Rethinking Metadynamics: from bias potential to probability distribution

Michele Invernizzi\footnote{1}\footnote{2} and Michele Parrinello\footnote{3}\footnote{2} \\

\textit{1 Department of Physics, ETH Zurich, c/o Università della Svizzera italiana,} \\
\textit{Via Giuseppe Buffi 13, 6900, Lugano, Switzerland} \\
\textit{2 Facoltà di Informatica, Institute of Computational Science,} \\
\textit{National Center for Computational Design and Discovery of Novel Materials (MARVEL),} \\
\textit{Università della Svizzera italiana, Via Giuseppe Buffi 13, 6900 Lugano, Switzerland} \\
\textit{3 Department of Chemistry and Applied Biosciences,} \\
\textit{ETH Zurich, c/o Università della Svizzera italiana,} \\
\textit{Via Giuseppe Buffi 13, 6900 Lugano, Switzerland,} \\
\textit{and Italian Institute of Technology, Via Morego 30, 16163 Genova, Italy} \\
(Dated: November 13, 2019)

Metadynamics is an enhanced sampling method of great popularity, based on the on-the-fly construction of a bias potential that is function of a selected number of collective variables. We propose here a drastic change in perspective that shifts the focus from the bias to the probability distribution reconstruction, leading to an improvement in usability and convergence speed. The resulting enhanced sampling method can be seen as a combination of metadynamics and adaptive umbrella sampling approaches, that aims at taking the best from the two worlds. This new method has a straightforward reweighting scheme and allows for efficient importance sampling, avoiding uninteresting high free energy regions. Thanks to a compressed kernel density estimation it can handle a higher dimensional collective variable space, and does not require any prior knowledge of the boundaries of such space. The new method comes in two variants. The first aims at a quick convergence, avoiding oscillations and maximizing the quasi-static bias regime, while in the second the main focus is on a rapid exploration of the free energy landscape. We demonstrate the performance of the method in a number of representative examples.

Enhanced sampling plays a crucial role in modern simulation techniques, and is a very active area of research\cite{1}. Of particular importance has been the work of Torrie and Valleau\cite{2}. They consider a system with an interaction potential $U(R)$, where $R$ denotes the atomic coordinates. Sampling is accelerated by adding a bias potential $V(s)$ that depends on $R$ via a set of collective variables (CVs), $s = s(R)$. The CVs are chosen so as to describe the modes of the system that are more difficult to sample. The choice of a proper set of CVs is critical, as it determines the efficiency of the method. The properties of the unbiased system are then calculated by using a reweighting procedure. In fact the unbiased probability density $P(s) = \langle \delta[s-s(R)] \rangle \propto \int dR e^{-\beta U(R)} \delta[s-s(R)]$ can be written as an average over the biased ensemble:

$$P(s) = \frac{\langle \delta[s-s(R)] e^{\beta V(s)} \rangle_V}{\langle e^{\beta V(s)} \rangle_V},$$

where $\beta$ is the inverse temperature. Since this work, a large number of CV based methods have been proposed.

Adaptive umbrella sampling\footnote{3} (AUS) was the first to build iteratively $V(s)$ so that it would compensate for the free energy surface (FES), $F(s) = -\frac{1}{\beta} \log P(s)$. At the $n$-th iteration the bias is given by:

$$V_n(s) = \frac{1}{\beta} \log P_n(s),$$

where the probability estimate $P_n(s)$ can be obtained via a weighted histogram, with weights $w_k = e^{\beta V_k}$, or some more elaborate estimator\cite{4}, and can be updated iteratively or on the fly\cite{5} \cite{6}. At convergence one has $V(s) = -F(s)$.

A different approach has been introduced by metadynamics\cite{7, 8} (MetaD). Here one builds $V(s)$ directly, instead of first reconstructing the probability distribution. The bias is updated on the fly by adding a Gaussian centered at every new point sampled $s_k$:

$$V_n(s) = \sum_k^n e^{-\beta V_{k-1}(s_k) / (\gamma - 1)} G(s, s_k),$$

where the parameter $\gamma > 1$ is called the bias factor, and the Gaussian function is defined as $G(s, s') = h \exp \left[-\frac{1}{2} (s - s')^T \Sigma^{-1} (s - s')\right]$, with height $h$ and variance $\Sigma$ set by the user. Typically only diagonal variances $\Sigma_{ij} = \sigma^2_i \delta_{ij}$ are employed, but more general choices have also been suggested\cite{9}.

A great advantage of this scheme over AUS, is its speed in adapting to the local free energy landscape and thus quickly escaping metastable states, without remaining stuck. It does so by allowing the bias to change also in a non-adiabatic way, i.e. without waiting for the system to fully relax back to equilibrium. In the first version of metadynamics, where the bias factor $\gamma$ is put to infinity, the bias never ceases to adapt locally and this results in convergence issues\cite{10}, that where solved with the introduction of the well-tempered variant\cite{11}. For a more thorough discussion we refer the reader to the many reviews on the subject\cite{12} \cite{13}.
Another difference with AUS is the convergence relation \( V(s) = -(1 - 1/\gamma)F(s) \). In the biased ensemble it leads to a probability distribution \( p^{WT}(s) \) that is not flat, but rather a smoother version of the starting one, \( p^{WT}(s) \propto [P(s)]^{1/\gamma} \), where FES barriers are lowered by a factor \( \gamma \). Thus well-tempered metadynamics implicitly introduced the possibility of choosing a non-uniform target distribution. The idea of aiming at a specific target distribution has been later generalized \[14\] and various possibilities have been explored \[15\]–\[18\], in particular within the framework of variationally enhanced sampling (VES) method \[19\].

This is an often overlooked point, especially in AUS-like methods. A common misconception is that a flat target distribution is to be preferred for enhanced sampling, because it ensures that there are no free energy barriers left that can slow down the transition rate between metastable states. However in most applications the CVs are suboptimal \[20\] and thus the transition rate is limited by the presence of slow modes not accelerated by \( V(s) \). In this real-life scenario also a well-tempered target leads to the same transition rate as a flat one. Furthermore the well-tempered distribution has the advantage of leading to a more efficient importance sampling, because it is closer to the unbiased distribution and does not push the system to physically uninteresting regions of high free energy.

Also in metadynamics, as in umbrella sampling, one can calculate any quantity in the unbiased ensemble via a reweighting procedure, that can be carried out as soon as the bias changes in an adiabatic fashion. It is common practice \[13\] to use reweighting also to obtain an alternative estimate of the free energy, since the one obtained directly from the bias can oscillate considerably, especially with suboptimal CVs \[20\].

One could argue that the perspective taken by AUS is more natural, being based on the probability distribution which can be directly sampled. And in fact MetaD had to wait more than ten years before seeing a general mathematical proof of its convergence, and research is still going on about the best way to perform reweighting \[21\]–\[24\]. In this letter we combine the probability-centered approach of AUS with some key MetaD ideas, such as the non-adiabatic bias update and a non-flat target distribution. This results in a new method with fast convergence, few intuitive parameters and a straightforward reweighting scheme.

As in MetaD we add Gaussians at constant pace as the simulation proceeds, but instead of directly updating the bias, we update the probability distribution estimate as in AUS. We also want to explicitly introduce a target distribution \( p^t(s) \), thus we write the bias at convergence as:

\[
V(s) = \frac{1}{\beta} \log \frac{P(s)}{p^t(s)},
\]

This is a rather general expression and opens to many possibilities. Here we shall restrict ourselves to considering only well-tempered target distributions:

\[
p^t(s) = p^{WT}(s) \propto [P(s)]^{1/\gamma}.
\]

Since at the beginning of the simulation \( P(s) \) is not known, we adopt a self-consistent scheme, similar to what is done in well-tempered VES \[25\]. This can be achieved in two ways, giving rise to two variants of the proposed method. The first is to estimate on the fly the unbiased distribution \( P(s) \) via reweighting, and then obtain an estimate of \( p^{WT}(s) \) using Eq. \[5\]. The other is to monitor the biased distribution \( p^{WT}(s) \) that is being sampled, and retrieve an estimate of \( P(s) \) by simply inverting Eq. \[5\],

\[
P(s) \propto [p^{WT}(s)]^{\gamma}.
\]

In the first case we write the bias at step \( n \) as:

\[
V_n(s) = (1 - 1/\gamma) \frac{1}{\beta} \log \tilde{P}_n(s),
\]

while in the second:

\[
V_n(s) = (\gamma - 1) \frac{1}{\beta} \log \tilde{p}^{WT}_n(s).
\]

Here and in the following we indicate with the tilde that the distributions in Eq. \[6\] and \[7\] need not be normalized, since normalization adds only an irrelevant constant (see also supplemental material - SM). The two different iterative schemes converge to the desired result (Eq. \[4\]), but they do so in very different ways. In both cases we update the probability distribution on the fly by periodically depositing a Gaussian. This is indeed a common way of reconstructing a probability, known as kernel density estimation (KDE), and we shall draw from the vast literature on the subject \[29\].

The first variant (convergence variant) is conceptually similar to AUS-like methods \[10\] and shares with them some of its strengths and weaknesses (see e.g. Ref. \[27\]). However, our approach presents some key innovations on how the probability density is estimated, which can greatly improve its performance, and simplify the choice of free parameters. In the convergence variant the probability estimate is written as:

\[
\tilde{P}_n(s) = \frac{\sum_{k=1}^{n} w_k G(s, s_k) + \epsilon}{\sum_{k=1}^{n} w_k + 1},
\]

where the weights \( w_k \) are given by

\[
w_k = e^{\beta V_{k-1}(s_k)},
\]

and \( \epsilon \ll 1 \) is an important regularization parameter that ensures the argument of the logarithm in Eq. \[6\] is always greater than zero. As suggested in Ref. \[3\], \( \epsilon \) can be thought of as an initial guess for the probability distribution over the whole space, \( \tilde{P}_0(s) = \epsilon \), but we found it more useful to link \( \epsilon \) to the free energy barrier height \( \Delta E \)
that one would like to overcome. By setting $\epsilon = e^{-\beta \Delta E}$ we indirectly set an upper limit to the bias and thus avoid exploring regions too high in free energy (see [SM] for more details). The $G(s, s_k)$ are unnormalized Gaussians, as those defined previously for MetaD, with diagonal variance $\Sigma_{ij} = \sigma_i^2 \delta_{ij}$ and fixed height $h = 1$. Changing the height $h$ simply corresponds to introducing an overall normalization and thus, contrary to MetaD, $h$ is an irrelevant parameter and can be dropped.

It has been shown [26] that in KDE the most relevant parameter is the bandwidth $\sigma$. A good choice of the bandwidth should depend on the amount of available data: the more the sampling the smaller the bandwidth. Thus we choose to shrink the bandwidth as the simulation proceeds, according to the popular Silverman’s rule of thumb [26]:

$$\sigma_i = \sigma_i^{(0)} [N_{\text{eff}}(d + 2)/4]^{-1/(d+4)},$$

(10)

where $\sigma_i^{(0)}$ is the standard deviation estimated from a short unbiased simulation, $d$ is the dimensionality of the CV space, and $N_{\text{eff}} = (\sum_k w_k)^2 / \sum_k w_k^2$ is the effective sample size. The KDE literature presents many other promising bandwidth selection rules, but we leave them for future investigation.

The number of kernels accumulated during the simulation quickly becomes very large and summing all of them at each time step is prohibitive. To avoid this problem we implement a simple on-the-fly kernel compression algorithm [25], that allows for the insertion of new kernels only in newly explored regions, otherwise merges them with existing ones. In the supplemental material we discuss this choice in detail, and we show the advantages over the more common approach of storing the bias on a grid [27].

Our choice of the probability estimator aims at quickly obtaining a coarse representation of the FES, and then slowly converging the finer details. This allows for a faster exploration of the relevant metastable states, which is one of the critical issues in AUS-like approaches [27]. Despite this improvement, the exploratory phase can still be challenging, especially when a suboptimal CV drives the system towards secondary unimportant states, or simply when a large number of CVs is used (see examples in [SM]). One way of addressing this issue is to extend the exploration time via an ad hoc parameter, but ultimately this only shifts the problem to the user, which usually can guess a good value for such parameter only after some trial and error. Here we prefer to adopt a different strategy, and instead of complicating the method, we introduce a different variant, designed to focus on the rapid exploration of the CV space.

This second variant makes use of Eq. [7] to update the bias, thus is based on an estimate of the biased target probability. To build such estimate we use KDE as in the convergence variant (Eq. [8]), but without the need for the reweighting weights, thus:

$$\hat{p}_n^{WT}(s) = \frac{\sum_n G(s, s_k) + \epsilon}{(n + 1)}.$$  

(11)

The choice of the parameters is identical to the previous case, apart from some trivial rescaling (see [SM]). This facilitates the typical workflow of enhanced sampling practitioners, who generally always perform an exploratory run before proceeding to a convergence run.

This second variant is more explorative because it is directly based on the sampled distribution, thus it is quick to realize the presence of a new metastable minimum and adapt to it. However, contrary e.g. to non-tempered ($\gamma = \infty$) MetaD, it diminishes the bias variation in time and eventually does converge. In the supplemental material we quantify the exploration speed of the two variants in a simple toy model, and compare them with well-tempered and non-tempered MetaD.

Due to the broader features of the target distribution, fewer kernels are generally needed in this exploration variant, compared to the convergence one. As a result the direct FES estimate is generally more noisy, but a simple post-processing reweighting procedure (Eq. [1]) can be carried, in order to retrieve a more precise estimate.

We implemented the new method, called on-the-fly probability enhanced sampling (OPES), in the enhanced sampling library PLUMED [29] and tested it on a variety of different systems. The code and all the files needed to reproduce the simulations are openly available in the PLUMED-NEST website [30], as plumID:19.068. Here we only present the results obtained on the benchmark system of alanine dipeptide in vacuum, using both the convergence variant (OPES-c) and the exploration (OPES-e) variant. In the supplemental material we also apply the method to two different Langevin toy models, based on 2D potentials that we bias using only a suboptimal CV. We also show the results obtained by biasing alanine tetrapeptide, using 3 or 6 dihedral angles as CVs, and the solid-liquid transition of sodium.

Alanine dipeptide is a small molecule with two main metastable basins well described by the two Ramachandran angles $\phi$ and $\psi$, which are nearly optimal CVs for this system. Comparing different enhanced sampling methods is always non-trivial, and one should not draw too broad a conclusion from just a few examples. But in order to give a better idea of our new method, in Fig. [1] and [2] we compare it with a typical well-tempered meta-dynamics run. We did not try to optimize manually the MetaD or the OPES input parameters, but rather we wanted to test them in an agnostic fashion, using very standard input values. For displaying purposes we show only one representative simulation, but we ran multiple independent replicas, and the full results can be found in the [SM] together with all the computational details. It can be seen how OPES-c has a steady convergence, and
FIG. 1. The φ trajectory for alanine dipeptide, obtained by biasing the two dihedral angles, φ and ψ, using MetaD and the two variants of OPES. The same bias factor γ = 10 and the same initial conditions are used in the three simulations.

FIG. 2. Time evolution of the free energy difference between the two basins of alanine dipeptide, obtained from the same simulations shown in Fig. 1. The reference blue stripe is 1 k_B T thick.

OPES-e provides an extremely quick exploration. The evolution of the free energy estimate for the convergence variant is shown in Fig. 3. A similar figure for the exploration variant can be found in the SM.

One strength of the method proposed here is the fact that it requires a minimal set of input parameters from the user. In the above simulations we only need to specify three quantities: the pace at which the bias is updated, the initial bandwidth σ of the Gaussian kernels, and the approximate height of the barriers we wish to cross. In our tests we always keep the deposition pace equal to the one used in MetaD, for the alanine system that is 500 simulation steps (1 ps). The initial bandwidth should be equal to the smaller standard deviation of the CVs in the minima, which can be measured in a short unbiased run. The choice of the barrier ΔE requires a minimal knowledge of the system under study, but only a vague idea is usually enough. This parameter is used to set both the value of the regularization factor ϵ and the bias factor γ. The choice of γ is not as critical as in MetaD, since here it does not directly influence the convergence speed. In the convergence variant the free energy is explored up to this barrier value and only a little more (see Fig. 3), so that no computational resources are wasted in highly unlikely regions. In the exploration variant instead, higher regions are also sampled, and the barrier parameter ΔE decides how aggressive the exploration will be.

In conclusion in this paper we present a new enhanced sampling method, based on a on-the-fly reconstruction of the probability distribution. It combines ideas from adaptive umbrella sampling and metadynam-
ics, and builds a bias potential through a self consistent procedure. This method provides a general framework, in which different target distributions could be implemented. We focused in this work on the well-tempered distribution, and showed how it allows for two different variants of the method, one aiming at convergence and the other at exploration, but both capable of either tasks. The method uses KDE with an on-the-fly kernel merging algorithm for the probability density estimation, which allows keeping the number of input parameters needed from the user to a bare minimum.

We believe this new method can become a handy tool in addressing enhanced sampling problems, and has the potential of opening to further interesting developments.

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. Calculations were carried out on Euler cluster at ETH Zurich and are openly available in the Materials Cloud Archive (www.materialscloud.org) with ID materialscloud:2019.0063.

The authors thank Sandro Bottaro and Barak Hirshberg for useful feedback on the manuscript. M.I. thanks Dario Azzimonti and Manuel Schuerch for drawing his attention to kernel compression algorithms.

* parrinello@phys.chem.ethz.ch

[1] B. Peters, Reaction rate theory and rare events (Elsevier, 2017) p. 619.
[2] G. Torrie and J. Valleau, Journal of Computational Physics 23, 187 (1977).
[3] M. Mezei, Journal of Computational Physics 68, 237 (1987).
[4] P. Maragakis, A. van der Vaart, and M. Karplus, The Journal of Physical Chemistry B 113, 4664 (2009).
[5] S. Marsili, A. Barducci, R. Chelli, P. Procacci, and V. Schettino, The Journal of Physical Chemistry B 110, 14011 (2006).
[6] B. M. Dickson, F. Legoll, T. Lelièvre, G. Stoltz, and P. Fleurat-Lessard, The Journal of Physical Chemistry B 114, 5823 (2010).
[7] A. Laio and M. Parrinello, Proceedings of the National Academy of Sciences 99, 12562 (2002).
[8] A. Barducci, G. Bussi, and M. Parrinello, Physical Review Letters 100, 020603 (2008) arXiv:0803.3861.
[9] D. Branduardi, G. Bussi, and M. Parrinello, Journal of Chemical Theory and Computation 8, 2247 (2012) arXiv:1205.4300.
[10] T. Lelièvre, M. Rousset, and G. Stoltz, Free Energy Computations (IMPERIAL COLLEGE PRESS, 2010).
[11] J. F. Dama, M. Parrinello, and G. A. Voth, Physical review letters 112, 240602 (2014).
[12] O. Valsson, P. Tiwary, and M. Parrinello, Annual Review of Physical Chemistry 67, 159 (2016).
[13] G. Bussi, A. Laio, and P. Tiwary, in Handbook of Materials Modeling (Springer International Publishing, Cham, 2018) pp. 1–31.
[14] A. D. White, J. F. Dama, and G. A. Voth, Journal of Chemical Theory and Computation 11, 2451 (2015).
[15] P. Shaffer, O. Valsson, and M. Parrinello, Proceedings of the National Academy of Sciences 113, 1150 (2016).
[16] M. Invernizzi, O. Valsson, and M. Parrinello, Proceedings of the National Academy of Sciences 114, 3370 (2017).
[17] J. Debnath, M. Invernizzi, and M. Parrinello, Journal of Chemical Theory and Computation 15, 2454 (2019).
[18] P. M. Piaggi and M. Parrinello, Physical Review Letters 122, 050601 (2019).
[19] O. Valsson and M. Parrinello, Physical Review Letters 113, 090601 (2014) arXiv:1407.0477.
[20] M. Invernizzi and M. Parrinello, Journal of Chemical Theory and Computation 15, 2157 (2019).
[21] P. Tiwary and M. Parrinello, Journal of Physical Chemistry B 119, 736 (2015).
[22] L. Donati and B. G. Keller, The Journal of Chemical Physics 149, 072335 (2018).
[23] V. Marinova and M. Salvalaglio, The Journal of Chemical Physics 151, 164115 (2019).
[24] P. Giberti, B. Cheng, G. A. Tribello, and M. Ceriotti, (2019) arXiv:1911.01140.
[25] O. Valsson and M. Parrinello, Journal of Chemical Theory and Computation 11, 1996 (2015) arXiv:1407.0477.
[26] B. Silverman, Density Estimation for Statistics and Data Analysis (Routledge, New York, 1998).
[27] G. Fort, B. Jourdain, T. Lelièvre, and G. Stoltz, Statistics and Computing 27, 147 (2017).
[28] D. Sokkombham, D. Ciliberti, M. A. Wilson, K.-I. Fukui, K. Moriyama, M. Numao, and F. Kloosterman, Knowledge-Based Systems 94, 1 (2016).
[29] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, Computer Physics Communications 185, 604 (2014).
[30] The PLUMED consortium, Nature Methods 16, 670 (2019).