1 Convergence Rate Analysis

Here we explain the critical slowing down observed in Section 3.1 of the paper. We perform this analysis under the following conditions: we assume the target density is constant, \( \rho(x) = \rho \), use Gaussian kernels, \( \Phi(x) = e^{-|x|^2} \), restrict to \( \gamma = 2 \), i.e., \( L_2 \) energy,

\[
E(\{x_i\}_{i=1}^n) = \int (A(x) - \rho(x))^2 dx,
\]

and assume very small time steps, \( \Delta t \ll 1 \) in the Langevin update

\[
x_j^{k+1} = x_j^k - \frac{\Delta t}{|A|} \nabla x_j E(X^k) + \sqrt{2\Delta t} \xi_j.
\]

We also assume a periodic domain and that the current point configuration already possess blue noise spectrum, i.e., it is near equilibrium at reasonably low temperature. The analysis is performed by studying the evolution of the approximation function,

\[
\hat{A}^k(x) = \sum_{j=1}^n \frac{1}{\sigma^d} \Phi \left( \frac{|x - x_j^k|}{\sigma} \right) = \Phi(x) * \sum_{j=1}^n \delta(x - x_j^k),
\]

where we abuse notations and use, from now on and on, \( \Phi(x) = \sigma^{-d} \Phi(\|x\|/\sigma) \) in Fourier space, \( A \) becomes

\[
\hat{A}^k(\omega) = \hat{\Phi}(\omega) \sum_{j=1}^n e^{-i \omega \cdot x_j^k},
\]

Since we assume \( \Delta t \) is infinitesimally small, we can skip the Metropolis-Hasting correction step. We also assume that all the state coordinates \( x_j \) are updated at every step \( k \). In this case,

\[
\hat{A}^{k+1}(\omega) = \hat{\Phi}(\omega) \sum_{j=1}^n e^{-i \omega \cdot x_j^{k+1}} = \hat{\Phi}(\omega) \sum_{j=1}^n e^{-i \omega \cdot (x_j^k + z_j^k)}
\]

\[
\approx \hat{\Phi}(\omega) \sum_{j=1}^n e^{-i \omega \cdot x_j^k} (1 - i \omega \cdot z_j^k)
\]

\[
= \hat{A}^k(\omega) - i \hat{\Phi}(\omega) \omega \sum_{j=1}^n e^{-i \omega \cdot x_j^k} z_j^k,
\]

where \( z_j^k = \sqrt{\Delta t} \xi_j^k - A(x_j) \nabla x_j E(X^k) \) according to (2). In (4) we used a first-order Taylor approximation that relies on \( \Delta t \) being small. The energy gradient in \( x_j \) is given by

\[
\nabla x_j E(X^k) = 2 \int_{\Omega} \nabla x_j \Phi(x - x_j^k) (A(x) - \rho) dx,
\]

and becomes in Fourier space, according to Plancherel theorem,

\[
2 \int \left( -i \omega e^{-i \omega \cdot x_j^k} \hat{\Phi}(\omega) \right)^* (\hat{A}(\omega) - \delta(\omega)) d\omega,
\]

\[
= 2i \int \theta^* e^{i \omega \cdot x_j^k} \hat{\Phi}(\omega) (\hat{A}(\omega) - \delta(\omega)) d\omega,
\]

where \( \delta(\omega) = \delta(\omega) \rho \). By plugging this in (4), we get

\[
\hat{A}^{k+1}(\omega) = \hat{A}^k(\omega) - i \sqrt{\Delta t} \hat{\Phi}(\omega) \omega \sum_{j=1}^n e^{-i \omega \cdot x_j^k} \xi_j^k \]

\[
- \omega \cdot \frac{\Delta t}{T} \hat{\Phi}(\omega) \sum_{j=1}^n e^{-i \omega \cdot x_j^k} \theta^* e^{i \omega \cdot x_j^k} \hat{\Phi}(\omega) (\hat{A}(\omega) - \delta(\omega)) d\omega.
\]

The last term equal to

\[
\omega \cdot \frac{\Delta t}{T} \int \theta^* \sum_{j=1}^n e^{-i(\omega - \theta) \cdot x_j^k} \hat{\Phi}(\omega) \hat{\Phi}(\omega) (\hat{A}(\omega) - \delta(\omega)) d\omega
\]

\[
= \omega \Delta t \int \theta^* \hat{A}^k(\omega - \theta) e^{-\sigma^2(\omega^2 + \omega' \omega'')/4} (\hat{A}(\omega) - \delta(\omega)) d\omega,
\]

since

\[
\hat{\Phi}(\omega) \hat{\Phi}(\omega) = e^{-\sigma^2(\omega^2 + \omega' \omega'')/4} \hat{\Phi}(\omega - \theta) e^{-\sigma^2(\omega^2 + \omega' \omega'')/4}.
\]

Assuming \( X^k \) is near equilibrium then, at reasonably low temperature, the power spectrum of the point configuration shows the blue-noise properties, i.e., it has a delta at the origin and vanishes in an annulus around it. The approximation function \( \hat{A}(\omega) \) equals to this function times \( \hat{\Phi}(\omega) \) which is zero outside the angulus and hence \( \hat{A}(\omega) \approx \delta(\omega) \hat{\Phi}(\omega) \). Therefore, the term in (8) is approximately

\[
\|\omega\|^2 \frac{\Delta t}{T} e^{-\sigma^2(\omega^2 + \omega' \omega'')/4} (\hat{A}(\omega) - \delta(\omega)) \hat{\Phi}(\omega) \hat{\Phi}(\omega)
\]

\[
= \|\omega\|^2 \frac{\Delta t}{T} e^{-\sigma^2(\omega^2 + \omega' \omega'')/4} (\hat{A}(\omega) - \delta(\omega)) \hat{\Phi}(\omega) \hat{\Phi}(\omega),
\]

For small \( \|\omega\| \), the factor \( e^{-\sigma^2(\omega^2 + \omega' \omega'')/4} \) is close to one and therefore, the evolution equation of \( A^k \) can be modeled by

\[
\hat{A}(\omega)^{k+1} = \hat{A}(\omega)^k - \|\omega\|^2 \frac{\Delta t}{T} \hat{A}(\omega)^k - \delta(\omega) \hat{\Phi}(\omega) \hat{\Phi}(\omega)
\]

\[
- \sqrt{\Delta t} \|\omega\| \xi(\omega)^k,
\]

where the noise term \( \xi(\omega)^k = i \omega/\|\omega\| \cdot \sum_{j=1}^n e^{-i \omega \cdot z_j^k} \) is a linear combination of independent normal random variables \( \xi_j^k \) with weights of unit amplitude \( e^{-i \omega \cdot z_j^k} \).

To ease notations in what follows, we assume that \( \hat{A} \) is normalized such that \( \hat{A}(0) = 1 \) by redefining \( \hat{A}(\omega) = \hat{A}(\omega) \hat{\Phi}(\omega) \hat{\Phi}(\omega) \). Note that this normalization does not intervene with the recursive relation we investigated so far since \( \hat{A}(0) \) is constant and independent of \( k \) as the number of kernels \( n \), and their scale \( \sigma \) do not change with \( k \). In this case the relation in (11) becomes, for every \( \omega \neq 0 \),

\[
\hat{A}(\omega)^{k+1} = \hat{A}(\omega)^k (1 - \frac{\Delta t \|\omega\|^2}{T}) + \sqrt{\Delta t} \|\omega\| \xi(\omega)^k,
\]

where we abuse notations again and denote the added noise by \( \xi(\omega)^k \) (to which we absorb the unit-circle factor \( \|\omega\| \)).

Unfolding this recursion yields

\[
\hat{A}(\omega)^k = \sqrt{\Delta t} \|\omega\| \sum_{l=1}^k \xi(\omega)^k (1 - \frac{\Delta t \|\omega\|^2}{T})^{k-l} + (1 - \frac{\Delta t \|\omega\|^2}{T})^k \hat{A}(\omega)^0.
\]
Since $\hat{A}(\omega)^k$ is a linear combination of normal random variables, we conclude that it is itself a normal variable. Therefore, we can measure the convergence of its distribution by examining its first two moments. Equation (12) is a discretized stochastic differential equation, therefore we can consider a 'physical' time $\tau$, such that $k = \tau / \Delta t$ and by allowing $\Delta t \to 0$ (and hence $k \to \infty$) we get

$$\mathbb{E}[\hat{A}(\omega)^k] = \left(\frac{1 - \frac{\Delta t\|\omega\|^2}{T}}{T}\right) \hat{A}(\omega)^0 \xrightarrow{\Delta t \to 0} e^{-\tau\|\omega\|^2/T} \hat{A}(\omega)^0,$$

which results from $\mathbb{E}[\xi] = 0$, and

$$\mathbb{V}[\hat{A}(\omega)^k] = \Delta t\|\omega\|^2 \sum_{l=1}^{k} \left(1 - \frac{\Delta t\|\omega\|^2}{T}\right)^{l} \mathbb{V}[\xi] \approx \Delta t\|\omega\|^2 \sum_{l=1}^{k} l! \left(1 - \frac{2\Delta t\|\omega\|^2}{T}\right)^{l} \mathbb{V}[\xi] = \Delta t\|\omega\|^2 \frac{1 - (1 - \frac{2\Delta t\|\omega\|^2}{T})^k}{2\Delta t\|\omega\|^2} \mathbb{V}[\xi] \xrightarrow{\Delta t \to 0} T \left(1 - e^{-\tau\|\omega\|^2/T}\right) \mathbb{V}[\xi].$$

From this analysis we conclude that the moments of $P(\hat{A}(\omega))$ converge to their equilibrium values at rate $\tau = O(\|\omega\|^{-2})$. The predicted number of Markov-chain steps $k$, needed to achieve stochastic relaxation, is therefore dictated by the lowest mode, $2\pi/m$ (along one axis). Hence, assuming a fixed small time step $\Delta t$, the number of required chain steps

$$k = \tau / \Delta t = O((2\pi/m)^{-2}) / \Delta t = O(m^2),$$

or equivalently $O(n)$ in two-dimensional space.

This theoretical finding is in agreement with the empirical results presented in Section 3.1 of the paper. This analysis also offers an intuitive explanation for this behavior. The product $\|\omega\|^2 = \omega^* \cdot \omega$ results from two reasons. In (4) the effect of moving the points independently, by $\mathbf{z}_j$, affects $\hat{A}$ like $O(\|\omega\|)$. This makes sense because as the frequency decreases, the response to small point movement becomes small (e.g., it is not felt at all at $\omega = 0$). The second reason follows from the fact that the energy gradient gradient, in (5), is $O(\|\omega\|)$, meaning that the energy is less sensitive to points movements at lower frequencies. These effects are illustrated in Figure 1.

2 Refinement Spectral Accuracy

Here we estimate the spectral agreement between the approximation function $A_C$ corresponding to $\{\mathbf{y}_j\}_{j=1}^{n_C}$ and $\hat{A}$ that results from the refined configuration $\{\mathbf{x}_j\}_{j=1}^{n_R}$ obtained from the refinement scheme we described in Section 3.2 of the paper. This refinement consists of splitting every point $\mathbf{y}_j$ into $2d$ points, given by $2\mathbf{y}_j + \mathbf{z}_j$, where $\mathbf{z}_j$ for $1 \leq l \leq 2^d$ are the offset vectors that sum to zero. In this derivation we assume small $\omega$ and use the Taylor approximation $e^{i\omega \mathbf{z}_j} \approx 1 - i\omega \cdot \mathbf{z}_j$. We also assume and use the following admissibility condition on the kernel functions

$$\hat{\Phi}'(0) = 0 \Rightarrow \hat{\Phi}(\|\omega\|) = \hat{\Phi}(0) + O(\|\omega\|^2),$$

which is obeyed by any symmetric function that is differentiable at zero, such the Gaussian kernel function we use. We use $\sigma_j$ to denote $\sigma(\mathbf{y}_j)$ at the coarse scale and $\sigma(2\mathbf{y}_j + \mathbf{z}_j)$ at the fine scale, assuming $\sigma$ does not change by much between scales and along small perturbations. Thus, for small $\omega$ we get

$$\hat{A}(\omega) = \sum_{j=1}^{n_C} \sum_{l=1}^{2^d} \hat{\Phi}(\|\mathbf{z}_j/2\|/e^{-i\omega \cdot (2\mathbf{y}_j + \mathbf{z}_j)})
\approx \sum_{j=1}^{n_C} \sum_{l=1}^{2^d} \hat{\Phi}(0) e^{-i\omega \cdot 2\mathbf{y}_j} (1 - i\omega \cdot \mathbf{z}_j) + O(\|\omega\|^2)
= 2^d \hat{A}_C(2\omega) + O(\|\omega\|^2),$$

where the second term in the line before the last disappeared since $\sum_{l=1}^{2^d} \mathbf{z}_j = 0.$