IL NUOVO CIMENTO **44 C** (2021) 112
DOI 10.1393/ncc/i2021-21112-8

COMMUNICATIONS: SIF Congress 2020

OPES: On-the-fly Probability Enhanced Sampling method

MICHELE INVERNIZZI([1])([2])([3])

([1]) *Department of Physics, ETH Zurich - Lugano, Switzerland*
([2]) *Institute of Computational Science, Università Svizzera Italiana - Lugano, Switzerland*
([3]) *Italian Institute of Technology - Genova, Italy*

received 30 January 2021

Summary. — Molecular simulations are playing an ever-increasing role, finding applications in fields as varied as physics, chemistry, biology and material science. However, many phenomena of interest take place on time scales that are out of reach for standard molecular simulations. This is known as the sampling problem and over the years several enhanced sampling methods have been developed to mitigate this issue. We propose a unified approach that puts on the same footing two of the most popular families of enhanced sampling methods and paves the way for novel combined approaches. The on-the-fly probability enhanced sampling method provides an efficient implementation of such generalized approach, while also focusing on simplicity and robustness.

1. – Unified approach

We present here in a synthetic fashion a recently developed method, called on-the-fly probability enhanced sampling (OPES), that implements a unified approach to rare events sampling in molecular simulations. The method has been introduced in refs. [1] and [2], where it is described in full detail.

The goal of enhanced sampling is to increase the probability of observing in a simulation certain rare events, and to do so in such a way that it is still possible to retrieve statistics about the original system. We call *target distribution* the modified probability distribution that is sampled instead of the Boltzmann one. Apart for some notable exceptions such as path sampling methods [3], we can group most of enhanced sampling into two main families according to how the target distribution is defined.

A first family is the one of methods like umbrella sampling [4] and metadynamics [5]. These methods make use of collective variables (CVs) or order parameters that are smooth functions of the atomistic coordinates and encode the slow modes of the system. In this family, the target distribution is defined by requiring that its marginal probability distribution over such CVs has a given functional form. Typically the marginal probability is chosen to be a constant flat distribution, as in adaptive umbrella sampling [6], but

other choices are possible, such as the well-tempered distribution often used in metadynamics [7] and the many others that have been explored within the variationally enhanced sampling method [8].

A second family includes tempering methods like simulated tempering [9] and replica exchange [10]. These methods define their target distribution as the combination of slightly different versions of the original system, for example the same system but at higher temperatures. These target distributions are also known as generalized ensembles or expanded ensembles [11].

The OPES method can be used to sample either kind of target distributions. It does so by adding to the potential energy of the system $U(\mathbf{x})$ a bias potential $V(\mathbf{x})$ such that the sampled distribution is not the equilibrium Boltzmann distribution, $P(\mathbf{x}) \propto e^{-\beta U(\mathbf{x})}$, but the target one, $p^{\mathrm{tg}}(\mathbf{x})$. This bias potential is defined as

$$V(\mathbf{x}) = -\frac{1}{\beta} \log \frac{p^{\mathrm{tg}}(\mathbf{x})}{P(\mathbf{x})}, \tag{1}$$

where β is the inverse Boltzmann temperature. The bias potential is not known *a priori*, but it is self-consistently learned during the simulation, via an on-the-fly estimate of the probability distributions. Statistics of the unbiased system can be retrieved via a reweighting procedure, by assuming that the bias is updated in an adiabatic way. Given any observable $O(\mathbf{x})$, its ensemble average $\langle O \rangle$ over the unbiased system can be estimated with ensemble averages over the sampled biased system $\langle O \rangle = \frac{\langle O e^{\beta V} \rangle_V}{\langle e^{\beta V} \rangle_V}$. In this way also free energy differences and free energy surfaces can be estimated [1, 2].

2. – OPES for collective variables sampling

Given a set of collective variables $\mathbf{s} = \mathbf{s}(\mathbf{x})$, one can define the marginal probability $P(\mathbf{s}) = \int P(\mathbf{x}) \delta[\mathbf{s}(\mathbf{x}) - \mathbf{s}] \mathrm{d}\mathbf{x}$. The well-tempered ensemble with respect to these CVs is obtained by requiring that the marginal probability of the target distribution is $p^{\mathrm{WT}}(\mathbf{s}) \propto [P(\mathbf{s})]^{1/\gamma}$, where $\gamma > 1$ is known as bias factor. It should be noted that the exact target distribution $p^{\mathrm{tg}}(\mathbf{x})$ is unknown, but this does not constitute a problem. In fact, the core requirements are that the corresponding bias potential can be explicitly expressed and that the target distribution is easy to sample, *i.e.*, that the kinetic bottlenecks between metastable states are removed. This is indeed guaranteed for the well-tempered distribution, given that the CVs are chosen properly and the bias factor is large enough. The case of uniform marginal target distribution can be seen as a special case of the well-tempered one, where $\gamma = \infty$.

When using OPES for CVs sampling we need to estimate $P(\mathbf{s})$. To do so, we use a weighted kernel density estimation with an automatic kernel-merging algorithm, that is explained in detail in ref. [1]. We also introduce a regularization term ϵ and a normalization Z over the explored CV-space. At step n the bias, eq. (1), can be written as a function of the CVs,

$$V_n(\mathbf{s}) = (1 - 1/\gamma) \frac{1}{\beta} \log \left(\frac{P_n(\mathbf{s})}{Z_n} + \epsilon \right), \tag{2}$$

where $P_n(\mathbf{s})$ is the estimate of $P(\mathbf{s})$ obtained via reweighting. Reference [1] presents the full derivation of this expression.

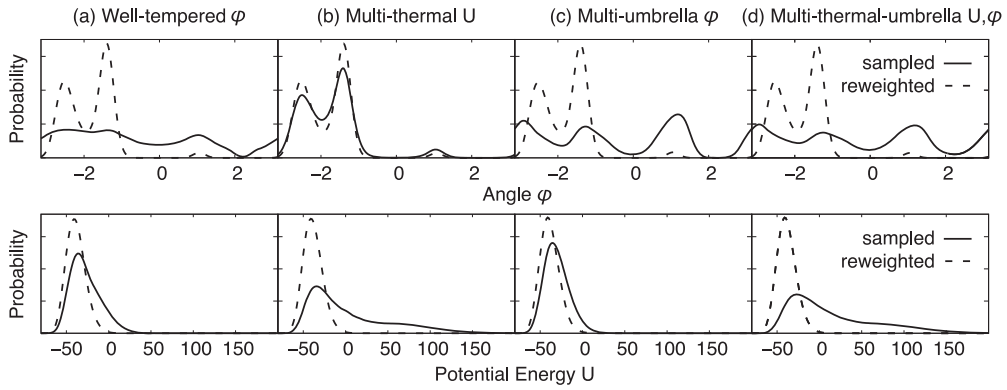

Fig. 1. – Marginal probabilities over the ϕ angle and the potential energy U for OPES simulations of alanine dipeptide with different target distributions. The unbiased distribution as obtained via reweighting is also shown. For each target distribution the biased CVs are specified.

3. – OPES for expanded ensembles sampling

To define the expanded ensemble target distribution, we first define a class of probability distributions $P_\lambda(\mathbf{x}) \propto e^{-\beta U_\lambda(\mathbf{x})}$, where λ can be a parameter (*e.g.,* the temperature) or a set of parameters, and U_0 is the original system potential. For simplicity we only consider nonweighted expanded ensembles, as done in ref. [2]. The expanded ensemble contains a discrete set $\{\lambda\}$ of $N_{\{\lambda\}}$ parameters such that the corresponding $P_\lambda(\mathbf{x})$ have an overlap in the configuration space. We can write the expanded target distribution as $p_{\{\lambda\}}(\mathbf{x}) = \frac{1}{N_{\{\lambda\}}} \sum_\lambda P_\lambda(\mathbf{x})$. One can then define the *expansion collective variables* as $\Delta u_\lambda(\mathbf{x}) = \beta U_\lambda(\mathbf{x}) - \beta U_0(\mathbf{x})$ and use them to write the bias potential at step n,

$$V_n(\mathbf{x}) = -\frac{1}{\beta} \log \left(\frac{1}{N_{\{\lambda\}}} \sum_\lambda e^{-\Delta u_\lambda(\mathbf{x}) + \beta \Delta F_n(\lambda)} \right), \tag{3}$$

where $\Delta F_n(\lambda)$ are the estimates of the free energy differences between the unbiased system U_0 and the one at a given λ. These are obtained via on-the-fly reweighting, similarly to $P_n(\mathbf{s})$ in sect. **2**, but this time without the need for kernel density estimation as $\{\lambda\}$ is a discrete set. The details of the derivation are explained in ref. [2].

Finally, we notice that often it is possible to rewrite eq. (3) so that, similarly to eq. (2), the bias is a function of only a small number of CVs. For example, in case of a multithermal expanded target distribution the bias can be expressed as a function of the potential energy only [2].

4. – Example: alanine dipeptide

As an example we consider alanine dipeptide *in vacuum* at 300 K, which is a prototypical system for enhanced sampling. It presents two main metastable basins, that can be characterized using as CV the torsion angle ϕ. The most stable basin contains two minima and lays in the region where ϕ is negative, while the second basin has one minimum at $\phi \simeq 1$. In fig. 1 we show the marginal probabilities along ϕ and the potential energy U obtained from OPES simulations that aim at different target distributions. Despite being very different, all these target distributions increase the probability of observing

transitions between the metastable states and make it possible to retrieve via reweighting the underlying Boltzmann probability. The well-tempered ($\gamma = 50$), fig. 1(a), and the multiumbrella (43 umbrellas), fig. 1(c), both employ a bias that is a function of the same CV ϕ, but they would look similar only in the limit of $\gamma = \infty$ and many umbrellas respectively [2]. The multithermal distribution, fig. 1(b), spans a range of temperatures between 300 and 1000 K, and can be reweighted at any temperature in that range. It can also be combined with the multiumbrella expansion along ϕ, fig. 1(d), as done in ref. [12].

All the simulation details, together with the inputs and the trajectories are available online on the PLUMED-NEST repository (`www.plumed-nest.org`, plumID:21.006) [13].

5. – Conclusion

We presented a target distribution perspective on enhanced sampling and the on-the-fly probability enhanced sampling method, that have been developed in refs. [1,2]. OPES is a general and flexible method that can be used to sample different types of target distributions. It is also easy to use and robust with respect to suboptimal collective variables [14]. It has been implemented as a contributed module in the open source library PLUMED [15,16], and has been already used for various applications [17-21]. We believe OPES can be a handy tool for anyone interested in enhanced sampling while also having the potential of supporting novel types of target distributions.

* * *

The author thanks Luigi Bonati for carefully reading the manuscript. This research was supported by the NCCR MARVEL, funded by the Swiss National Science Foundation, and European Union Grant No. ERC-2014-AdG-670227/VARMET.

REFERENCES

[1] Invernizzi M. and Parrinello M., *J. Phys. Chem. Lett.*, **11** (2020) 2731.
[2] Invernizzi M., Piaggi P. M. and Parrinello M., *Phys. Rev. X*, **10** (2020) 41034.
[3] Bolhuis P. G. *et al.*, *Annu. Rev. Phys. Chem.*, **53** (2002) 291.
[4] Torrie G. M. and Valleau J. P., *Chem. Phys. Lett.*, **28** (1974) 578.
[5] Laio A. and Parrinello M., *Proc. Natl. Acad. Sci. U.S.A.*, **99** (2002) 12562.
[6] Mezei M., *J. Comput. Phys.*, **68** (1987) 237.
[7] Barducci A., Bussi G. and Parrinello M., *Phys. Rev. Lett.*, **100** (2008) 020603.
[8] Valsson O. and Parrinello M., *Phys. Rev. Lett.*, **113** (2014) 090601.
[9] Marinari E. and Parisi G., *Europhys. Lett.*, **19** (1992) 451.
[10] Sugita Y. and Okamoto Y., *Chem. Phys. Lett.*, **314** (1999) 141.
[11] Lyubartsev A. P. *et al.*, *J. Chem. Phys.*, **96** (1992) 1776.
[12] Sugita Y., Kitao A. and Okamoto Y., *J. Chem. Phys.*, **113** (2000) 6042.
[13] The PLUMED consortium, *Nat. Methods*, **16** (2019) 670.
[14] Invernizzi M. and Parrinello M., *J. Chem. Theory Comput.*, **15** (2019) 2187.
[15] Tribello G. A. *et al.*, *Comput. Phys. Commun.*, **185** (2014) 604.
[16] `www.plumed.org/doc-master/user-doc/html/_o_p_e_s.html`.
[17] Mandelli D., Hirshberg B. and Parrinello M., *Phys. Rev. Lett.*, **125** (2020) 026001.
[18] Bonati L., Rizzi V. and Parrinello M., *J. Phys. Chem. Lett.*, **11** (2020) 2998.
[19] Rizzi V. *et al.*, *Nat. Commun.*, **12** (2021) 93.
[20] Piaggi P. M. *et al.*, *J. Chem. Theory Comput.*, **17** (2021) 3065.
[21] Karmakar T. *et al.*, *Mol. Phys.* (2021) `https://doi.org/10.1080/00268976.2021.1893848`.