mathematical Physics 29 (1991) 109-121.

[29] R. Mrugala, J. D. Nulton, J. C. Schn, and P. Salamon: "A Statistical Approach to the Geometric Structure of Thermodynamics", Phys. Rev. A 41 (1990) 3156.

[30] P. Salamon, E. Ihrig, and R.S. Berry: "A Group of Coordinate Transformations Preserving the Metric of Weinhold" J. Math. Phys 24 (1983) 2515.

[31] B.C. Eu: "Note on the nonequilibrium partition function and generalized potentials", J. Chem. Phys. 105 (1996) 5525.

[32] K.H. Hoffmann, B. Andresen, and P. Salamon: "Measures of Dissipation", Phys. Rev. A 40 (1989) 3618-3630.

[33] J. Dieudonne: Linear algebra and geometry, Houghton Mifflin, Boston (1969).

[34] T. Feldmann, B. Andresen, Anmin Qi, and P. Salamon: "Thermodynamic Lengths and Intrinsic Time Scales in Molecular Relaxation", J. Chem. Phys. 83 (1985) 5849-5853.

[35] S.R. deGroot and P. Mazur: Non-Equilibrium Thermodynamics, North Holland, Amsterdam (1962).

[36] J. Milnor: Morse Theory, Princeton University Press, Princeton (1969) p. 70-73.

[37] L. Diosi, K. Kulacsy, B. Lukacs, and A. Racz: "Thermodynamic length, time, speed, and optimum path to minimize entropy production", J. Chem. Phys. 105 (1996) 11220-11225.

[38] W. Spirkl and H. Ries, "Optimal finite time endoreversible processes", Phys. Rev. E 52 (1995) 3485-3489.

[39] K. Olah: "The entropy production: new results", Magyar kemiai folyoirat 103 (1997) 411.

[40] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller: "Equation of State Calculations by Fast Computing Machines", J. Chem. Phys. 21 (1953) 1087.

[41] P. Salamon, J. Nulton, J. Robinson, J. Pedersen, G. Ruppeiner, and L. Liao: "Simulated Annealing with Constant Thermodynamic Speed", Comp. Phys. Comm. 49 (1988) 423.

[42] K.H. Hoffmann and P. Salamon: "The Optimal Simulated Annealing Schedule for a Simple Model", J. Phys. A 23 (1990) 3511.

[43] R. Azencott: Simulated annealing : parallelization techniques, J. Wiley, New York (1992).

[44] C. Sechen and A. Sangiovanni-Vincentelli: "The TimberWolf Placement and Routing Package", IEEE Custom Integrated Circuits Conference (1984).

[45] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli: "Convergence and finite-time behavior of simulated annealing", Adv. Appl. Prob. 18 (1986) 747.

[46] T. Zimmermann and P. Salamon, "The Demon Algorithm", Int. J. Comp. Math. 42 (1992) 21.

[47] P. Salamon, K.H. Hoffmann, J.R. Harland, and J.D. Nulton: "An Information Theoretic Bound on the Performance of Simulated Annealing Algorithms", Research Report IRC.89.2 (1989) Interdisciplinary Research Center, San Diego State University.

[48] Y. Nourani and B. Andresen: "Simulated annealing with optimal cooling strategy and the natural time scale", to appear, J. Phys. A.

[49] K. Mosegaard and P.D. Vestergaard: "A simulated annealing approach to seismic model optimization with sparse prior information", Geophysical prospecting 39 (1991) 599.

[50] J. Nulton and P. Salamon: "Statistical Mechanics of Combinatorial Optimization", Phys. Rev. A 37 (1988) 1351-1356.

[51] R.K. Pathria: Statistical Mechanics, Butterworth-Heinemann, Oxford (1996).

[52] B. Andresen and J.M. Gordon: "Constant Thermodynamic Speed Simulated Annealing", Lecture notes in earth sciences 63 (1996) 303.

[53] P. Salamon and J.D. Nulton: The Geometry of Separation Processes: The Horse-Carrot Theorem for Steady Flow Systems, to appear, Europhys. Lett..

[54] C.J. King: Separation Processes, McGraw-Hill, New York (1971).

[55] J.S. Rowlinson: Liquids and Liquid Mixtures, Plenum Press, New York (1969).

[56] A. de Vos: "Some examples of thermodynamic processes in finite time", Berlin colloquium on finite-time thermodynamics, 13-15 Nov. (1997).