数理解析研究所講究録
980 巻 1997 年 56-65

Detailed Balance and Free Energy

R.F.Streater,
Dept. of Mathematics,
King's College London,
Strand WC2R 2LS

October 27, 1993

Abstract

We formulate the notion of detailed balance at a given beta for a discrete Markov chain, and show that it implies that one time-step in the chain is a contraction in the norm $\|\bullet\|_\beta$; this is defined on the set of states as the dual to the norm on the random variables given by the classical KMS state.

Key words: Free energy; relative entropy

1 Introduction and Notation

Suppose that Ω is a countable space, with points labelled by integers $i, j \ldots$. Denote by $\mathcal{A}$ the (abelian) algebra of bounded real-valued random variables f on Ω with sup norm, and by $\mathcal{A}^*$ the dual space (of continuous linear maps, $\mathcal{A} \rightarrow \mathbf{R}$). The positive, normalised elements of $\mathcal{A}^*$ are the probability measures on Ω, also called the states on $\mathcal{A}$. A typical state will be denoted by p; they form the convex set $\Sigma(\Omega)$. In fact, $\Sigma(\Omega)$ is a simplex, whose extreme points, the corners, are the Dirac measures on the points of Ω, thus for any subset $\Omega_0 \subseteq \Omega$,

$$\begin{aligned} \delta_j(\Omega_0) &= 1 \qquad \text{if } j \in \Omega_0 && (1)\\ &= 0 \qquad \text{otherwise} && (2) \end{aligned}$$

Any probability measure $p \in \Sigma$ can be written uniquely as a sum

$$p = \sum_j p_j \delta_j \qquad \text{where } p_j \geq 0, \ \sum_j p_j = 1. \tag{3}$$

The p_j will be called the components of p. The duality between $\mathcal{A}$ and $\mathcal{A}^*$ reduces on Σ to the expectation:

$$\langle p, f\rangle = \sum_{j\in\Omega} p_j f_j \qquad p\in\Sigma,\ f\in\mathcal{A}. \tag{4}$$

Let $T:\mathcal{A}\to\mathcal{A}$ be a linear map. Its adjoint, T^*, maps $\mathcal{A}^*$ to $\mathcal{A}^*$ and is defined by

$$\langle T^*p, f\rangle = \langle p, Tf\rangle \qquad p\in\mathcal{A}^*,\ f\in\mathcal{A}. \tag{5}$$

In order for T^* to map Σ to itself it is necessary and sufficient that T be a stochastic map, that is

1. If $f\geq 0$, then $Tf\geq 0$
2. $T1=1$.

The components of an element of $\mathcal{A}^*$ define a bounded function on Ω, and so give us an element of $\mathcal{A}$. The components of the Dirac measure δ_i define a function that we shall denote by Δ_i. Then δ_i,Δ_j form a dual basis as we vary i and j: $\langle\delta_i,\Delta_j\rangle=\delta_{ij}$. If T is a stochastic map, then T^* is determined by its matrix elements

$$T^*_{ij} = \langle T^*\delta_j,\Delta_i\rangle = \langle\delta_j, T\Delta_i\rangle. \tag{6}$$

Clearly, $T^*_{ij}\geq 0$. Also

$$\begin{aligned}\sum_{i\in\Omega} T^*_{ij} &= \langle T^*\delta_j,\sum_i\Delta_i\rangle = \langle T^*\delta_j, 1\rangle &(7)\\ &= \langle\delta_j, T1\rangle = \langle\delta_j,1\rangle = 1 &(8)\end{aligned}$$

for all j. Thus T^*_{ij} is a Markov matrix in the usual sense.

Suppose now that $\mathcal{E}:\Omega\to\mathbf{R}$ is a positive random variable, not necessarily bounded, interpreted as the energy. We shall assume that $e^{-\beta\mathcal{E}}$ is summable, so that the partition function

$$Z_\beta = \sum_\omega e^{-\beta\mathcal{E}(\omega)} \tag{9}$$

is finite. A Markov matrix T^*_{ij}, $i,j\in\Omega$, is usually said to obey "detailed balance" at beta β if

$$T^*_{ij} = e^{-\beta(\mathcal{E}_i-\mathcal{E}_j)}T^*_{ji} \qquad (i,j\in\Omega). \tag{10}$$

If $\beta = 0$ Eqn.(10) says that T^* is symmetric. For some applications of Markov chains such as neural nets there is no good reason for T^* to be symmetric, even for $\beta = 0$. In this paper we suggest a generalisation of (10) which avoids the symmetry, but which nevertheless enables us to derive thermodynamics. We show that the Gibbs state is a fixed point of the chain if and only if generalised detailed balance holds. We use the "dual KMS norm" coming from the Gibbs state and show that the Gibbs state is the state of minimum norm, and that T^* is a contraction. We obtain an estimate for the loss in free energy at each time-step in terms of this norm. Finally, a short proof of the second law of thermodynamics for a quantum Markov chain is presented.

2 The Detailed Balance Condition

Let $\mathcal{A}_c$ be the subalgebra of functions on Ω zero except at a finite number of points, and for $f, g \in \mathcal{A}$ define the scalar product

$$\langle f, g\rangle_\beta = Z_\beta^{-1} \sum_{i\in\Omega} e^{-\beta\mathcal{E}_i} f_i g_i. \tag{11}$$

This idea is borrowed from the theory of quantum statistical mechanics, [1]. In this paper, the importance of the analyticity of the Green's functions found by Kubo, and by Martin and Schwinger, is emphasized. So we call the corresponding scale of norms

$$\|f\|_\beta = \left(Z_\beta^{-1} \sum_j e^{-\beta\mathcal{E}_j} |f(j)|^2 \right)^{1/2} \tag{12}$$

the KMS norms. Let $\mathcal{A}_\beta$ be the completion of $\mathcal{A}_c$ in the metric given by $\|\bullet\|$. The dual spaces of these Banach spaces are the completions of $\mathrm{Span}\Sigma(\Omega)_c$, of real measures of compact support, in the topology defined by the dual norms

$$\|p\|_{-\beta} = \left(Z_\beta \sum_j e^{\beta\mathcal{E}_j} |p_j|^2 \right)^{1/2} . \tag{13}$$

We shall see that these are the natural norms governing the convergence of states under isothermal conditions, rather than the L^1-norms usually used for measures. Convergence in the KMS-norm $\| \bullet \|_{-\beta}$ with $\beta > 0$ is a much stronger notion than L^1 convergence since the rising exponential $e^{\beta\mathcal{E}_j}$ forces the high-energy tail of the probability to converge very fast, and captures some of the information in the theory of large deviations.

Since we assume that Z_β is finite, the unit function lies in $\mathcal{A}_\beta$. Let T be a bounded operator on $\mathcal{A}_\beta$, and denote by $T^{(\beta)}$ its adjoint relative to $\langle \bullet, \bullet \rangle$. Note that $T^{(\beta)}$ acts on $\mathcal{A}_\beta$, which is a Hilbert space.

Definition 1 *We say that $T : \mathcal{A}_\beta \to \mathcal{A}_\beta$ obeys* detailed balance *(relative to $\mathcal{E}$) if T and $T^{(\beta)}$ are both stochastic maps.*

In particular, if T satisfies detailed balance, then $T1 = T^{(\beta)}1 = 1$. We note that the Gibbs measure p_β lies comfortably in $\mathcal{A}^*_\beta$. We have an easy lemma:

Lemma 2 *A stochastic matrix T obeys detailed balance if and only if p_β is a fixed point of T^*.*

For the proof, note that if $T^* p_\beta = p_\beta$, then we have

$$\begin{aligned}\langle 1, f\rangle_\beta &= \langle p_\beta, f\rangle = \langle T^* p_\beta, f\rangle \\ &= \langle p_\beta, Tf\rangle = \langle 1, Tf\rangle_\beta = \langle T^{(\beta)}1, f\rangle_\beta\end{aligned}$$

for all $f \in \mathcal{A}_c$. Hence, as $\mathcal{A}_c$ is dense in $\mathcal{A}_\beta$, we get $T^{(\beta)}1 = 1$ as required. The converse is obtained by working backwards.

So our formulation of detailed balance is the most general that could lead to the convergence of the Markov chain $\{(T^*)^n p\}_{n=0,1\ldots}$ to the thermal state p_β. At $\beta = 0$, the definition (1) reduces to the condition that T is bistochastic. This is enough to ensure that entropy is increased by T^*. At $\beta > 0$, we shall get the corresponding result, that T^* reduces the free energy.

Theorem 3 *The unique probability measure with the smallest norm $\|p\|_\beta$ is the Gibbs state p_β.*

This is proved by using a Lagrange multiplier $\lambda \sum_j p_j$ to fix the L^1 norm.

We now come to the main result of this section.

Theorem 4 *If T obeys detailed balance, then T^* is a contraction in $\|\bullet\|_{-\beta}$. If in addition T is irreducible, then p_β is the only fixed point of T^* acting on $\mathcal{A}^*_\beta$ and if $p \neq p_\beta$ then $\|T^* p\|_{-\beta} < \|p\|_{-\beta}$.*

Note: T irreducible means that for any $i, j \in \Omega$ there is a chain of finite length $i_1 = i$, $i_2 i_3 \ldots i_N = j$ such that $T^*_{i_k, i_{k+1}} > 0, \quad k = 1, \ldots N$.

Proof.
Obviously for any sum over a finite subset $\Omega_0 \subseteq \Omega$ we have for any $p, q \in \Sigma(\Omega)$:

$$\sum_{i,j} Z_\beta T^*_{ji} e^{-\beta\mathcal{E}_i} \left(e^{\beta\mathcal{E}_j} q_j - e^{\beta\mathcal{E}_i} p_i\right)^2 \geq 0 \tag{14}$$

and this inequality persists in the limit as Ω_0 increases to Ω if the terms converge. Let $q = T^* p$. The first term is the sum of non-negative terms

$$Z_\beta \sum_{i,j} T^*_{ji} e^{-\beta(\mathcal{E}_i - \mathcal{E}_j)} e^{\beta\mathcal{E}_j} q_j^2 = Z_\beta \sum_{i,j} T^{(\beta)}_{ji} e^{\beta\mathcal{E}_j} q_j^2.$$

By detailed balance the sum of $T^{(\beta)}_{ji}$ over i converges to 1 as Ω_0 increases to Ω. So the first term reduces to

$$Z_\beta \sum_j e^{\beta\mathcal{E}_j} q_j^2 = ||T^* p||^2_{-\beta}.$$

The cross-terms are also of one sign,

$$-2Z_\beta \sum_{ij} T^*_{ji} e^{\beta\mathcal{E}_j} q_j p_i$$

and the sum over i is

$$\sum_{i\in\Omega} T^*_{ji} p_i = q_j = (T^* p)_j.$$

The cross-terms thus give us

$$-2||T^* p||^2_{-\beta}.$$

The last term is exactly $||p||^2_{-\beta}$, since T^* is stochastic. We conclude that

$$-||T^* p||^2_{-\beta} + ||p||^2_{-\beta} \geq 0$$

as desired. For the strict inequality we follow [3]. We get a proper contraction unless

$$e^{\beta\mathcal{E}_j} q_j = e^{\beta\mathcal{E}_i} p_i \text{ for all } i,\, j \text{ with } T^*_{ji} > 0.$$

Starting with $i = 0$ and taking $\mathcal{E}_0 = 0$ we get

$$q_j = e^{-\beta\mathcal{E}_j} p_0$$

for all j linked to 0 by one step. But then we get

$$q_{j'} = e^{-\beta(\mathcal{E}_{j'}-\mathcal{E}_j)}e^{-\beta\mathcal{E}_j}p_0$$

for all j' linked to j by one step, so

$$q_{j'} = e^{-\beta\mathcal{E}_{j'}}p_0$$

for all j' linked to 0 by two steps. In this way we get

$$q_k = e^{-\beta\mathcal{E}_k}p_0 \text{ for all } k.$$

Similarly we get

$$p_k = e^{-\beta\mathcal{E}_k}q_0 \text{ for all } k$$

so both p and q are p_β.

3 The Free-energy Theorem

Definition 5 *Let us say that a state $p \in \Sigma(\Omega)$ is* close to equilibrium (C,β) *if there exists $C \geq 1$ such that*

$$p(j) \leq CZ_\beta^{-1}e^{-\beta\mathcal{E}_j} \textit{ for all } j \in \Omega.$$

This includes all Gibbs states of beta greater than β, all states with a maximum energy, and all states of the form fp_β with $f \in \mathcal{A}$.

Lemma 6 *If p is close to equilibrium (C,β) and T obeys detailed balance (β), then T^*p is close to equilibrium (C,β).*

For, T^* being positivity preserving,

$$T^*\left(CZ_\beta^{-1}e^{-\beta\mathcal{E}} - p\right)(j) \geq 0.$$

Since T^* leaves p_β fixed, we get

$$CZ_\beta^{-1}e^{-\beta\mathcal{E}_j} - (T^*p)_j \geq 0$$

which proves the lemma. The next theorem relates the relative entropy to the KMS-norm, and is proved in [2].

Theorem 7 *If p is close to equilibrium (C,β) then the relative entropy obeys*

$$S\left(p_\beta|p\right) = \sum_j p_\beta(j)(\log p_\beta(j) - \log p_j) \geq (2C^2)^{-1}\|p - p_\beta\|_{-\beta}^2. \qquad (15)$$

This says that for a sequence of states close to equilibrium, if the relative entropy converges to zero, then the states converge to equilibrium in $\|\bullet\|_{-\beta}$. We now give an estimate for the decrease in free energy under a map obeying detailed balance, following [3] where a special case is proved.

Theorem 8 *Suppose that T obeys detailed balance (β) and let*

$$F(p) = \beta\langle p, \mathcal{E}\rangle - S(p)$$

denote the free energy, defined on a subset of Ω. Then

$$F(T^*p) \leq F(p).$$

If further p is close to equilibrium (C, β), then

$$F(T^*p) - F(p) \leq -(2C)^{-1}\left(\|p\|_{-\beta}^2 - \|T^*p\|_{-\beta}^2\right).$$

Our proof uses the

Lemma 9 *Let $s(x) = -x\log x$ and $x = \sum \lambda_j x_j$, $\lambda_j \geq 0, \sum \lambda_j = 1$. Then if $x \neq x_j, j = 1 \ldots n$ we have*

$$\sum \lambda_j s(x_j) = s(x) - \frac{1}{2}\sum \lambda_j (x_j - x)^2/\xi_j$$

where $x_j < \xi_j < x$ or $x < \xi_j < x_j$.

Proof.
By Taylor's theorem with Lagrange remainder,

$$s(x_i) = s(x) + (x_i - x)[-1 - \log x] + (x_i - x)^2(-1/(2\xi_i).$$

Multiply by λ_i and sum; we get the lemma, since $\sum \lambda_i(x_i - x) = 0$.

Try this lemma with

$$\lambda_i = T_{ji}^* e^{-\beta(\mathcal{E}_i - \mathcal{E}_j)} = T_{ji}^{(\beta)}$$

for each j, and with $x_i = e^{\beta\mathcal{E}_i}$. Then

$$x = \sum_i T_{ji}^* e^{-\beta\mathcal{E}_i + \beta\mathcal{E}_j} e^{\beta\mathcal{E}_i} p_i = e^{\beta\mathcal{E}_j} q_j$$

where $q = T^*p$. So for each j

$$
\begin{aligned}
-\sum_i T^*_{ji} e^{-\beta(\mathcal{E}_i-\mathcal{E}_j)} \left(e^{\beta\mathcal{E}_i} p_i(\log p_i + \beta\mathcal{E}_i)\right) &= -e^{-\beta\mathcal{E}_j} q_j(\log q_j + \beta\mathcal{E}_j) \\
&\quad - \sum_i T^*_{ji} e^{-\beta(\mathcal{E}_i-\mathcal{E}_j)} \left(e^{\beta\mathcal{E}_i} p_i - e^{\beta\mathcal{E}_j} q_j\right)^2 /(2\xi_{ij}).
\end{aligned}
$$

Cancel $e^{\beta\mathcal{E}_j}$ to get

$$
\begin{aligned}
\sum_i T^*_{ji}(-p_i \log p_i - \beta p_i \mathcal{E}_i) &= -q_j \log q_j - \beta q_j \mathcal{E}_j \\
&- \sum_i T^*_{ji} e^{-\beta\mathcal{E}_i} \left(e^{\beta\mathcal{E}_i} p_i - e^{\beta\mathcal{E}_j} q_j\right)^2 /(2\xi_{ij}).
\end{aligned}
$$

Now either

$$
e^{\beta\mathcal{E}_i} p_i < \xi_{ij} < e^{\beta\mathcal{E}_j} q_j < Z_\beta^{-1} C
$$

or

$$
e^{\beta\mathcal{E}_j} q_j < \xi_{ij} < e^{\beta\mathcal{E}_i} p_i < Z_\beta^{-1} C.
$$

In either case, $1/\xi_{ij} > Z_\beta C^{-1}$. Summing over j we get

$$
\begin{aligned}
S(p) - \beta\langle p, \mathcal{E}\rangle &\le S(q) - \beta\langle q, \mathcal{E}\rangle \\
&- \frac{Z_\beta}{2C} \sum_{ij} T^*_{ji} e^{-\beta\mathcal{E}_i} \left(e^{\beta\mathcal{E}_i} p_i - e^{\beta\mathcal{E}_j} q_j\right)^2,
\end{aligned}
$$

which gives the result, in view of the proof of theorem(4).

It follows that if p is close to equilibrium and T is irreducible, then $T^{*n}p \to p_\beta$ as $n \to \infty$ in KMS-norm. Our theorem is the classical, discrete-time version of the deep result of Majewski [4], who shows that a continuous quantum stochastic semigroup on a C^*-algebra, obeying a suitable detailed balance condition, is a contraction in the Gelfand-Naimark-Segal Hilbert space given by the KMS state. It is remarkable that he does not need to assume complete positivity. Majewski's version of "detailed balance" expresses microscopic reversibility in quantum mechanics, and is similar to the formulation of Agarwal [5]. A useful account of quantum dynanimical semigroups can be found in [6].

4 An Easy Result on Quantum Entropy

There is a parallel theory in quantum mechanics, and our notation has been designed to bring out the similarities. The observables form a C^*-algebra

$\mathcal{A}$ and the states, the positive normalised linear maps $\rho : \mathcal{A} \to \mathbf{R}$, make up a convex subset Σ of the dual space $\mathcal{A}^*$. A stochastis map is a linear map $T : \mathcal{A} \to \mathcal{A}$ which takes positive operators to positive operators, and takes 1 to 1. It then follows that T^* takes Σ to Σ. We say that T is bistochastic if it is stochastic and trace-preserving, which makes sense on the class of trace-class operators. For a state ρ in a suitable domain we can define the quantum entropy of a ρ by $S(\rho) = -\mathrm{Tr}(\rho \log \rho)$.

I. Amah [7] has noticed that we can reduce some questions in quantum statistical dynamics to the corresponding classical question.

Theorem 10 *Let T be a bistochastic map on a represented C^*-algebra. Then for any normal state ρ we have*

$$S(T^*\rho) \geq S(\rho).$$

Proof.
The states ρ and $\sigma = T^*\rho$ are of trace class; let $\rho = \sum \rho_i P_i$ and $\sigma = \sum_j \sigma_j Q_j$ be their spectral resolutions. We take it that P_i and Q_j are one-dimensional projections, and that multiplicity is accounted for by repetition. Since both are self-adjoint, the P_i can be chosen mutually orthogonal. The same goes for the Q_j. Define the infinite matrix

$$T_{ij} = \mathrm{Tr}\left(Q_i(TP_j)\right). \tag{16}$$

The point is that this is a Markov matrix: $\sum_j T_{ij} = \mathrm{Tr}\left(Q_i(T\sum_j P_j)\right) = \mathrm{Tr}(Q_i T1) = \mathrm{Tr}(Q_i) = 1$. Its elements are non-negative, since TP_j is a positive operator, and Q_i is a density matrix. In the same way, we show that if T is bistochastic in the quantum sense, then $[T_{ij}]$ is bistochastic. Moreover, it is immediate that $\sigma_i = \sum_j T_{ij}\rho_j$. Therefore the entropies of ρ and σ obey the inequality $S(T^*\rho) \geq S(\rho)$ by appealing to the classical result. The gain in quantum entropy can also be estimated using a classical estimate.

5 Conclusions

We have advocated the use of the dual KMS-norms for studying the convergence of Markov chains obeying detailed balance. Such an operator is shown to be a contraction in this norm. The same proof works for a generalisation of the notion of detailed balance, which reduces to bistochasticity when beta is zero. The concept of a state's being close to equilibrium is formulated, and it is shown that this condition is stable under one time-step. The definition gives a restriction on the class of initial states for which convergence of the chain holds in the KMS-norm. The condition always holds if Ω is finite, and for all states in the same representation space as the Gibbs state in question.

References

[1] R. Haag, N.M. Hugenholtz and M. Winnick, On the equilibrium states in quantum statistical mechanics, Commun. Math. Phys., **5**, 215-236 (1967). See G.L. Sewell, *Quantum Theory of Collective Phenomena*, Clarendon Press, Oxford, 1986.

[2] R.F. Streater, Entropy and KMS-Norms; Proceedings of the 1993 Madeira Advanced Study Institute "Stochastic Analysis and Applications to Physics", (L.Streit, Editor), Kluwer, 1994.

[3] S.Koseki, Thesis, Ph.D., King's College London University, 1993.

[4] W.A. Majewski, Dynamical Semigroups in the Algebraic Formulation of Statistical Mechanics, Fortschritte der Physik, **25**, 89-133, (1984). See also W.A. Majewski, Journ. of Statistical Physics, **55** 417 (1989).

[5] G.S. Agarwal, Open Quantum Markovian Systems and Microreversibility; Zeits. fur Physik **258**,409-422, 1973.

[6] R. Alicki and K. Lendl, *Quantum Dynamical Semigroups and Applications.* Lecture Notes in Physics, **286**, 1987.

[7] I. Amah, M.Sc. Thesis, Imperial College, London University, (1993).