

Comment on two distinct notions of free energy

Carsten Hartmann*, Christof Schütte

Freie Universität Berlin, Institut für Mathematik II, Arnimallee 6, 14195 Berlin, Germany

Received 8 December 2006; received in revised form 12 February 2007; accepted 26 February 2007

Available online 2 March 2007

Communicated by M. Vergassola

Abstract

In classical molecular dynamics free energy is one of the most important quantities in analyzing a molecular system. In addition to the standard free energy there is a related free energy concept that is prevalent in transition state theory, but which relies on a different ensemble concept. We show that problems that rely on either definition can be treated in a uniform way using constrained molecular dynamics with no need for unbiasing the respective probability ensembles. Not only does this prove useful in designing algorithms that sample the free energy landscape, but also it clarifies the relation between various results that are available in the literature. In particular we explain that the famous Blue Moon formula is an instance of Federer's co-area formula, and can easily be generalized to phase space observables. We moreover argue that Blue Moon reweighting also becomes an issue for first-order dynamical systems (e.g., Brownian motion).

© 2007 Elsevier B.V. All rights reserved.

Keywords: Molecular dynamics; Free energy calculation; Rare events; Thermodynamic integration; Constrained Hamiltonian systems; Conditional expectation; Blue Moon formula; Fixman theorem; Force of constraint

1. Introduction

The calculation of free energy profiles along certain prescribed (reaction) coordinates plays an essential role in physical chemistry and dynamical systems. In particular in molecular dynamics applications there are a variety of phenomena, for instance, molecular solvation, enzyme catalysis, and conformation dynamics, the understanding of which is directly related to the corresponding free energy landscape.

In recent years progress has been made towards algorithms that efficiently sample free energy profiles, yet the question has remained of in which sense the derivative of the free energy can be considered a force. In principle free energy profiles could be easily computed from the marginal probability distribution of the reaction coordinate. However the reaction dynamics is typically slow, and so reliably sampling the distribution is rather tedious. One way out is to constrain the system to fixed values of the reaction coordinate, and then sample the average force acting upon them. The free energy is recovered afterwards by

numerical integration with respect to the reaction coordinate. This widely used technique, which exploits the dichotomy of *free energy as the potential of mean force*, is known as thermodynamic integration and goes back to Kirkwood [1]. There is a related concept which has been often confused with the standard free energy, and which is prevalently used in transition state theory [2]. As we will argue, the term *potential of mean constraint force* applies to this second type of free energy (geometric free energy). Intriguingly it is possible to treat both concepts within a unifying algorithmic framework, where the mean force is simply computed by directly averaging over the respective Lagrange multiplier that is obtained from constraining the reaction coordinate. In contrast to the famous Blue Moon ensemble method [3] our approach requires no reweighting or unbiasing of the expectation values. Moreover, and in contrast to the case for available approaches, there is no need for computing second-order derivatives of the reaction coordinates.

We briefly review the concept of free energy as it is used in molecular dynamics. Then we explain in which sense the free energy can be considered as a potential of mean constraint force by investigating the relation between the two most important definitions. This allows for linking the various results that circulate in the literature, and finally leads to a deeper

* Corresponding author.

E-mail addresses: chartman@math.fu-berlin.de (C. Hartmann), schuette@math.fu-berlin.de (Ch. Schütte).

understanding of the Blue Moon ensemble method for the sampling of rare events.

2. Two definitions

Consider a Hamiltonian $H : T^*\mathbf{R}^n \rightarrow \mathbf{R}$ on the phase space $T^*\mathbf{R}^n \cong \mathbf{R}^n \times \mathbf{R}^n$,

$$H(q, p) = \frac{1}{2} \langle M^{-1}p, p \rangle + V(q), \quad (2.1)$$

which is the sum of kinetic and potential energy. Here $\langle \cdot, \cdot \rangle$ denotes the usual Euclidean inner product, M is the positive-definite mass matrix containing the molecular masses, and $V : \mathbf{R}^n \rightarrow \mathbf{R}$ is the molecular interaction potential. For the sake of simplicity we introduce mass-scaled coordinates $(q, p) \mapsto (M^{1/2}q, M^{-1/2}p)$, which allows us to set $M = \mathbf{1}$ and to identify the tangent and cotangent space (phase space) in the following.

Let $\xi : \mathbf{R}^n \rightarrow \mathbf{R}^s$ be a smooth function (reaction coordinate), and consider its level sets $\Sigma_x = \xi^{-1}(x)$ for regular values $x \in \mathbf{R}^s$ of ξ . The Σ_x are smooth submanifolds of codimension s in \mathbf{R}^n . (The submanifold property requires that the Jacobian $J_\xi(q) = \mathbf{D}\xi(q)$ has maximum rank s almost everywhere on Σ_x .) Moreover we assume that the dynamics generated by $H(q, p)$ has the unique invariant distribution $\rho \propto \exp(-\beta H(q, p))$, where $\beta = 1/T$ denotes the inverse temperature.

The traditional way to define free energy is by means of the marginal density of the reaction coordinate

$$Z(x) = \int_{\mathbf{R}^n \times \mathbf{R}^n} \exp(-\beta H(q, p)) \delta(\xi(q) - x) dq dp.$$

Employing Federer's co-area formula [4], the marginal density can be equivalently expressed as a surface integral

$$Z(x) = \int_{\Sigma_x \times \mathbf{R}^n} \exp(-\beta H) (\text{vol } J_\xi)^{-1} d\mathcal{H}_\xi, \quad (2.2)$$

where $d\mathcal{H}_\xi$ is the surface element (Hausdorff measure) of $\Sigma_x \times \mathbf{R}^n$ considered as a submanifold of phase space $T^*\mathbf{R}^n \cong \mathbf{R}^n \times \mathbf{R}^n$. Since the reaction coordinate is merely defined on configuration space, we have $d\mathcal{H}_\xi = d\sigma_\xi dp$ with $d\sigma_\xi$ being the surface element of Σ_x considered as a submanifold of \mathbf{R}^n . The volume of the rectangular matrix J_ξ is defined as

$$\text{vol } J_\xi(q) = \sqrt{\det J_\xi^T(q) J_\xi(q)}, \quad q \in \Sigma_x.$$

The *standard free energy* is obtained as the logarithm of the partition function $Z(x)$,

$$F(x) = -\beta^{-1} \ln Z(x). \quad (2.3)$$

Another way to define a free energy that is important in the context of transition state theory [5,6] and optimal prediction [7, 8] utilizes the probability density of the submanifold $\Sigma_x \times \mathbf{R}^n \subset \mathbf{R}^n \times \mathbf{R}^n$:

$$Q(x) = \int_{\Sigma_x \times \mathbf{R}^n} \exp(-\beta H) d\mathcal{H}_\xi. \quad (2.4)$$

Here, the free energy is given by the logarithm of $Q(x)$,

$$G(x) = -\beta^{-1} \ln Q(x). \quad (2.5)$$

In contrast to the standard free energy the last quantity is defined only by the surface Σ_x and does not depend on the specific form of the function ξ , as has been pointed out on various occasions [2,9]. On the other hand the free energy F is directly available from the marginal density of the reaction coordinate ξ , whereas G is not; in fact the authors of [9] call only F a proper free energy. Nevertheless it is easy to see that

$$F(x) = G(x) - \beta^{-1} \ln \mathbf{E}_Q(\text{vol } J_\xi)^{-1}, \quad (2.6)$$

where \mathbf{E}_Q denotes the conditional expectation

$$\mathbf{E}_Q f = \frac{1}{Q(x)} \int_{\Sigma_x \times \mathbf{R}^n} f \exp(-\beta H) d\mathcal{H}_\xi.$$

According to this we can easily switch between the two free energies, F and G , just by adding or subtracting the averaged *Fixman potential*

$$U(x) = -\beta^{-1} \ln \mathbf{E}_Q(\text{vol } J_\xi)^{-1}. \quad (2.7)$$

For the sake of clarity, we shall call G the *geometric free energy*, since it marks an intrinsic property of the foliation defined by the reaction coordinate. Since the quantity F is certainly the most prevalent free energy in the molecular dynamics literature, it will be referred to as the *standard free energy*.

3. Thermodynamic integration

The method of thermodynamic integration takes advantage of the fact that the derivative of the (standard) free energy can be computed from an average force (or pseudo-force) acting upon the fibres $\Sigma_x = \xi^{-1}(x)$. Eventually the free energy is recovered by numerical integration of the thus obtained force field.

We shall now try to clarify in which sense the derivative of both standard and geometric free energy can be considered a force. To this end we first establish a relation between the geometric free energy $G(x)$ and the average force acting transversally to the level sets Σ_x . Before stating the main result we introduce some notation: For each $\sigma \in \Sigma_x$ we can decompose the momentum space $T_\sigma^* \mathbf{R}^n = T_\sigma^* \Sigma_x \oplus N_\sigma^* \Sigma_x$ into tangential and normal components. Upon identifying $T_\sigma^* \mathbf{R}^n$ with \mathbf{R}^n , this is a decomposition of \mathbf{R}^n . In the same way, $p = p_t + p_n$ is a decomposition of an arbitrary momentum vector $p \in \mathbf{R}^n$, such that $p_t \in T_\sigma^* \Sigma_x$ denotes the tangential momentum component, and $p_n \in N_\sigma^* \Sigma_x$ labels its normal part. Then the following statement holds true.

Theorem 3.1. *Let the geometric free energy $G(x)$ be defined according to (2.5). Its derivative with respect to the reaction coordinate is given by the conditional average*

$$\nabla G(x) = \mathbf{E}_Q(\mathbf{d}H)_x, \quad (3.1)$$

where $(\mathbf{d}H)_x = \partial H / \partial \xi|_{\xi=x}$ denotes the generalized force along the reaction coordinate evaluated on the subspace $(T^*\mathbf{R}^n)|_{\Sigma_x} \cong \Sigma_x \times \mathbf{R}^n$. It admits the decomposition

$$(\mathbf{d}H)_x = -\lambda(q, p_t) + \omega(q, p_t, p_n), \quad q \in \Sigma_x, \quad (3.2)$$

where the first term is the constraint force that is needed to constrain a particle with the Hamiltonian H to the surface Σ_x ; it depends only on points on the constrained phase space,

$$T^*\Sigma_x = \{(q, p) \in \mathbf{R}^n \times \mathbf{R}^n \mid \xi(q) = x \text{ and } \langle \nabla \xi(q), \partial H / \partial p \rangle = 0\}.$$

The rightmost term in (3.2) is a coupling term which is linear in both tangential and normal momenta, and that averages to zero, i.e., $\mathbf{E}_Q \omega = 0$.

Proof. We give only a short sketch of the proof which relies on defining appropriate local coordinates, and refer the reader to [10,8] for the details. Introducing Fermi coordinates [11] in a small tubular neighbourhood of each level set, the Hamiltonian (2.1) takes the form

$$H(z, \rho, w, \zeta) = \frac{1}{2} \left\langle (g(z, \rho))^{-1} (w, \zeta)^T, (w, \zeta)^T \right\rangle + V(z, \rho),$$

where $z = (z^1, \dots, z^{n-s})$ are local coordinates on Σ_x and $\rho = (\rho^1, \dots, \rho^s)$ measures the distance from Σ_x with respect to an orthonormal frame attached to Σ_x . The matrix g^{-1} is the inverse of the local metric tensor,

$$g(z, \rho) = \begin{pmatrix} A_\Sigma(z) + C(z, \rho) & \Omega(z, \rho) \\ \Omega^T(z, \rho) & \mathbf{1} \end{pmatrix}.$$

Here $A_\Sigma \in \mathbf{R}^{(n-s) \times (n-s)}$ denotes the metric on Σ_x (first fundamental form), and $C \in \mathbf{R}^{(n-s) \times (n-s)}$ involves the extrinsic curvature of Σ_x in \mathbf{R}^n (second fundamental form). The off-diagonal matrices $\Omega \in \mathbf{R}^{(n-s) \times s}$, $\Omega^T \in \mathbf{R}^{s \times (n-s)}$ are linear in ρ and are related to the orthonormal frame that spans the normal space over Σ_x (so-called normal fundamental forms or normal connections). The generalized momenta (w, ζ) are the local coordinate expressions of tangential and normal momenta, p_t and p_n , above.

The result follows by differentiation of (2.5) upon noting that $\xi(z, \rho) - x = (J_\xi^T Q)(z)\rho$,

where the matrix $Q \in \mathbf{R}^{n \times s}$ contains the vectors of the normal frame (the orthonormalized columns of the Jacobian J_ξ). \square

Roughly speaking, the generalized force along the reaction coordinate is the contribution of the Hamiltonian vector field perpendicular to the level set Σ_x , where the decomposition (3.2) is an instance of the *Fundamental Equations for Submanifolds* [12]: The constraint force contains the extrinsic curvature of the submanifold Σ_x in the ambient space \mathbf{R}^n , which is quadratic in the tangential momenta. The components of ω involve the normal connection stemming from the off-diagonal terms of the metric tensor; physically speaking, the normal connection couples vectors fields that are tangential and normal to Σ_x , satisfying $\omega(q, p_t, 0) = 0$. It is linear in both p_t and p_n ; hence it vanishes upon taking the average in (3.1), for the momenta follow the Maxwell distribution, i.e., a zero-mean Gaussian distribution. For details on the constraint force geometry, see the textbook [13].

In the following we can simply ignore the connection term and focus on the constraint force part. Formally the constrained

Hamilton equations read

$$\dot{q}^i = \frac{\partial H}{\partial p_i} \quad (3.3)$$

$$\dot{p}_i = -\frac{\partial H}{\partial q^i} - \sum_{k=1}^s \lambda_k \frac{\partial \xi^k}{\partial q^i}, \quad \xi(q) = x, \quad (3.4)$$

where $i = 1, \dots, n$, and λ_k denotes the Lagrange undetermined multipliers. The fact that ξ explicitly appears in the equations of motion should not conceal that the dynamics is intrinsic to the constrained phase space $T^*\Sigma_x$. That is, a constrained system is completely specified by the submanifold Σ_x , not by the, in some way arbitrary, function ξ . As $\Sigma_x \subset \mathbf{R}^n$, we can take advantage of the natural inclusion of the respective phase spaces $T^*\Sigma_x \subset T^*\mathbf{R}^n$, and (similar to the viewpoint of Lagrangian mechanics) construct a constrained Hamiltonian H_Σ by simply restricting it to the constrained phase space, i.e., $H_\Sigma = H|_{T^*\Sigma_x}$. The canonical density generated by the equations of motion (3.3) and (3.4) thus reads

$$\rho_\Sigma = Q_\Sigma^{-1} \exp(-\beta H_\Sigma).$$

The normalization constant

$$Q_\Sigma = \int_{T^*\Sigma_x} \exp(-\beta H_\Sigma) d\mathcal{L}_\Sigma$$

should be distinguished from the partition function $Q(x)$ in Eq. (2.4). The respective constrained expectation is denoted by \mathbf{E}_Σ . Here $d\mathcal{L}_\Sigma$ labels the surface element (Liouville measure) of the constrained phase space $T^*\Sigma_x \subset \mathbf{R}^n \times \mathbf{R}^n$, i.e., the former surface element $d\mathcal{H}_\xi$ plus the momentum constraint.

The reader may notice that the Lagrange multiplier $\lambda \in \mathbf{R}^s$ in the constrained equations of motion (3.3) and (3.4) equals the function λ in Theorem 3.1. This can be seen by differentiating the constraint twice and inserting the equations of motion (3.3) and (3.4), which amounts to writing the Lagrange multiplier as a phase space function $\lambda = \lambda(q, p)$ with (q, p) satisfying the constraint.

Notice that only the normal momenta $p_n \in N^*\Sigma_x$ distinguish the constrained expectation \mathbf{E}_Σ from the conditional expectation \mathbf{E}_Q . Anyway the constraint force does not depend on the normal momenta, and nor does the Lagrange multiplier which is defined subject to the constraint. Hence we can replace the conditional expectation in (3.1) by \mathbf{E}_Σ , and find for the derivative of the geometric free energy

$$\nabla G(x) = -\mathbf{E}_\Sigma \lambda, \quad (3.5)$$

or,

$$\nabla G(x) = -\mathbf{E}_Q \lambda. \quad (3.6)$$

Remark 3.2. A special situation arises, when ξ is scalar. Then the level sets Σ_x are codimension-1 submanifolds of \mathbf{R}^n (hypersurfaces), in which case the normal connection in (3.2) vanishes identically; see [10].

4. Potential of mean force

By definition, we can recover $G(x)$ by numerical integration of $\nabla G(x)$ over the components of the reaction coordinate.

Eq. (3.5) is particularly useful for computing the mean force during the course of the numerical integration of a constrained system: the mean force is simply the time average of the Lagrange multiplier $\lambda(t) = \lambda(q(t), p(t))$ along the constrained orbits $(q(t), p(t)) \in T^*\Sigma_x$, provided the dynamics is ergodic with respect to the constrained canonical density ρ_Σ ; no further function evaluations are required. Taking moreover advantage of (2.6) we find $F(x)$ after adding the Fixman potential (2.7) upon replacing \mathbf{E}_Q by \mathbf{E}_Σ . Neither does this require any sort of reweighting or the calculation of second-order derivatives.

The result is in perfect agreement with the formulae that have been derived in various articles for the case when the reaction coordinate is one dimensional [14–17], but also in the multidimensional case [10,18–20]. Notice that averaging the Lagrange multiplier over the tangential momenta reduces Eq. (3.1) to the expression derived in [21,22]; this becomes evident if one bears in mind that the average of the second fundamental form, that is contained in the constraint force, is the negative (extrinsic) mean curvature which appears there.

It remains to understand in which sense ∇F and ∇G can be considered a force. It is a straightforward consequence of the previous considerations that ∇G transforms like a 1-form, i.e., a force. The same does not hold true for the standard free energy, as was pointed out in [2] for the first time: Suppose we define a new reaction coordinate $\zeta = h(\xi)$, where h is a smooth, strictly monotonic function. Clearly ζ has the same level sets $\Sigma_x = (\zeta^{-1} \circ h)(x)$ as the former reaction coordinate, but

$$F(h(x)) = F(x) + \beta^{-1} \ln |\det \mathbf{D}h(x)|. \quad (4.1)$$

Consequently, $\nabla F(x)$ does not transform like a 1-form, and, strictly speaking, cannot be considered a force. (Even worse, the additional term in the transformed standard free energy depends on temperature.) It is easy to see that the gauge term which degrades the transformation properties of the standard free energy stems from the matrix volume in the partition function (2.2). Indeed it follows immediately from the definitions of both standard and geometric free energy that the underlying expectation values are related by the famous Blue Moon reweighting formula of Carter et al. [3]: Let $f = f(q, p)$ be an integrable phase space function. Then

$$\mathbf{E}_Z f = \frac{\mathbf{E}_Q f(\text{vol } J_\xi)^{-1}}{\mathbf{E}_Q(\text{vol } J_\xi)^{-1}}, \quad (4.2)$$

where \mathbf{E}_Z is the conditional expectation that is associated with the marginal probability in (2.2). If moreover $f = f(q, p_t)$ with $p_t \in T_{q_t}^*\Sigma$ satisfying the constraint we can replace \mathbf{E}_Q by the constrained expectation \mathbf{E}_Σ , such that

$$\mathbf{E}_Z f = \frac{\mathbf{E}_\Sigma f(\text{vol } J_\xi)^{-1}}{\mathbf{E}_\Sigma(\text{vol } J_\xi)^{-1}}. \quad (4.3)$$

Hence the Blue Moon relation (4.3) turns out to be an instance of Federer's co-area formula; see the work of Fixman [23] for further comparison. This is an intriguing insight, for the typical reasoning that leads to the Blue Moon formula involves reference to a specific type of dynamical system that is used to sample the respective probability densities. However we find

that the Blue Moon relation can be understood as a matter of defining certain expectation values. In particular (4.2) holds true, although there are no (hidden) constraints on the momenta at all. This remains true, even if the system does not involve any momenta or velocities. As an example, consider the particular case of constrained Brownian motion as has been addressed recently in [22]: Upon omitting momenta, we have $\mathbf{E}_Q = \mathbf{E}_\Sigma$ which amounts to the proper constraint $q \in \Sigma_x$. In contrast to that, \mathbf{E}_Z corresponds to the expectation with respect to the unconstrained canonical density conditional on $\xi(q) = x$.

Acknowledgments

We thank Tony Lelièvre, Eric Vanden-Eijnden and Giovanni Ciccotti for vibrant and fruitful discussions on the present topic. The work was supported by the DFG Research Center MATH-EON “Mathematics for Key Technologies” (FZT86) in Berlin.

References

- [1] J.G. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* 3 (1935) 300–313.
- [2] W. E, E. Vanden-Eijnden, Metastability, conformation dynamics, and transition pathways in complex systems, in: S. Attinger, P. Koumoutsakos (Eds.), *Multiscale, Modelling, and Simulation*, Springer, Berlin, 2004, pp. 35–68.
- [3] E.A. Carter, G. Ciccotti, J.T. Hynes, R. Kapral, Constrained reaction coordinate dynamics for the simulation of rare events, *Chem. Phys. Lett.* 156 (5) (1989) 472–477.
- [4] H. Federer, *Geometric Measure Theory*, Springer, Berlin, 1969.
- [5] C.H. Bennett, Molecular dynamics and transition state theory: The simulation of infrequent events, in: R.E. Christoffersen (Ed.), *Algorithms for Chemical Computations*, Amer. Chem. Soc., Washington, DC, 1977, pp. 63–97.
- [6] D. Chandler, Statistical mechanics of isomerization dynamics in liquids and the transition state approximation, *J. Chem. Phys.* 68 (1978) 2959–2970.
- [7] A.J. Chorin, O.H. Hald, R. Kupferman, Optimal prediction and the Mori–Zwanzig representation of irreversible processes, *Proc. Natl. Acad. Sci. USA* 97 (7) (2000) 2969–2973.
- [8] C. Hartmann, Model reduction in classical molecular dynamics, Ph.D. Thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2006 (in press). Available soon at: <http://jorge.ub.fu-berlin.de/work/DissSearch>.
- [9] E. Vanden-Eijnden, F.A. Tal, Transition state theory: Variational formulation, dynamical corrections, and error estimates, *J. Chem. Phys.* 123 (2005) 184103.
- [10] C. Hartmann, Ch. Schütte, A constrained hybrid Monte-Carlo algorithm and the problem of calculating the free energy in several variables, *ZAMM Z. Angew. Math. Mech.* 85 (10) (2005) 700–710.
- [11] P. Maraner, A complete perturbative expansion for quantum mechanics with constraints, *J. Phys. A* 28 (1995) 2939–2951.
- [12] M. Spivak, *Differential Geometry*, vol. 4, Publish or Perish, Boston, 1975.
- [13] J.E. Marsden, T.S. Ratiu, *Introduction to Mechanics and Symmetry*, Springer, New York, 1999.
- [14] M. Sprik, G. Ciccotti, Free energy from constrained molecular dynamics, *J. Chem. Phys.* 109 (18) (1998) 7737–7744.
- [15] W.K. den Otter, Thermodynamic integration of the free energy along a reaction coordinate in Cartesian coordinates, *J. Chem. Phys.* 112 (17) (2000) 7283–7286.
- [16] E. Darve, M.A. Wilson, A. Pohorille, Calculating free energies using a scaled-force molecular dynamics algorithm, *Mol. Sim.* 28 (1–2) (2002) 113–144.
- [17] C. Hartmann, Ch. Schütte, A geometric approach to free energy calculations, *Commun. Math. Sci.* 3 (1) (2005) 1–20.

- [18] W.K. den Otter, Free energy from molecular dynamics with multiple constraints, *Mol. Phys.* 98 (12) (2000) 773–781.
- [19] E. Darve, A. Pohorille, Calculating free energies using average force, *J. Chem. Phys.* 115 (20) (2001) 9169–9183.
- [20] I. Coluzza, M. Sprik, G. Ciccotti, Constrained reaction coordinate dynamics for systems with constraints, *Mol. Phys.* 101 (18) (2003) 2885–2894.
- [21] G. Ciccotti, R. Kapral, E. Vanden-Eijnden, Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics, *Chem. Phys. Chem.* 6 (9) (2005) 1809–1814.
- [22] G. Ciccotti, T. Lelièvre, E. Vanden-Eijnden, Sampling Boltzmann–Gibbs distributions restricted on a manifold with diffusions: Application to free energy calculations, *Rapport de recherche du CERMICS* 2006-309, 2006.
- [23] M. Fixman, Classical statistical mechanics of constraints: A theorem and applications to polymers, *Proc. Natl. Acad. Sci. USA* 71 (1974) 3050–3053.