NAMT
96-011

FREE ENERGY AND THE FOKKER-PLANCK EQUATION

Richard Jordan
David Kinderlehrer
Carnegie Mellon University

Felix Otto
Carnegie Mellon University and
University of Bonn

Research Report No. 96-NA-011

July 1996

Sponsors

U.S. Army Research Office
Research Triangle Park
NC 27709

National Science Foundation
1800 G Street, N.W.
Washington, DC 20550

Submitted to Physica D

Free Energy and the Fokker–Planck Equation

Richard Jordan*

David Kinderlehrer†

Center for Nonlinear Analysis

Carnegie Mellon University

Felix Otto‡

Center for Nonlinear Analysis

Carnegie Mellon University

and

Department of Applied Mathematics

University of Bonn

June 10, 1996

*rjordan@andrew.cmu.edu

†dk3L@andrew.cmu.edu

‡otto@andrew.cmu.edu

Abstract

We establish a new and intriguing connection between the Fokker–Planck equation with gradient drift term and an associated free energy functional. Namely, we demonstrate that such a Fokker–Planck equation may be interpreted as a gradient flux, or a steepest descent, of a free energy functional with respect to a certain metric. This is accomplished through the construction of a time–discrete iterative variational scheme whose solutions converge to the solution of the Fokker–Planck equation. The time step in this scheme is governed by the Wasserstein metric on probability measures.

Keywords: Fokker–Planck equation, gradient flux, free energy, Wasserstein metric.

PACS numbers: 02.30.Jr, 02.30.Wd, 02.50.Ey, 05.40.+j.

1 Introduction and overview

Many important questions in statistical mechanics and in the theory of stochastic processes may be cast in terms of a free energy functional of the form $F = E + \beta^{-1}S$, where E is the potential energy, $-S$ is the entropy, and β plays the role of the inverse temperature. For a system with a finite or a countable number of states $x_1, x_2, \cdots$, the familiar expression for the free energy is

$$F = \sum \Psi(x_i)p_i + \beta^{-1} \sum p_i \log p_i \,,$$

where Ψ is a potential and p_i is the probability of state x_i. The form of the entropy $-S = -\sum p_i \log p_i$ arises from the usual Gibbs–Boltzmann statistical description. For a system with continuous state–space, say $\boldsymbol{R}^n$, the free energy

is given by the obvious analogous expression

$$F(\rho) = \int_{R^n} \Psi \rho \, dx + \beta^{-1} \int_{R^n} \rho \log \rho \, dx \,, \tag{1}$$

where ρ is a probability density on $\boldsymbol{R}^n$,

$$E(\rho) := \int_{R^n} \Psi \rho \, dx \,, \tag{2}$$

is the energy functional, and

$$S(\rho) := \int_{R^n} \rho \log \rho \, dx \,, \tag{3}$$

is the negative of the Gibbs–Boltzmann entropy functional. The potential Ψ is, in general, a nonconvex function.

In an attempt to go beyond the present framework of the calculus of variations, we have initiated an investigation [1] on functionals of the form (1), and on generalized free energy functionals of the form

$$\mathcal{F}(\nu) = \int \Psi \, d\nu + \beta^{-1} \int \log \frac{d\nu}{d\mu} \, d\nu \,. \tag{4}$$

Here, ν is a probability measure, μ is a fixed reference measure (in the case of (1), μ is just Lebesgue measure), and $-\int \log \frac{d\nu}{d\mu} \, d\nu$ is the Kullback relative entropy of ν with respect to μ [2]. We recognize in (4) the Young measure formulation of a variational principle [3] coupled to an entropic stabilization principle. The extrema of (4) deliver in the limit as $\beta^{-1} \to 0$ the relaxation, or the Young measure distribution, of a classical but nonconvex variational principal. Put another way, the integral $\int \Psi \, d\nu$ represents the contemporary approach to the calculus of variations, and in our interpretation, it has become ineluctably wedded to an entropy functional. In general, there is an intriguing connection between variational functionals of type (1) and (4) and simulated annealing procedures for global optimization [4].

Let us now focus our attention on the classical free energy (1). The minimizer of this functional is the Gibbs density

$$\rho_s(x) = Z^{-1} \exp(-\beta\Psi(x)) \,, \quad \text{with} \quad Z = \int_{R^n} \exp(-\beta\Psi(x))\, dx. \tag{5}$$

Note that, in order for equation (5) to make sense, Ψ must grow rapidly enough to ensure that Z is finite.

It is known [5, 6] that the Gibbs density is the stationary solution of the Fokker–Planck equation

$$\frac{\partial \rho}{\partial t} = \operatorname{div}(\nabla\Psi(x)\rho) + \beta^{-1}\Delta\rho \,. \tag{6}$$

The Fokker–Planck equation (6) plays a prominent role in statistical physics and in the study of fluctuations in physical, chemical, and biological systems [5, 6, 7]. Its importance in the study of energy landscapes in physical and biological systems is evidenced by its conspicuous appearance in many presentations at this particular conference. It is inherently connected with the theory of stochastic differential equations: A (normalized) solution to the Fokker–Planck equation provides the probability density for a stochastic process whose dynamics is governed by the Ito stochastic differential equation

$$dX(t) = -\nabla\Psi(X(t))dt + \sqrt{2\beta^{-1}}\, dW(t) \,.$$

Here, $W(t)$ is a standard n–dimensional Brownian motion. The Gibbs measure $\rho_s(x)\, dx$ is the unique invariant measure for the Markov process $X(t)$ [5, 6]. This illustrates that a variational principle involving a free energy functional of the form (1) (or more generally of the form (4)) may be associated with a stochastic process.

It may be shown [6] that the free–energy is an H–function for the Fokker–Planck dynamics: If the probability density $\rho(t, x)$ is a solution of (6), then

$F(\rho(t,x))$ is a decreasing function of time. Our main purpose here is to make prominent a more compelling, and previously unrecognized, connection between the Fokker–Planck dynamics and the free energy functional. Specifically, we shall demonstrate that the Fokker–Planck dynamics may be interpreted as a gradient flux, or a steepest descent, of the free energy functional. Technically, the notion of a gradient flux or a steepest descent makes sense only in the presence of an appropriate metric. The requisite metric in the case of the Fokker–Planck equation is the Wasserstein metric (to be defined below) on the set of probability measures having finite second moments. To make precise the sense in which the Fokker–Planck equation may be regarded as a steepest descent, or gradient flux, of the free energy functional with respect to this metric, we switch to a discrete time formulation. Namely, we develop a discrete, iterative variational scheme whose solutions converge, in a sense to be made exact below, to the solution of the Fokker–Planck equation. The time-step in this iterative scheme is associated with the Wasserstein metric.

To illustrate these ideas, let us consider the classical diffusion equation

$$\frac{\partial \rho(t,x)}{\partial t} = \Delta \rho(t,x)\,, \quad t \in (0,\infty)\,, \quad x \in \boldsymbol{R}^n\,,$$

which is the Fokker–Planck equation associated with a standard n–dimensional Brownian motion. It is well–known (see, for example, [8, 9]) that this equation is the gradient flux of the Dirichlet integral $\frac{1}{2}\int_{R^n} |\nabla \rho|^2\, dx$ with respect to the $L^2(\boldsymbol{R}^n)$ metric. The classical discretization is given by the scheme

$$\left.\begin{array}{l} \text{Determine } \rho^{(k)} \text{ that minimizes} \\ \frac{1}{2}\,\|\rho^{k-1} - \rho\|^2_{L^2(R^n)} \;+\; \frac{h}{2}\displaystyle\int_{R^n} |\nabla \rho|^2\, dx\,, \end{array}\right\}$$

over an appropriate class of densities ρ. Here, h is the time step size. On the

other hand, it is a special case of our results below that the scheme

$$\left.\begin{array}{l} \text{Determine } \rho^{(k)} \text{ that minimizes} \\ \quad \frac{1}{2}\, d(\rho^{k-1}, \rho)^2 \;+\; h \displaystyle\int_{R^n} \rho \log \rho \; dx \\ \text{over all } \rho \in K\,, \end{array}\right\}$$

is also a discretization of the diffusion equation when d is the Wasserstein metric. Here, K is the set of all probability densities on $\boldsymbol{R}^n$ having finite second moments. In particular, this allows us to regard the diffusion equation as a steepest descent of the functional $\int_{R^n} \rho \log \rho \; dx$ with respect to the Wasserstein metric. This reveals a novel link between the diffusion equation and the Gibbs–Boltzmann entropy of the density ρ. Furthermore, this formulation allows us to attach a precise interpretation to the conventional notion that diffusion arises from the tendency of the system to maximize entropy.

We shall now define the Wasserstein metric and briefly review its properties and interpretations.

2 The Wasserstein metric

The Wasserstein distance, $d(\mu_1, \mu_2)$, between two probability measures μ_1 and μ_2 on $\boldsymbol{R}^n$ is defined by the formula

$$d(\mu_1, \mu_2)^2 = \inf_{p \in \mathcal{P}(\mu_1, \mu_2)} \int_{R^n \times R^n} |x - y|^2 \; p(dxdy), \tag{7}$$

where $\mathcal{P}(\mu_1, \mu_2)$ is the set of all probability measures on $\boldsymbol{R}^n \times \boldsymbol{R}^n$ with first marginal μ_1 and second marginal μ_2. More precisely, a probability measure p is in $\mathcal{P}(\mu_1, \mu_2)$ if for each set $A \subset \boldsymbol{R}^n$ there holds

$$p(A \times \boldsymbol{R}^n) = \mu_1(A) \;\;,\;\; p(\boldsymbol{R}^n \times A) = \mu_2(A).$$

It is well known that d defines a metric on the set of probability measures μ on R^n having finite second moments: $\int_{R^n} |x|^2 \mu(dx) < \infty$ [10, 11]. We note that the Wasserstein metric may be equivalently defined by [10]

$$d(\mu_1, \mu_2)^2 = \inf \mathbf{E}|X - Y|^2 , \tag{8}$$

where $\mathbf{E}(U)$ denotes the expectation of the random variable U, and the infimum is taken over all random variables X and Y such that X has distribution μ_1 and Y has distribution μ_2. Since $\mathbf{E}(|X-Y|^2) = \mathbf{E}(|X|^2) + \mathbf{E}(|Y|^2) - 2\,\mathbf{E}(X \cdot Y)$, and since $\mathbf{E}(|X|^2) = \int_{R^n} |x|^2 \, \mu_1(dx)$ and $\mathbf{E}(|Y|^2) = \int_{R^n} |y|^2 \, \mu_2(dy)$ are fixed, we see that the infimum in (8) is attained when the correlation between the random variables X and Y is maximal. The Wasserstein metric induces a *weak* topology on the space of probability measures.

The variational problem (7) is an example of a Monge–Kantorovich mass transference problem with the particular cost function $c(x, y) = |x - y|^2$ [10]. Variational principles of this type find applications in many disciplines, including economics, statistics, and differential geometry [10, 12]. An infimizer $p^* \in \mathcal{P}(\mu_1, \mu_2)$ is referred to as an optimal transference plan. When μ_1 and μ_2 have finite second moments, the existence of such a p^* for (7) is readily verified [11, 13]. In the case that the measures μ_1 and μ_2 have densities ρ_1 and ρ_2 with respect to the Lebesgue measure, we will write $\mathcal{P}(\rho_1, \rho_2)$ for the set of probability measures having first marginal μ_1 and second marginal μ_2. Correspondingly, we will denote by $d(\rho_1, \rho_2)$ the Wasserstein distance between μ_1 and μ_2. This is the situation that we will be concerned with in the sequel.

3 The discrete scheme

We shall now construct a time–discrete scheme that is designed to converge in an appropriate sense (to be made precise below) to a solution of the Fokker–Planck equation. Our construction is motivated by a similar scheme developed by Otto in an investigation of pattern formation in magnetic fluids [14]. It was in [14] that the connection between the Wasserstein metric and dynamical problems involving diffusion or dissipation was first illustrated.

We assume for simplicity that the potential Ψ is smooth and nonnegative and that it satisfies appropriate growth conditions. In particular, we allow for cases in which $\int_{R^n} \exp(-\beta\Psi)\, dx$ is not defined, so that the stationary density ρ_s given by (5) does not exist. These assumptions allow us to treat a wide class of Fokker–Planck equations, including the classical diffusion equation $\frac{\partial \rho}{\partial t} = \beta^{-1}\Delta\rho$. We also introduce the set K of admissible probability densities:

$$K := \left\{ \rho \colon \boldsymbol{R}^n \to [0,\infty) \;\middle|\; \int_{R^n} \rho(x)\, dx = 1\,,\, \int_{R^n} |x|^2\, \rho(x)\, dx < \infty \right\}.$$

With these assumptions and definitions in hand, we now formulate the iterative discrete scheme:

$$\left.\begin{array}{l} \text{Determine } \rho^{(k)} \text{ that minimizes} \\ \\ \frac{1}{2}\, d(\rho^{(k-1)}, \rho)^2 \;+\; h\, F(\rho) \\ \\ \text{over all } \rho \in K\,. \end{array}\right\} \tag{9}$$

where $\rho^{(0)} = \rho^0$ is a fixed probability density in K. This scheme exhibits two competing tendencies. On the one hand, there is the tendency to minimize the free energy functional F at each step, while, on the other hand, there is the tendency to maintain a large correlation between successive iterates (see eq.(8)).

The existence and uniqueness of solutions essentially follow from the fact

that (9) is a strictly convex variational scheme. In fact, the constraint set K is convex, and the objective functional $\frac{1}{2}\, d(\rho^{(k-1)}, \rho)^2 \,+\, h\, F(\rho)$ is strictly convex. Details are provided in [15].

4 Convergence to the solution of the Fokker–Planck equation

Our main result is that an appropriate interpolation of the solution to the scheme (9) converges to the unique solution of the Fokker–Planck equation (6). The convergence result that we can establish is:

THEOREM *Let $\rho^0 \in K$ satisfy $F(\rho^0) < \infty$. For given $h > 0$, let $\{\rho_h^{(k)}\}_{k \in N}$ be the solution of (9). Define the interpolation $\rho_h : (0,\infty) \times \boldsymbol{R}^n \to [0,\infty)$ by*

$$\rho_h(t) \;=\; \rho_h^{(k)} \quad \textit{for } t \in [k\,h, (k+1)\,h) \;\; \textit{and } \; k \in \boldsymbol{N} \cup \{0\}\,.$$

Then as $h \downarrow 0$,

$$\rho_h(t) \;\rightharpoonup\; \rho(t) \quad \textit{weakly in } L^1(\boldsymbol{R}^n) \quad \textit{for all } \; t \in (0,\infty)\,, \tag{10}$$

where $\rho \in C^\infty((0,\infty) \times \boldsymbol{R}^n)$ is the unique solution of the Fokker–Planck equation (6) with the initial condition $\rho(0,x) = \rho^0(x)$ interpreted in the sense that

$$\rho(t) \;\rightharpoonup\; \rho^0 \quad \textit{strongly in } \; L^1(\boldsymbol{R}^n) \quad \textit{for } \;\; t \downarrow 0\,, \tag{11}$$

and with

$$M(\rho),\, E(\rho) \;\in\; L^\infty((0,T)) \quad \textit{for all } \; T < \infty\,, \tag{12}$$

where $M(\rho) = \int_{R^n} |x|^2 \rho(x)\, dx$.

The proof of this Theorem, which is given in [15], is rather lengthy and technical. The crucial step in the analysis, however, is to demonstrate that the

first variation of the functional

$$K \ni \rho \mapsto \frac{1}{2} d(\rho^{(k-1)}, \rho)^2 + hF(\rho) \tag{13}$$

with respect to the independent variables does indeed yield a time–discrete scheme for the Fokker–Planck equation. We shall now illustrate this fact. For simplicity, we will set $\beta = 1$.

Let $\xi \in C_0^\infty(\boldsymbol{R}^n, \boldsymbol{R}^n)$ be a smooth vector field with bounded support, and define the associated *flux* $\{\Phi_\tau\}_{\tau \in R}$, by

$$\partial_\tau \Phi_\tau \;=\; \xi \circ \Phi_\tau \;\text{ for all } \tau \in \boldsymbol{R} \quad \text{and} \quad \Phi_0 \;=\; \mathrm{id}\,.$$

For any $\tau \in \boldsymbol{R}$, let the measure $\rho_\tau(y)\,dy$ be the *push forward* of $\rho^{(k)}(y)\,dy$ under Φ_τ. This means that

$$\int_{R^n} \rho_\tau(y)\,\zeta(y)\,dy \;=\; \int_{R^n} \rho^{(k)}(y)\,\zeta(\Phi_\tau(y))\,dy \quad \text{for all } \zeta \in C_0^0(\boldsymbol{R}^n)\,. \tag{14}$$

Since Φ_τ is invertible, (14) is equivalent to the following relation for the densities:

$$\det \nabla \Phi_\tau \; \rho_\tau \circ \Phi_\tau \;=\; \rho^{(k)}\,. \tag{15}$$

Because $\rho^{(k)}$ minimizes (13), we have for each $\tau > 0$

$$\tfrac{1}{\tau}\,\Big(\Big(\tfrac{1}{2}\,d(\rho^{(k-1)}, \rho_\tau)^2 + h\,F(\rho_\tau)\Big) - \Big(\tfrac{1}{2}\,d(\rho^{(k-1)}, \rho^{(k)})^2 + h\,F(\rho^{(k)})\Big)\Big) \;\geq\; 0, \tag{16}$$

We now examine this inequality in the limit $\tau \downarrow 0$. To begin with, we investigate the first variation of the square of the Wasserstein distance. To this end, let p be optimal in the definition of $d(\rho^{(k-1)}, \rho^{(k)})^2$ (see Section 2). The formula

$$\int_{R^n \times R^n} \zeta(x,y)\,p_\tau(dx\,dy) = \int_{R^n \times R^n} \zeta(x, \Phi_\tau(y))\,p(dx\,dy)\;, \zeta \in C_0^0(\boldsymbol{R}^n \times \boldsymbol{R}^n)\,,$$

then defines a $p_\tau \in \mathcal{P}(\rho^{(k-1)}, \rho_\tau)$. Consequently, we have

$$\tfrac{1}{\tau}\,\Big(\tfrac{1}{2}\,d(\rho^{(k-1)}, \rho_\tau)^2 \;-\; \tfrac{1}{2}\,d(\rho^{(k-1)}, \rho^{(k)})^2\Big)$$

$$\leq \int_{R^n \times R^n} \tfrac{1}{\tau} \left(\tfrac{1}{2} |\Phi_\tau(y) - x|^2 - \tfrac{1}{2} |y - x|^2 \right) p(dx\, dy) ,$$

which implies that

$$\limsup_{\tau \downarrow 0} \tfrac{1}{\tau} \left(\tfrac{1}{2} d(\rho^{(k-1)}, \rho_\tau)^2 - \tfrac{1}{2} d(\rho^{(k-1)}, \rho^{(k)})^2 \right)$$
$$\leq \int_{R^n \times R^n} (y - x) \cdot \xi(y)\, p(dx\, dy) . \tag{17}$$

A straightforward but tedious analysis using (14) and (15) reveals that the first variations of E and S are given by

$$\tfrac{\mathrm{d}}{\mathrm{d}\tau} [E(\rho_\tau)]_{\tau=0} = \int_{R^n} \nabla\Psi(y) \cdot \xi(y)\, \rho^{(k)}(y)\, dy , \tag{18}$$

and

$$\tfrac{\mathrm{d}}{\mathrm{d}\tau} [S(\rho_\tau)]_{\tau=0} = - \int_{R^n} \rho^{(k)} \operatorname{div}\xi \, dy . \tag{19}$$

Appealing now to (16),(17), (18), and (19) (and the symmetry in $\xi \to -\xi$), we see that

$$\left. \begin{aligned} &\int_{R^n \times R^n} (y - x) \cdot \xi(y)\, p(dx\, dy) + h \int_{R^n} (\nabla\Psi \cdot \xi - \operatorname{div}\xi)\, \rho^{(k)}\, dy = 0 \\ &\text{for all } \xi \in C_0^\infty(\boldsymbol{R}^n, \boldsymbol{R}^n) . \end{aligned} \right\} \tag{20}$$

Also, using the fact that $p \in \mathcal{P}(\rho^{(k-1)}, \rho^{(k)})$ together with Hölder's inequality, we find that

$$\left| \int_{R^n} (\rho^{(k)} - \rho^{(k-1)})\, \zeta \, dy - \int_{R^n \times R^n} (y - x) \cdot \nabla\zeta(y)\, p(dx\, dy) \right|$$
$$= \left| \int_{R^n \times R^n} \Big(\zeta(y) - \zeta(x) + (x - y) \cdot \nabla\zeta(y) \Big)\, p(dx\, dy) \right|$$
$$\leq \tfrac{1}{2} \sup_{R^n} |\nabla^2 \zeta|\, d(\rho^{(k-1)}, \rho^{(k)})^2 ,$$

for all $\zeta \in C_0^\infty(\boldsymbol{R}^n)$. Now, choosing $\xi = \nabla\zeta$ in (20), we obtain

$$\left. \begin{aligned} &\left| \int_{R^n} \left\{ \tfrac{1}{h} (\rho^{(k)} - \rho^{(k-1)})\, \zeta + (\nabla\Psi \cdot \nabla\zeta - \Delta\zeta)\, \rho^{(k)} \right\} dy \right| \\ &\quad \leq \tfrac{1}{2} \sup_{R^n} |\nabla^2 \zeta|\, \tfrac{1}{h}\, d(\rho^{(k-1)}, \rho^{(k)})^2 \text{ for all } \zeta \in C_0^\infty(\boldsymbol{R}^n) . \end{aligned} \right. \tag{21}$$

Upon (formally) integrating by parts, the integral on the left hand side of (21) becomes

$$\int_{R^n} \zeta(\tfrac{1}{h}\,(\rho^{(k)} - \rho^{(k-1)}) - \mathrm{div}(\rho^{(k)}\nabla\Psi) - \Delta\rho^{(k)})\,dy\,.$$

In (21), therefore, we recognize a time-discrete version of the Fokker–Planck equation.

The conclusions of the Theorem now follow from a careful analysis of the limit $h \downarrow 0$. We refer the reader to [15] for the mathematical details.

REMARKS

1. A finer analysis reveals that

$$\rho_h \;\to\; \rho \quad \textit{strongly in } L^1((0,T)\times \boldsymbol{R}^n) \quad \textit{for all } T < \infty\,.$$

2. With minor modifications to the scheme (9), we can establish an analogous convergence theorem for Fokker–Planck equations with time dependent diffusion coefficient (i.e., for $\beta = \beta(t)$). This could perhaps lead to some new insight into or improvement upon cooling schedules for the simulated annealing type algorithms for global optimization. We plan to explore these possibilities in future investigations.

Acknowledgments

The research of all three authors is partially supported by the ARO and the NSF through grants to the Center for Nonlinear Analysis. In addition, FO is partially supported by the Deutsche Forschungsgemeinschaft (German Science Foundation), and DK is partially supported by grants NSF/DMS 9505078 and DAAH04–96–0060.

References

[1] R. Jordan and D. Kinderlehrer, in *Partial differential equations and applications*, P. Marcellini, G. G. Talenti, and E. Vesintini, eds., Lecture Notes in Pure and Applied Mathematics, **177**, Marcel Dekker, 1996.

[2] R. S. Ellis, *Entropy, large deviations, and statistical mechanics*, Springer–Verlag, 1985.

[3] L. C. Young, *Lectures on the calculus of variations and optimal control*, W. B. Saunders, 1969.

[4] F. Aluffi-Pentini, V. Parisi and F. Zarilli, *J. Optim. Theo. Appl.* **47** (1986) 1.

[5] C. W. Gardiner, *Handbook of stochastic methods, 2nd ed.*, Springer–Verlag, 1985.

[6] H. Risken, *The Fokker–Planck equation: Methods of solution and applications, 2nd ed.*, Springer–Verlag, 1989.

[7] Z. Schuss, *SIAM J. Review* **22**(1980) 119.

[8] R. Courant, K. Friedrichs and H. Lewy, *Math. Ann.* **100**(1928) 1.

[9] J. C. Strikwerda, *Finite difference schemes and partial differential equations*, Wardsworth & Brooks/Cole, 1989.

[10] S. T. Rachev, *Probability metrics and the stability of stochastic models*, Wiley, 1991.

[11] C. R. Givens and R. M. Shortt, *Michigan Math. J.* **31**(1984) 231.

[12] W. Gangbo and R. J. McCann, preprint (1996).

[13] Y. Brenier, *Comm. Pure Appl. Math.* **44**(1991) 375.

[14] F. Otto, *Archive Rat. Mech. Anal.* in press.

[15] R. Jordan, D. Kinderlehrer and F. Otto, submitted.