

NOTES ON GRAPH UQ SCALING

Abstract. The purpose of this paper is to investigate the scaling of pCN and Gibbs sampler with respect to the number of data points under the assumption that the data points are sampled i.i.d from a fixed distribution $\mathbb{P}(X)$. The problem has both a continuum and a discrete version. On the discrete side, one could study scaling limits. On the continuum side, we can 1) formulate the continuum problem under $N \rightarrow \infty$, and $\epsilon \rightarrow 0$. 2) Study/prove whether pCN or Gibbs is well-defined in the limit.

1. Introduction. Graph semi-supervised learning has many applications xxx. We state the problem setting for graph SSL. Let $Z = \{1 \dots N\}$, and $Z' \subset Z$. $\{Y_j \in \{-1, 1\} | j \in Z'\}$ be the labels. The task is to infer the missing labels via the graph Laplacian L . The optimization approach is to minimize a data fidelity term with regularization by q quadratic form $\langle u, Lu, \rangle$.

In [?], this problem was formulated as a Bayesian inference problem, where the prior is a Gaussian $\mathcal{N}(0, C)$. Given some likelihood function $\exp(-\Phi(u; y, \theta))$, we can form the joint probability distribution as $\mathbb{P}(u|y) \propto \exp(-J(u))$, where

$$J(u) = \frac{1}{2} \langle u, Pu \rangle + \sum_{j \in Z'} \Phi(u(j); y(j), \theta).$$

In [?], the authors sampled the posterior distribution to compute quantities that reflect the uncertainty of the prediction and the system. The algorithm used was pCN, an MCMC algorithm designed for Bayesian inverse problems that scales well with dimensionality. The Auxiliary variable Gibbs method is a popular alternative used in [?] when the likelihood is probit. The goal of this paper is to study the scaling of these algorithms as $N \rightarrow \infty$, both theoretically and empirically. We need to specify how the feature vectors are obtained as the total number of points N large. the convention in [?] and assume the points $\{x_i\}$ are sampled iid from some distribution ρ supported on the domain D . We follow [?] and assume the fidelity Z' is finite. Moreover, to simplify the analysis, we assume the first J points of the i.i.d. samples $\{x_1, \dots, x_J\}$ are the fidelity samples.

In order to have a consistent limit, we need to scale the graph Laplacian. We also scale the in [4], so that the limiting dirichlet functional gamma converges to $\langle u, \mathcal{L}u \rangle$, where \mathcal{L} is a diffusion operator. The reason for choosing this scaling is that due to locality, the eigenvalues have asymptotic growth $O(j^{-d/2})$. Thus we can define a continuum Gaussian measure $\mu = N(0, (1 + \mathcal{L})^{-\alpha})$, and define a continuum version of the problem. In particular, if the eigenfunctions and eigenvalues of the continuous Laplacian operator satisfy

- λ_k has growth $N^{\frac{d}{2}}$
- Eigenfunctions ϕ_k are smooth up to boundary.
- $\sum_k k^{-\frac{\alpha d}{2}} \|\phi_k\|_{\infty} < \infty$.

(Andrew: Refer to pdeML paper for references on the exact assumptions made on the spectrum and eigenvalues, and how they are reasonable if ρ is smooth etc.)

The Gaussian is defined on the space of continuous function, and it is meaningful to evaluate u at the fidelity points.

We would hope that μ^N approximates μ meaningfully. However, due to the fact that the eigenvectors of the Graph Laplacian converges at different rates, the full operator doesn't necessarily converge, as shown in the numerical experiments. Instead, we separate the limits, and define μ_{NK} and μ_K independently. Namely, we approximate μ in two different limits. We call

Our contribution in this paper is listed below:

1. we study the discrete approximations to the continuum.
2. we prove a bound on pCN..
3. We study variations of the Gibbs method.
4. Numerical Experiments.

Finally, some important notations are listed below.

- μ_N, μ : The Gaussian prior on \mathbb{R}^N and on the continuum H^t .
- μ_K^N, μ_K : The degenerate measure induced by truncating the spectrum up to level K .
- π_N, π : The posterior. Assuming the $\pi \ll \mu$ and $\pi_N \ll \mu_0^N$.
- η as the joint probability of u, v under the posterior and the proposal $\mathbb{P}(v|u)$.
- D : The compact domain in \mathbb{R}^N on which the density function ρ is supported.
- C : The set on which the labels y are observed.

2. Discrete Approximation. To study precisely the discretization error from μ_{NK} to μ_K , we need study the measures from the perspective of the eigenfunctions q_k and q_k^N via the KL expansion. For notational simplicity, we assume that the spectrum of the continuous operator L has distinct eigenvalues. However it is straightforward to generalize to the general case. We summarize the convergence analysis Slepcev 17 in Proposition 1, which is proved in the appendix.

PROPOSITION 1. *Let ρ be the density function defined on the domain D . Let \mathcal{L} be the continuous Laplacian operator, $\lambda_1 < \lambda_k$ be eigenvalues of \mathcal{L} , and q_k be the eigenvectors of \mathcal{L} . Given $\omega = \{x_i\}$ is a sequence of i.i.d. sample from ρ , we can define the discrete graph Laplacian as $s_n L_n$. We denote λ_{NK} as the discrete eigenvalues, and q_k^N , a choice of the eigenvectors corresponding to λ_{NK} . Then for any fixed K , the following holds almost surely with respect to ω .*

1. $\lim_{N \rightarrow \infty} \lambda_{NK} = \lambda_K$
2. *There exists a choice of eigenvectors $\{q_{NK}\}$ for each N such that*

$$\lim_{N \rightarrow \infty} \|q_{NK} - \pi_N(q_K)\|_{\nu_N} \rightarrow 0$$

We apply proposition 1 to prove the convergence of expectations for a class of functionals F_N and F . Namely, we want them to satisfy the continuity condition

DEFINITION 2. *Suppose F_N be functionals defined on $u \in \mathbb{R}^N$, and F be functionals on $u \in \mathbb{R}^N$. We say F_N and F satisfy the continuity condition if $\|u_n - \pi_n(u)\|_{\nu_n} = 0$ implies $F(u) = F_N(u)$.*

PROPOSITION 3. *Let the variables and notations be the same as in Proposition 1, and F_N, F are functional defined in 2 satisfying the continuity condition. If F_N and F are bounded, then*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\mu_{NK}} F_N = \mathbb{E}_{\mu_K} F$$

Proof. Given the choice of basis q_k , we can express the expectations in terms of the eigenbasis q^k . $E(F) = \int \exp(-\|\xi\|^2) G_N(\xi)$. Since q_{NK} and q_K are close, G_N converges to G point-wise. By dominated convergence, we have E_{NK} converges to E_K . \square

Since the likelihood is defined on the set of fidelity points, We need to control q_{NK} and q_K on the set of fidelity points. To avoid L^∞ estimates, we make the assumption that the labels are given on the first J samples of x .

PROPOSITION 4. Let q^k be the continuous eigenvectors. For all k and $\epsilon > 0$, we have with probability $1 - o(1)$ that $\exists \{q_{NK}\}$ eigenvectors of L_N such that $\|\pi_J^N(q_{NK}) - \pi_J(q_K)\|_\infty < \epsilon$.

Proof. 55qzyu

As an immediate corollary, we show the restriction of the measures on a finite fidelity set are close.

COROLLARY 5. Let $\mu_K^J := \pi_J^*(\mu_K)$, and $\mu_{NK}^J := \pi_J^*(\mu_{NK})$. Let C_{NK}^J and C_K^J be the covariance matrix of these Gaussian measures on \mathbb{R}^J . For all k and $\epsilon > 0$, we have with probability $1 - o(1)$ that $\|C_K^J - C_{NK}^J\| < \epsilon$.

3. Analysis of pCN. Given a Gaussian measure, the the pCN algorithm stated below:

$$(1) \quad \begin{aligned} v &= (1 - \beta^2)^{\frac{1}{2}}u + \beta w \\ a(u, v) &= \exp(\Phi(u) - \Phi(v)) \end{aligned}$$

pCN is well defined in the continuum, as well as on the truncated space μ_K and the diffuse limit theory in [3] is applicable.

In this paper, we use basic calculations to show that the expected acceptance probability and mean square jump have a non-zero upper bound on the continuum problem, and show this bound holds under the truncation error and the discretization error is small.

We first prove a bound on the acceptance probability of the pCN method. The quantity $Z = \int_u \mathbf{1}_A(u) d\mu_J(u) = \int_u \mathbf{1}_A(u) d\mu(u)$ needs to be positive to ensure the existence of a conditional probability, and positivity. We assume that the label set y is chosen such that Z is positive. We call such assignments a *compatible* label assignment.

PROPOSITION 6. Let $u \sim \mu_p$ the posterior, and v according to the pCN proposal. Suppose the log likelihood Φ satisfies $0 < \Phi(u; y) < C$ if $\text{Sign}(u) = y$. Then we have the estimate

$$\mathbb{E}_{\mu_p} a(u, v) \geq \frac{Z}{C}$$

where $A = \{u | \text{Sign}(u) = y\}$, $Z = \int_u \mathbf{1}_A(u) d\mu_J(u) = \int_u \mathbf{1}_A(u) d\mu(u)$

(Check the constants more carefully and make sure they are in $[0, 1]$)

Proof.

$$(2) \quad \begin{aligned} \mathbb{E}_{\mu_p} a(u, v) &= \int_u \left(\int_v \exp(\Phi(u) - \Phi(v)) d\mathbb{P}(v|u) \right) d\mu_p(u) \\ &\geq \frac{1}{C} \int_u (\mathbf{1}_A(v) d\eta_u(v)) d\mu_p(u) = \frac{1}{CZ'} \int_u ((\mathbf{1}_A(v) d\eta_u(v))) \mathbf{1}_A(u) d\mu(u) \end{aligned}$$

Where $Z' = \int_u \exp(-\Phi(u)) d\mu(u) \geq \int_u \mathbf{1}_A(u) d\mu_J(u)$. Note that the integrant function

factors through the projection π_J ,

$$\begin{aligned}
\int_u (\mathbb{P}(v \in A|u)) \mathbf{1}_A(u) d\mu(u) &= \int_u \mathbb{P}(v > 0|u) \mathbf{1}_A(u) d\mu_J(u) \\
&= \int_u \left(\int_w \mathbf{1}_A(w + \frac{\sqrt{1-\beta^2}}{\beta} u) d\mu_J(w) \right) \mathbf{1}_A(u) d\mu_J(u) \\
&\geq \int_u \left(\int_w \mathbf{1}_A(w) d\mu_J(w) \right) \mathbf{1}_A(u) d\mu_J(u) \\
&= \left(\int_u \mathbf{1}_A(u) d\mu_J(u) \right)^2
\end{aligned}
\tag{3}$$

Since the analysis holds regardless of the Gaussian measure and the space it was defined on, the same estimates holds if we replace μ with μ_{NK} and μ_K . Hence to prove an upper bound for the discrete and continuous truncated measure, we need to show the quantity Z is close to Z_{NK} and Z_K .

PROPOSITION 7. *Let $Z_K = \int_u \exp(-\Phi(u)) d\mu_K(u) = \int_u \exp(-\Phi(u)) d\mu_K^J(u)$. Then $\lim_{K \rightarrow \infty} Z_K = Z$.*

Proof. This comes from the fact that μ is trace class, and hence μ_K^J converges to μ^J . \square

For the discrete approximation, we have

PROPOSITION 8. *Let $Z_{NK} = \int_u \exp(-\Phi(u)) d\mu_K(u) = \int_u \exp(-\Phi(u)) d\mu_{NK}^J(u)$. Then for all $\epsilon > 0$, with probability $1 - o(1)$, $|Z_{NK} - Z_K| < \epsilon$.*

Proof. Since $\|\mu_{NK}^J - \mu_K^J\| < \epsilon$, we have Z_{NK} and Z_K are close with probability $1 - o(1)$. \square

We prove an estimate for the mean square jump as well.

PROPOSITION 9. *Define the mean square jump as $\mathbb{E}_{\mu_p} \|u - v\|^2$. Then for $0 < \Phi(u; y) < C$ for $S(u) = y$, we have $\mathbb{E}_{\mu_p} \|u - v\|^2 > C'$ for some positive constant C' independent of beta, Φ .*

Proof.

$$\begin{aligned}
\mathbb{E}_{\mu_p} \|u - v\|^2 &= \int_u \left(\int_{v \in A} \|u - v\|^2 d\mathbb{P}_{v|u} \right) d\mu_p(u) \\
&= \int_u \left(\int_v \|u - v\|^2 \mathbf{1}_A(v) \mathbb{P}_{v|u} \right) \mathbf{1}_A(u) d\mu(u) \\
&= \int_u \left(\int_v \|\beta w - \beta' u\|^2 \mathbf{1}_A(w + \frac{\sqrt{1-\beta^2}}{\beta} u) d\mu(w) \right) \mathbf{1}_A(u) d\mu(u) \\
&\geq \int_u \left(\int_w \|\beta w - \beta' u\|^2 \mathbf{1}_A(w) d\mu(w) \right) \mathbf{1}_A(u) d\mu(u) \\
&\geq \beta^2 \int_u \left(\int_w \|w - \frac{\beta'}{\beta} u\|^2 \mathbf{1}_A(w) d\mu(w) \right) \mathbf{1}_A(u) \mathbf{1}_{\|u\| \geq 2\delta} d\mu(u) \\
&\geq \beta^2 \int_u \left(\int_w \|w - \frac{\beta'}{\beta} u\|^2 \mathbf{1}_A(w) \mathbf{1}_{\|w\| \leq \delta} d\mu(w) \right) \mathbf{1}_A(u) \mathbf{1}_{\|u\| \geq 2\delta} d\mu(u) \\
&\geq \delta^2 \beta^2 \int_u \left(\int_w \mathbf{1}_A(w) \mathbf{1}_{\|w\| \leq \delta} d\mu(w) \right) \mathbf{1}_A(u) \mathbf{1}_{\|u\| \geq 2\delta} d\mu(u) \\
&= \delta^2 \beta^2 \left(\int_w \mathbf{1}_A(w) \mathbf{1}_{\|w\| \leq \delta} d\mu(w) \right) \times \left(\int_u \mathbf{1}_A(u) \mathbf{1}_{\|u\| \geq 2\delta} d\mu(u) \right)
\end{aligned}
\tag{4}$$

where $\beta' = 1 - \sqrt{1 - \beta^2}$, and $\delta > 0$. Hence we have to show that for $\delta > 0$, $\int_w \mathbf{1}_A(w) \mathbf{1}_{\|w\| \leq \delta} d\mu(w)$ and $\int_u \mathbf{1}_A(u) \mathbf{1}_{\|u\| \geq 2\delta} d\mu(u)$ are both greater than 0. \square

We show this bound is stable under truncation and discretization.

PROPOSITION 10. *Let $I = \int_w \mathbf{1}_A(w) \mathbf{1}_{\|w\| \leq \delta} d\mu$. Then $\lim_{K \rightarrow \infty} I_K = I$.*

Proof. Cite standard theory for weak convergence of measures μ_K to μ ? \square

PROPOSITION 11. *Let $I_K = \int_w \mathbf{1}_A(w) \mathbf{1}_{\|w\| \leq \delta} d\mu_K$, and $I_{NK} = \int_w \mathbf{1}_A(w) \mathbf{1}_{\|w\| \leq \delta} d\mu_{NK}$. Then $|I_K - I_{NK}| < \epsilon$ with probability $1 - o(1)$.*

Proof. We use Proposition 2. Under such a basis q_{NK} , we have

$$I_{NK} = \int \mathbf{1}_A(Q_N^J \xi) \mathbf{1}_{B_\delta}(\xi) \exp(-\|\xi\|_{\Lambda_N}^2) d\xi$$

. By convergene of eigenvalues $|I_{NK} - I_K| < \epsilon$. \square

4. Auxiliary Variable Methods.

4.1. Gibbs Method. The Auxiliary variable Gibbs approach is to model the variable $u^*(j) = u(j) + \epsilon$. There are two ways. One is to model u^* on all nodes, the second is to only model u^* on labelled nodes. We call these two approach Gibbs(I) and Gibbs(II). We show Gibbs(I) does not scale well with N .

$$\begin{aligned} P' &= P + \gamma^{-2} I, \\ P' m &= \gamma^{-2} u^*. \end{aligned}$$

The Standard Gibbs method(I) is ill defined on the continuum. This should reflect in the slowing down of Gibbs sampler as $N \rightarrow \infty$. This should be similar to [1, 2]. We quantify this slowing down precisely in the next proposition.

PROPOSITION 12. *Given fixed u , let u^+ be the value of the next sample. Let $\hat{u}_k^* = \langle u^*, q_{NK} \rangle$. Then $\hat{u}_k^+ = \hat{u}_k^*(1 + o(1))$ with large probability.*

(Andrew: Formulate this more elegantly as the “white noise disappearance problem” in the continuum.

Let $u^*(j) = u(j) + \eta(j)$, $j \in Z'$ be only on the fidelity set. Then $\mathbb{P}(u^*|u, y)$ as before is independent truncated Gaussians. $\mathbb{P}(u|u^*, y) = N(m, P')$, where

$$\begin{aligned} P' &= P + \gamma^{-2} D_J, \\ P' m &= \gamma^{-2} E_J u^*, \end{aligned}$$

where $I_J = [I_J, 0; 0, 0]$, and $E_J = [I_J; 0]$.

Standard Gibbs method(II) seems well-defined on the continuum, and empirically behaves well when N large. However, when γ is small, the method freezes on the observed labels $\{x_1, \dots, x_J\}$.

PROPOSITION 13. *Let $u \in \mathbb{R}^N$. Then $\|u_J^+ - u_J\|_\infty \leq \gamma \|u_J\|_\infty$.*

4.2. Holmes and Held. In Holmes and Held, the standard Gibbs method was revised to handle the case of small noise γ , by considering the factorization $\mathbb{P}(u, u^*|y) = \mathbb{P}(u * |y) \mathbb{P}(u|u^*)$. The method first samples from $\mathbb{P}(u^*|y)$, and then samples u according to $\mathbb{P}(u|u^*)$, which is a Gaussian.

To study Holmes and Held, and compare with other methods, it is necessary to list two key components of this method. The first is the auxiliary variable approach,

to sample $\mathbb{P}(u, u * | y)$ jointly. This is the same with the Gibbs sampler. The second idea which is the key for small γ is to marginalize u and sample from $\mathbb{P}(u^* | y)$.

Since the HH methods is designed for small γ , it is telling to study the algorithm in the small γ limit. In the small γ limit, HH reduces to sampling $\mathbb{P}(u_1 | y)$, and then sampling $\mathbb{P}(u_2 | u_1)$, where $u = (u_1, u_2)$, where u_1 is the value of u on the observed nodes. The efficacy of the algorithm depends on the sampler used for $\mathbb{P}(u_1 | y)$. In fact, one can use pCN to sample from $\mathbb{P}(u_1 | y)$. The proposition below shows a connection between the pCN method, and HH-pCN.

PROPOSITION 14. *Let $u = (u_1, u_2)$. Suppose $v = (v_1, v_2)$ the proposal from pCN, $w = (w_1, w_2)$ be from the HH-pCN method, and $acc_1(u, v)$, $acc_2(u, w)$ be the corresponding acceptance probabilities. Then $acc_1 = acc_2$ as a function of (u, v) , $w_1 = v_1$ in law. Moreover, $w_1 = (1 - \beta)xxw_2$ in law.*

Proof.

5. Numerical Experiments.

5.1. Large N limits of pCN, Gibbs. This section studies the large N behaviour of pCN and Gibbs (I) and (II). We sample $N = 2000 : 1000 : 8000$ samples from a uniform distribution on a square domain, and choose five points in the square as fidelity, with 2 points in class 1 and 3 points in class 2. We use the scaling in slepcev17 paper, and project to 60 eigenvectors. We vary $\gamma = [0.1, 0.3, 0.7, 1.0]$ over a moderate range, and plot the acceptance probability (for pCN only), the mean square jump, the empirical l_2 error over 2×10^4 iterations, as a function of N . Each experiment is averaged over 10 realizations of samples from the uniform distribution.

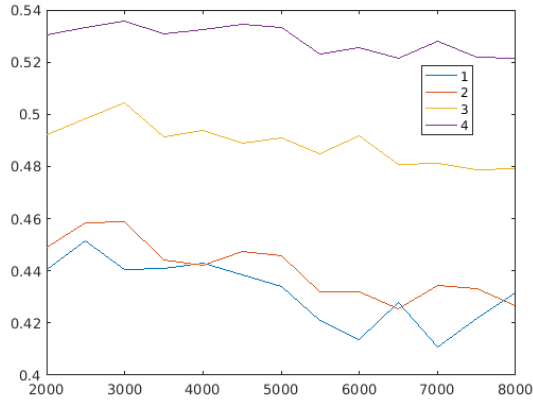


FIG. 1. acceptance probability of pCN

5.2. Small γ Limit. This section compares the Gibbs sampler and the probit under the small γ limit. We choose $\gamma = 1e - 3$, and plot the sampled values of $u(j)$ for $j \in J$ in the fidelity set. We see the Gibbs(II) method freezes on those nodes, while the pCN method does not.

5.3. Gaussian measures. In this section, we study the quality of the approximation of μ_N , μ_{NK} to the continuum by studying the pushforward of these measures on a finite set of points. We demonstrate empirically that pushforward of the full

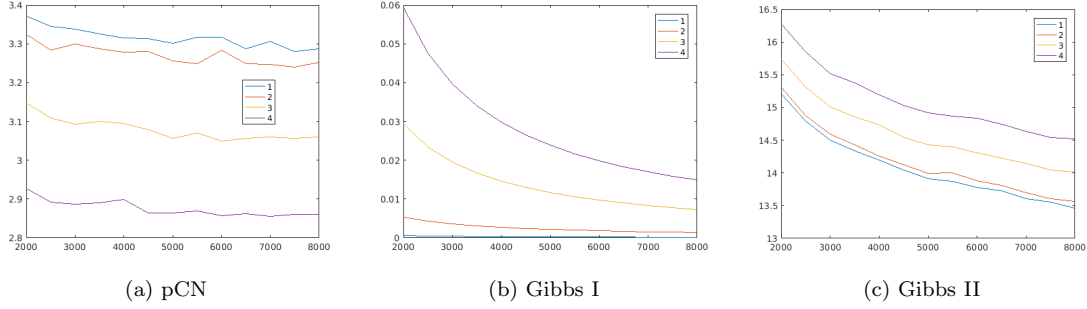


FIG. 2. *Mean Square Jump*

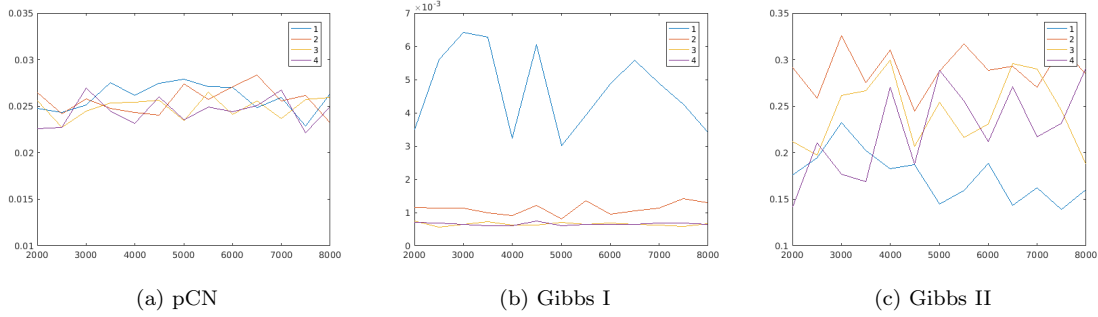


FIG. 3. *L2 error*

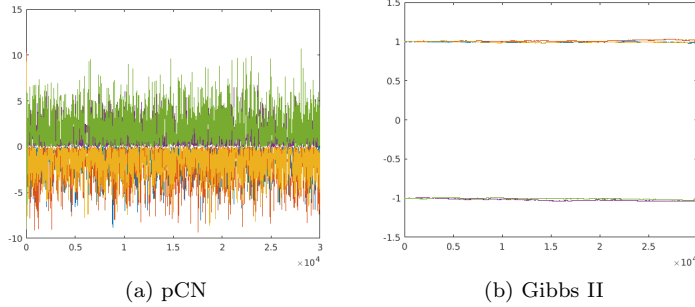


FIG. 4. *Samples of $u(j)$ for $j \in J$.*

discrete Gaussian measure μ_N does not converge the continuum measure μ when N increases. Moreover, we show that the truncated measure $\mu_{N,K}$ by choosing K such that the spectrum as converged, approximates the true continuum μ better than the full Laplacian.

This shows empirically that the eigenvectors/values of L_N which have not yet converged to the continuum induces a non-trivial error away from the true continuum limit.

Appendix A. Gaussian measures. We look at measures of the form mu whose

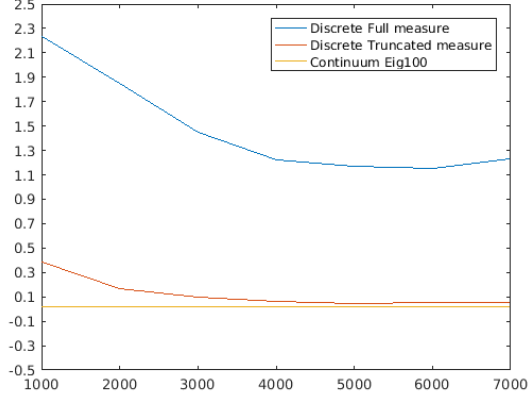


FIG. 5. l_2 err of Covariance matrix C_{JJ} as a function of N computed from L_N , L_{NK} and the continuum measure μ .

covariance operator is $C = (L + \tau I)^\alpha$ with $\alpha > 1$. From Slepcev17, we have that the eigenvalues and the eigenvector projections converge to its continuum limits. However, this convergence is not uniform among different eigenspace. In fact, we empirically observe that only a small fraction $K \ll N$ of the eigenvectors have actually converged. Moreover, the eigenvectors that have not converged can even hinder the convergence of certain quantities as $N \rightarrow \infty$. This motivates us to consider decoupling the number of nodes N and the number of eigenvalues K , and study separate limits.

- For fixed K , we show that for certain functions f , $|E_{\mu_k^N} f - E_{\mu_k} f|$ is small with high probability.
- We show that for certain f , $|E_{\mu_k} f - E_{\mu_k} f|$ is small for large k .

The second point follows from the fact that μ is trace class, while the first fact uses conclusions about the approximation of the continuum eigenvectors from the discrete.

A.1. Approximation of Eigenvectors. We first prove a proposition which shows that we can approximate an orthonormal eigenbasis of the discrete Laplacian that spans V_K^N by restricting a set of orthonormal eigenbasis in the continuous domain to the set of data points.

PROPOSITION 15. *Let V_k be the eigenspace of L of the eigenvalue $\bar{\lambda}_k = \lambda_{\hat{k}}, \dots, = \lambda_{\hat{k}+s(k)}$. Let $\{v_j\}$ be eigenvectors of L^n corresponding to the eigenvalues λ_j^n . Then almost surely, $\exists \{\phi_j^{(n)}\}$ an eigenbasis of V_k such that $\lim_{n \rightarrow \infty} \|\tilde{\phi}_j^{(n)} - v_j\|_{\nu_n} = 0$, where $\tilde{\phi}_j^n = \phi_j^n(x_1, \dots, x_n) \in \mathbb{R}^n$.*

Proof. Let $\{\phi^j\}$ be a choice of an eigenbasis of V^k , and $\phi_j^n = \phi_j(x_1, \dots, x_n)$. By smoothness of ϕ_j , $\phi_j^n \rightarrow \phi_j$ in the $TL2$ sense. Let $w_j^n = Proj_K^N(\phi_j^n)$, then $w_j^n \rightarrow Proj_k(\phi_j) = \phi_j$ in $TL2$. Hence $\|w_i^n - \phi_i^n\|_{\nu_n} \rightarrow 0$. Moreover, $G_{ij} = \langle w_i^n, w_j^n \rangle_{\nu_n} \rightarrow \delta_{ij}$ as $n \rightarrow \infty$. Since G is full rank, $\{w_j\}$ spans V_k^n . Let $\{w_j'^n\}$ be the Gram-Schmidt orthogonalization of $\{w_j^n\}$. Note that if the Gram matrix G is sufficiently close to the identity, $\|w_j'^n - w_j^n\|_{\nu_n}$ will be small. Hence we can choose N such that $\forall n > N$, $\|w_j'^n - w_j^n\|_{\nu_n} < \epsilon$, and $\|w_j'^n - \phi_j^n\| \leq 2\epsilon$.

Let Q be an orthogonal transform mapping $w_j'^n$ to v_j^n . Then letting $\phi_j' = \sum_i \phi_i q_{ij}$,

we have $\|\phi_j'^n - v_j\| = O(\epsilon)$. \square

It is useful to study the convergence of the eigenfunctions when restricted to the first J components. Since the convergence is in TL2, this does not follow immediately from the original theorem. However, we are able to prove:

PROPOSITION 16. *Let $\{x_i\}$ be an i.i.d. sequence sampled from ρ . Let v_i^n be eigenvectors of V_k^n . Then with probability $1 - o(1)$ with respect to N , $\exists \phi_j$ such that $|\phi_i(x_j) - v_i^n(j)| < \epsilon$, $\forall i = \{1, \dots, J\}, j = 1, \dots, s(k)$.*

Proof. Let $E_{n,\epsilon}^J$ be the event $\{w = \{x_i\} | \exists \phi_j \text{ s.t. } |\phi_j(x_i) - v_i^n(j)| < \epsilon\}$. Since $\{x_i\}$ is i.i.d, $P(E_{n,\epsilon}^J) = P(E_{n,\epsilon}^{J'})$, $\forall |J'| = |J|, J' \subset \{1, \dots, n\}$. Let dJ' be the uniform measure on subsets of size $|J|$, we have

$$(5) \quad \begin{aligned} P(E_{n,\epsilon}^J) &= \int_{J'} \int_w \mathbf{1}_{E_{n,\epsilon}^{J'}}(w) dw dJ' \\ &= \int_w \int_{J'} \mathbf{1}_{E_{n,\epsilon}^{J'}}(w) dJ' dw \end{aligned}$$

Define the set $F_{n,\delta} = \{w | \exists \{\phi_j\}, \text{ s.t. } \|\phi_j^n - v_j^n\|_{\nu_n} < \delta, \forall k \leq K\}$. Note that if $\|\phi_j^n - v_j^n(i)\|_{\nu_n} < \delta$, then $|\phi_j^n(i) - v_j^n(i)| < \epsilon$ for all i except on a set $K(w) \subset \{1, \dots, n\}$, $|K(w)| < \frac{\delta}{\epsilon}n$. Hence

$$(6) \quad \int_{J'} \mathbf{1}_{E_{n,\epsilon}^{J'}}(w) dJ' \geq \int_{J'} \mathbf{1}_{F_{n,\delta}}(w) \mathbf{1}_{J' \cap K(w) = \emptyset}(J') dJ'$$

Therefore

$$(7) \quad \begin{aligned} P(E_n^J) &\geq \int_w \mathbf{1}_{F_{n,\delta}}(w) \left(\int_{J'} \mathbf{1}_{J' \cap K(w) = \emptyset}(J') dJ' \right) dw \\ &= \int_w \mathbf{1}_{F_{n,\delta}}(w) \mathbf{1}_{J' \cap K(w) = \emptyset}(J') dJ' \\ &\geq \int_w \mathbf{1}_{F_{n,\delta}}(w) \left(1 - C \frac{\delta}{\epsilon}\right)^J \\ &= P(F_{n,\delta}) \left(1 - C \frac{\delta}{\epsilon}\right)^J \end{aligned}$$

where the above holds for any δ . We want to use the above to show $\lim_{n \rightarrow \infty} P(E_{n,\epsilon}^J) = 0$ by choosing a sequence of $\delta_n \rightarrow 0$. Note that from above, we have $P(\cap_{\delta} \cup_N \cap_{n > N} F_{n,\delta}) = 1$. Let $G_{N,\delta} = \cap_{n > N} F_{n,\delta}^J$. Note that $\lim_{N \rightarrow \infty} P(G_{N,\delta}) = 1$ monotonically, and $P(G_{N,\delta_1}) < P(G_{N,\delta_2})$ if $\delta_1 < \delta_2$. Given any σ , we choose δ such that $(1 - C \frac{\delta}{\epsilon}) > (1 - \sigma/2)$, and N such that $P(F_{n,\delta}) > 1 - \sigma/2$. Then $P(E_{n,\epsilon}^J) > 1 - \sigma$, $\forall n > N$. \square

A.2. Approximation of Gaussian Measures. We compare the Gaussian measures μ_K and μ_{NK} . Since the two measures are not defined on the same space, we do not directly compare the two measures, but instead map these measures to the Fourier space. Namely, given an eigenbasis $\{\phi_j\}$, we can identify R^K with V^K via $\Phi : (\xi_1, \dots, \xi_k) \mapsto \sum_i \xi_i \phi^i$. Define $\tilde{\mu}_K = \Phi^{-1*} \mu_K$, we show in the proposition below that there exists an eigenbasis Φ such that $\tilde{C}_K \approx \tilde{C}_K^N$, and $\phi^n \approx v^n$.

PROPOSITION 17. *Let ϕ_j be orthonormal basis of V_K . There almost surely exists an orthonormal basis $\{q_j^n\}$ of V_K^N such the following holds:*

1. $q_j^n \rightarrow \phi_j$ in TL2.

2. The measure $\tilde{\mu}_{NK} := Q^{-1*}(\mu_N K)$ defined on \mathbb{R}^K converges to $\tilde{\mu}_K := N(0, C)$, where C is the diagonal matrix.

Proof. (sketch): We show that under an orthonormal transform of the eigenbasis v_j^n to q_j^n , we have $\|q_j^n - \phi_j^n\|_{\nu_n} \rightarrow 0$. Thus $q_j^n \rightarrow \phi_j$ in TL2. Moreover, since the transform preserves V_N^K , $|\tilde{\mu}_{NK} - \tilde{\mu}_K| = o(1)$. \square

As a corollary, we can show that $\lim_{n \rightarrow \infty} E_{\mu_{nk}} f^n = E_{\mu_K} f$ almost surely for certain class of f .

COROLLARY 18. Let f^n be a bounded function defined on \mathbb{R}^N , f such that $f^n(u^n) \rightarrow f(u) \forall u^n \rightarrow u$. Then $\lim_{n \rightarrow \infty} E_{\mu_{nk}} f^n = E_{\mu_K} f$

Proof. (sketch): Choose $\{q_j^n\}$ the basis as above. Let $g^n(\xi) = f(\xi^T q^n)$. Then $g^n \rightarrow g$ point-wise by property of f . Since $\tilde{\mu}_{NK} \rightarrow \tilde{\mu}_K$, by dominated convergence, $E_{\mu_{NK}} f^n \rightarrow E_{\mu_K} f$. \square

Finally, we prove a similar proposition on the projection of the measures onto a finite set of sample points.

PROPOSITION 19. Let $\{x_i\}$ be an i.i.d. sequence sampled from ρ . Let $\mu_{NK}^J = \pi_N^{J*}(\mu_{NK})$, and $\mu_K^J = \pi_K^{J*}(\mu_K)$, where $\pi^J(u) = u(x_1, \dots, x_J)$ is the evaluation map on the first J indices. Then for any $\epsilon > 0$, we have $|\mu_{NK}^J - \mu_K^J| < \epsilon$ with probability $1 - o(1)$ with respect to N .

COROLLARY 20. Let q_N^k be the discrete eigenbasis of level k . Then for all $\epsilon > 0$, with probability $1 - o(1)$ exists $|q_N^k(j) - q^k(j)| = 0, \forall j \in J$.

REFERENCES

- [1] S. AGAPIOU, J. M. BARDSLEY, O. PAPASPILIOPOULOS, AND A. M. STUART, *Analysis of the gibbs sampler for hierarchical inverse problems*, SIAM/ASA Journal on Uncertainty Quantification, 2 (2014), pp. 511–544.
- [2] A. BESKOS, G. ROBERTS, AND A. STUART, *Optimal scalings for local metropolis-hastings chains on nonproduct targets in high dimensions*, The Annals of Applied Probability, (2009), pp. 863–898.
- [3] N. S. PILLAI, A. M. STUART, AND A. H. THIÉRY, *On the random walk metropolis algorithm for gaussian random field priors and the gradient flow*, Arxiv preprint, (2011).
- [4] N. G. TRILLOS AND D. SLEPČEV, *A variational approach to the consistency of spectral clustering*, Applied and Computational Harmonic Analysis, (2016).