Analysis for the Blue-Noise Point Sampling using Kernel Density Model

Raanan Fattal*
Hebrew University of Jerusalem, Israel

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, $\rho(\mathbf{x}) = \rho$, use Gaussian kernels, $\Phi(\mathbf{x}) = e^{-\|\mathbf{x}\|^2}$, restrict to $\gamma = 2$, i.e., L_2 energy,

$$E\big(\{\mathbf{x}_j\}_{j=1}^n\big) = \int_\Omega \left(A(\mathbf{x}) - \rho(\mathbf{x})\right)^2 d\mathbf{x}, \tag{1}$$

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

$$\mathbf{x}_j^{k+1} = \mathbf{x}_j^k - \frac{\Delta t}{2T}\nabla_{\mathbf{x}_j} E(\mathbf{X}^k) + \sqrt{\Delta t_j}\boldsymbol{\xi}_j^k. \tag{2}$$

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,

$$A^k(\mathbf{x}) = \sum_{j=1}^n \frac{1}{\sigma^d}\Phi\left(\frac{\|\mathbf{x}-\mathbf{x}_j^k\|}{\sigma}\right) = \Phi(\mathbf{x}) * \sum_{j=1}^n \delta(\mathbf{x}-\mathbf{x}_j^k),$$

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

$$\hat{A}^k(\boldsymbol{\omega}) = \hat{\Phi}(\boldsymbol{\omega})\sum_{j=1}^n e^{-i\boldsymbol{\omega}\cdot\mathbf{x}_j^k}. \tag{3}$$

Since we assume Δt is infinitesimally small, we can skip the Metropolis-Hasting correction step. We also assume that all the state coordinates $\mathbf{x}_j^k$ are updated at every step k. In this case,

$$\begin{aligned}\hat{A}^{k+1}(\boldsymbol{\omega}) &= \hat{\Phi}(\boldsymbol{\omega})\sum_{j=1}^n e^{-i\boldsymbol{\omega}\cdot\mathbf{x}_j^{k+1}} = \hat{\Phi}(\boldsymbol{\omega})\sum_{j=1}^n e^{-i\boldsymbol{\omega}\cdot(\mathbf{x}_j^k+\mathbf{z}_j^k)} \\ &\approx \hat{\Phi}(\boldsymbol{\omega})\sum_{j=1}^n e^{-i\boldsymbol{\omega}\cdot\mathbf{x}_j^k}(1 - i\boldsymbol{\omega}\cdot\mathbf{z}_j^k) \\ &= \hat{A}^k(\boldsymbol{\omega}) - i\hat{\Phi}(\boldsymbol{\omega})\boldsymbol{\omega}\cdot\sum_{j=1}^n e^{-i\boldsymbol{\omega}\cdot\mathbf{x}_j^k}\mathbf{z}_j^k\end{aligned} \tag{4}$$

where $\mathbf{z}_j^k = \sqrt{\Delta t}\boldsymbol{\xi}_j^k - \frac{\Delta t}{2T}\nabla_{\mathbf{x}_j}E(\mathbf{X}^k)$ according to (2). In (4) we used a first-order Taylor approximation that relies on Δt being small. The energy gradient in $\mathbf{z}_j$ is given by

$$\nabla_{\mathbf{x}_j}E(\mathbf{X}^k) = 2\int_\Omega \nabla_{\mathbf{x}_j}\Phi(\mathbf{x}-\mathbf{x}_j^k)\big(A(\mathbf{x})-\rho\big)d\mathbf{x}, \tag{5}$$

and becomes in Fourier space, according to Plancherel theorem,

$$\begin{aligned}&2\int\left(-i\boldsymbol{\theta}e^{-i\boldsymbol{\theta}\cdot\mathbf{x}_j^k}\hat{\Phi}(\boldsymbol{\theta})\right)^*\left(\hat{A}(\boldsymbol{\theta})-\delta_\rho(\boldsymbol{\theta})\right)d\boldsymbol{\theta}, \\ &= 2i\int\boldsymbol{\theta}^* e^{i\boldsymbol{\theta}\cdot\mathbf{x}_j^k}\hat{\Phi}(-\boldsymbol{\theta})\big(\hat{A}(\boldsymbol{\theta})-\delta_\rho(\boldsymbol{\theta})\big)d\boldsymbol{\theta},\end{aligned} \tag{6}$$

where $\delta_\rho(\boldsymbol{\omega}) = \delta(\boldsymbol{\omega})\rho$. By plugging this in (4), we get

$$\begin{aligned}\hat{A}^{k+1}(\boldsymbol{\omega}) &= \hat{A}^k(\boldsymbol{\omega}) - i\sqrt{\Delta t}\hat{\Phi}(\boldsymbol{\omega})\,\boldsymbol{\omega}\cdot\sum_{j=1}^n e^{-i\boldsymbol{\omega}\cdot\mathbf{x}_j^k}\boldsymbol{\xi}_j^k \\ &-\boldsymbol{\omega}\cdot\frac{\Delta t}{T}\hat{\Phi}(\boldsymbol{\omega})\sum_{j=1}^n e^{-i\boldsymbol{\omega}\cdot\mathbf{x}_j^k}\int\boldsymbol{\theta}^* e^{i\boldsymbol{\theta}\cdot\mathbf{x}_j^k}\hat{\Phi}(-\boldsymbol{\theta})\big(\hat{A}(\boldsymbol{\theta})-\delta_\rho(\boldsymbol{\theta})\big)d\boldsymbol{\theta}.\end{aligned} \tag{7}$$

The last term equal to

$$\begin{aligned}&\boldsymbol{\omega}\cdot\frac{\Delta t}{T}\int\boldsymbol{\theta}^*\sum_{j=1}^n e^{-i(\boldsymbol{\omega}-\boldsymbol{\theta})\cdot\mathbf{x}_j^k}\hat{\Phi}(\boldsymbol{\omega})\hat{\Phi}(-\boldsymbol{\theta})\big(\hat{A}^k(\boldsymbol{\theta})-\delta_\rho(\boldsymbol{\theta})\big)d\boldsymbol{\theta} \\ &=\boldsymbol{\omega}\cdot\frac{\Delta t}{T}\int\boldsymbol{\theta}^*\hat{A}^k(\boldsymbol{\omega}-\boldsymbol{\theta})e^{-\sigma^2(\boldsymbol{\omega}\cdot\boldsymbol{\theta}^*+\boldsymbol{\omega}^*\cdot\boldsymbol{\theta})/4}\big(\hat{A}^k(\boldsymbol{\theta})-\delta_\rho(\boldsymbol{\theta})\big)d\boldsymbol{\theta},\end{aligned} \tag{8}$$

since

$$\hat{\Phi}(\boldsymbol{\omega})\hat{\Phi}(-\boldsymbol{\theta}) = e^{-\sigma^2(\|\boldsymbol{\omega}\|^2+\|\boldsymbol{\theta}\|^2)/4} = \hat{\Phi}(\boldsymbol{\omega}-\boldsymbol{\theta})e^{-\sigma^2(\boldsymbol{\omega}\cdot\boldsymbol{\theta}^*+\boldsymbol{\omega}^*\boldsymbol{\theta})/4}. \tag{9}$$

Assuming $\mathbf{X}^k$ is near equilibrium then, at reasonably low temperature, the power spectrum of the point configuration shows the blue-noise properties, i.e., has a delta at the origin and vanishes in an annulus around it. The approximation function $\hat{A}(\boldsymbol{\omega})$ equals to this function times $\hat{\Phi}(\boldsymbol{\omega})$ which is zero outside the angulus and hence $\hat{A}(\boldsymbol{\omega}) \approx \delta(\boldsymbol{\omega})\hat{A}(\mathbf{0})$. Therefore, the term in (8) is approximately

$$\|\boldsymbol{\omega}\|^2\frac{\Delta t}{T}e^{-\sigma^2\|\boldsymbol{\omega}\|^2/2}\big(\hat{A}^k(\boldsymbol{\omega})-\delta_\rho(\boldsymbol{\omega})\big)\hat{A}(\mathbf{0}). \tag{10}$$

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

$$\begin{aligned}\hat{A}(\boldsymbol{\omega})^{k+1} &= \hat{A}(\boldsymbol{\omega})^k - \|\boldsymbol{\omega}\|^2\frac{\Delta t}{T}\big(\hat{A}^k(\boldsymbol{\omega})-\delta_\rho(\boldsymbol{\omega})\big)\hat{A}(\mathbf{0}) \\ &\quad - \sqrt{\Delta t}\|\boldsymbol{\omega}\|\hat{\xi}(\boldsymbol{\omega})^k,\end{aligned} \tag{11}$$

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

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

$$\hat{A}(\boldsymbol{\omega})^{k+1} = \hat{A}(\boldsymbol{\omega})^k\left(1-\frac{\Delta t\|\boldsymbol{\omega}\|^2}{T}\right) + \sqrt{\Delta t}\|\boldsymbol{\omega}\|\hat{\xi}^k, \tag{12}$$

where we abuse notations again and denote the added noise by $\hat{\xi}^k$ (to which we absorb the unit-circle factor $i\boldsymbol{\omega}/\|\boldsymbol{\omega}\|$), and disregard the fact that noise is correlated among different frequencies. Unfolding this recursion yields

$$\hat{A}(\boldsymbol{\omega})^k = \sqrt{\Delta t}\|\boldsymbol{\omega}\|\sum_{l=1}^k\hat{\xi}^l\left(1-\frac{\Delta t\|\boldsymbol{\omega}\|^2}{T}\right)^{k-l} + \left(1-\frac{\Delta t\|\boldsymbol{\omega}\|^2}{T}\right)^k\hat{A}(\boldsymbol{\omega})^0 \tag{13}$$

*e-mail: raananf@cs.huji.ac.il

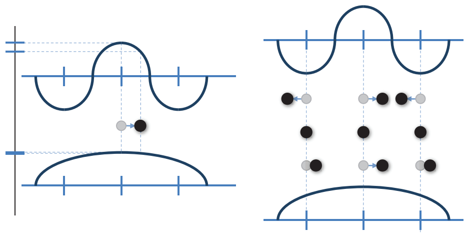

Figure 1: *(left) Show the effect a displacement of a single point has on low and high frequency modes. The high-frequency mode shows a greater change in response than the low-frequency mode. (right) The effect of low- and high-frequency displacements, of an equal magnitude, have on the inter-particle distance. Positive values correspond to right offsets and negative to left ones. The high-frequency offset has a greater effect over the inter-point distances and violates more the local point density.*

Since $\hat{A}(\boldsymbol{\omega})^k$ 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 τ, such that $k = \tau/\Delta t$ and by allowing $\Delta t \to 0$ (and hence $k \to \infty$) we get

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

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

$$\begin{aligned}\mathbf{V}\big[\hat{A}(\boldsymbol{\omega})^k\big] &= \Delta t\|\boldsymbol{\omega}\|^2 \sum_{l=1}^{k}\Big(1-\frac{\Delta t\|\boldsymbol{\omega}\|^2}{T}\Big)^{2(k-l)}\mathbf{V}[\hat{\xi}] \\ &\approx \Delta t\|\boldsymbol{\omega}\|^2 \sum_{l=1}^{k}\Big(1-\frac{2\Delta t\|\boldsymbol{\omega}\|^2}{T}\Big)^{k-l}\mathbf{V}[\hat{\xi}] \\ &= \Delta t\|\boldsymbol{\omega}\|^2 T\frac{1-\big(1-\frac{2\Delta t\|\boldsymbol{\omega}\|^2}{T}\big)^k}{2\Delta t\|\boldsymbol{\omega}\|^2}\mathbf{V}[\hat{\xi}] \xrightarrow[\Delta t\to 0]{} \frac{T}{2}(1-e^{-\tau 2\|\boldsymbol{\omega}\|^2/T})\mathbf{V}[\hat{\xi}].\end{aligned}$$

From this analysis we conclude that the moments of $P(\hat{A}(\boldsymbol{\omega}))$ converge to their equilibrium values at rate $\tau = O(\|\boldsymbol{\omega}\|^{-2})$. 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 Δt, the number of required chain steps

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

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 $\|\boldsymbol{\omega}\|^2 = \boldsymbol{\omega}^* \cdot \boldsymbol{\omega}$ results from two reasons. In (4) the effect of moving the points independently, by $\mathbf{z}_j$, affects $\hat{A}$ like $O(\|\boldsymbol{\omega}\|)$. 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 $\boldsymbol{\omega} = \mathbf{0}$). The second reason follows from the fact that the energy gradient gradient, in (5), is $O(\|\boldsymbol{\omega}\|)$, 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 $\hat{A}_{\mathcal{C}}$ corresponding to $\{\mathbf{y}_j\}_{j=1}^{n_{\mathcal{C}}}$ and $\hat{A}$ that results from the refined configuration $\{\mathbf{x}_j\}_{j=1}^{n}$ 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 2^d points, given by $2\mathbf{y}_j + \mathbf{z}_j^l$, where $\mathbf{z}^l$ for $1 \le l \le 2^d$, are the offset vectors that sum to zero. In this derivation we assume small $\boldsymbol{\omega}$ and use the Taylor approximation $e^{\boldsymbol{\omega}\cdot\mathbf{z}_j} \approx 1 - i\boldsymbol{\omega}\cdot\mathbf{z}_j$. We also assume and use the following *admissibility* condition on the kernel functions

$$\hat{\Phi}'(0) = 0 \;\Rightarrow\; \hat{\Phi}(\|\boldsymbol{\omega}\|) = \hat{\Phi}(0) + O\big(\|\boldsymbol{\omega}\|^2\big), \tag{14}$$

which is obeyed by any symmetric function that is differentiable at zero, such the Gaussian kernel function we use. We use σ_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 σ does not change by much between scales and along small perturbations. Thus, for small $\boldsymbol{\omega}$ we get

$$\begin{aligned}\hat{A}(\boldsymbol{\omega}) &= \sum_{j=1}^{n_{\mathcal{C}}}\sum_{l=1}^{2^d}\hat{\Phi}\big(\|\boldsymbol{\omega}\|\sigma_j/2\big)e^{-i\boldsymbol{\omega}\cdot(2\mathbf{y}_j+\mathbf{z}_j^l)} \\ &= \sum_{j=1}^{n_{\mathcal{C}}}\sum_{l=1}^{2^d}\hat{\Phi}(0)e^{-i2\boldsymbol{\omega}\cdot\mathbf{y}_j}(1-i\boldsymbol{\omega}\cdot\mathbf{z}_j^l) + O(\|\boldsymbol{\omega}\|^2) \\ &= 2^d\sum_{j=1}^{n_{\mathcal{C}}}\hat{\Phi}(0)e^{-i2\boldsymbol{\omega}\cdot\mathbf{y}_j} - i\boldsymbol{\omega}\cdot\sum_{l=1}^{2^d}\mathbf{z}_j^l\sum_{j=1}^{n_{\mathcal{C}}}\hat{\Phi}(0) + O(\|\boldsymbol{\omega}\|^2) \\ &= 2^d\hat{A}_{\mathcal{C}}(2\boldsymbol{\omega}) + O(\|\boldsymbol{\omega}\|^2),\end{aligned}$$

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