Bayesian Design of Sampling Set for Bandlimited Graph Signals

Xuan Xie, Junhao Yu, Hui Feng, Bo Hu

Research Center of Smart Networks and Systems, Department of Electronic Engineering, Fudan University

Introduction

The design of sampling set (DoS) for bandlimited graph signals (GS) has been extensively studied in recent years, but few of them exploit the benefits of the stochastic prior of GS. In this work, we introduce the optimization framework for Bayesian DoS of bandlimited GS. We also illustrate how the choice of different sampling sets affects the estimation error and how the prior knowledge influences the result of DoS compared with the non-Bayesian DoS by the aid of analyzing Gershgorin discs of error metric matrix. Finally, based on our analysis, we propose a heuristic algorithm for DoS to avoid solving the optimization problem directly.

Framework

Consider an N-vertex undirected connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, where $\mathcal{V}$ is the vertex set, $\mathcal{E}$ is the edge set and $\mathbf{W}$ is the weighted adjacency matrix. The graph Laplacian is defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$. The spectral decomposition of $\mathbf{L} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$ is $\mathbf{L} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$. The graph Fourier transform (GFT) can be expressed as $\mathbf{f} = \mathbf{V}\hat{\mathbf{f}}$.

Subspace prior: A graph signal can be represented by a linear combination of a subset of $\{\mathbf{v}_k\}$. Explicitly, if $\mathbf{f}$ is in the $\mathcal{K}$-subspace, where $\mathcal{K} \subset \mathcal{V}$ and $|\mathcal{K}| = K$, then it satisfies

$$\mathbf{f} = \mathbf{V}_{\mathcal{K}}\hat{\mathbf{f}}_{\mathcal{K}}. \tag{1}$$

Stochastic prior: $\hat{\mathbf{f}}_{\mathcal{K}}$ is known to be drawn from the following distribution

$$p(\hat{\mathbf{f}}_{\mathcal{K}}) \propto \exp(-(\hat{\mathbf{f}}_{\mathcal{K}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}}^{-1}(\hat{\mathbf{f}}_{\mathcal{K}} - \boldsymbol{\mu})), \tag{2}$$

where $\boldsymbol{\mu}$ is the mean of $\hat{\mathbf{f}}_{\mathcal{K}}$. Let $\boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}} = \text{diag}(\sigma_{\mathcal{K}_1}^2, \cdots, \sigma_{\mathcal{K}_K}^2)$, then each diagonal element represents the uncertainty of the corresponding mean value. The sampling operator $\boldsymbol{\Psi} : \mathbb{C}^N \mapsto \mathbb{C}^M$ is defined as

$$\boldsymbol{\Psi}_{i,j} = \begin{cases} 1, & j = \mathcal{S}_i; \\ 0, & \text{otherwise}. \end{cases} \tag{3}$$

The observation model is $\mathbf{y}_{\mathcal{S}} = \boldsymbol{\Psi}\mathbf{y} = \boldsymbol{\Psi}(\mathbf{V}_{\mathcal{K}}\hat{\mathbf{f}}_{\mathcal{K}} + \mathbf{w})$, where $\mathbf{w}$ is the *i.i.d* noise with zero mean and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{w}} = \sigma_{\mathbf{w}}^2\mathbf{I}$.

In this paper, the goal is to design a sampling set that can estimate $\hat{\mathbf{f}}_{\mathcal{K}}$ by samples on them with the least estimation error. If m_i observations are taken at the i-th vertex, then we define the design as $\eta_i = m_i/M$, which is a proportion of vertices being sampled. Let $\mathbf{u}_1^T, \cdots, \mathbf{u}_N^T$ be the rows of $\mathbf{V}_{\mathcal{K}}$, then $\boldsymbol{\Sigma}_{\text{B}}^*$ is a function of the design $\boldsymbol{\eta}$ as

$$\boldsymbol{\Sigma}_{\text{B}}^*(\boldsymbol{\eta}) \triangleq (\sigma_{\mathbf{w}}^{-2}M\sum_{i=1}^{N}\eta_i\mathbf{u}_i\mathbf{u}_i^T + \boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}}^{-1})^{-1}. \tag{4}$$

Since the prior distribution does not depends on the design $\boldsymbol{\eta}$, the design maximizes the expected utility is the one that maximizes

$$\begin{aligned} U_1(\boldsymbol{\eta}) &\triangleq \iint \log p(\hat{\mathbf{f}}_{\mathcal{K}}|\mathbf{y}_{\mathcal{S}}, \boldsymbol{\eta})p(\mathbf{y}_{\mathcal{S}}, \hat{\mathbf{f}}_{\mathcal{K}}|\boldsymbol{\eta})d\hat{\mathbf{f}}_{\mathcal{K}}d\mathbf{y}_{\mathcal{S}} \\ &= -\frac{K}{2}\log(2\pi) - \frac{K}{2} + \frac{1}{2}\log\det(\boldsymbol{\Sigma}_{\text{B}}^*(\boldsymbol{\eta}))^{-1}. \end{aligned} \tag{5}$$

The optimization problem can be expressed as follow

$$\begin{aligned} &\max_{\boldsymbol{\eta}} \quad U_1(\boldsymbol{\eta}) \\ &\text{s.t.} \quad 0 \le \eta_i \le 1, \ \sum_{i=1}^{N}\eta_i = 1, \\ &\qquad M\eta_i \in \mathbf{Z}. \end{aligned} \tag{6}$$

Algorithm

The optimization problem (6) is an intractable combinatorial problem, but it can be converted to a convex optimization problem by relaxing the constraint condition $M\eta_i \in \mathbf{Z}$. Instead of solving the relaxed problem directly, we analyze how $\boldsymbol{\eta}$ changes the bound of eigenvalues of $(\boldsymbol{\Sigma}_{\text{B}}^*(\boldsymbol{\eta}))^{-1}$ and find a heuristic method to decide $\boldsymbol{\eta}^*$ with low complexity.

The optimal $\boldsymbol{\eta}^*$ will maximize

$$\begin{aligned} &\log\det(\boldsymbol{\Sigma}_{\text{B}}^*(\boldsymbol{\eta}))^{-1} \\ &= \log\det(\sigma_{\mathbf{w}}^{-2}\boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}}^{-\frac{1}{2}}(M\boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}}^{\frac{1}{2}}\mathbf{V}_{\mathcal{K}}^T\text{diag}(\boldsymbol{\eta})\mathbf{V}_{\mathcal{K}}\boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}}^{\frac{1}{2}} + \sigma_{\mathbf{w}}^2\mathbf{I})\boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}}^{-\frac{1}{2}}) \\ &= \log\det(\sigma_{\mathbf{w}}^{-2}\boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}}^{-1}) + \sum_{i=1}^{K}\log(\lambda_i^{\text{B}}), \end{aligned} \tag{7}$$

where $\{\lambda_1^{\text{B}}, \cdots, \lambda_K^{\text{B}}\}$ are the eigenvalues of $(M\boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}}^{\frac{1}{2}}\mathbf{V}_{\mathcal{K}}^T\text{diag}(\boldsymbol{\eta})\mathbf{V}_{\mathcal{K}}\boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}}^{\frac{1}{2}} + \sigma_{\mathbf{w}}^2\mathbf{I})$. In order to analyze the bound of $\boldsymbol{\lambda}^{\text{B}}$ corresponding to $\boldsymbol{\eta}^*$, we define

$$\mathbf{G}_{\text{B}}(\boldsymbol{\eta}) \triangleq M\text{diag}(\boldsymbol{\eta})^{\frac{1}{2}}\mathbf{V}_{\mathcal{K}}\boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}}\mathbf{V}_{\mathcal{K}}^T\text{diag}(\boldsymbol{\eta})^{\frac{1}{2}} + \sigma_{\mathbf{w}}^2\mathbf{I},$$

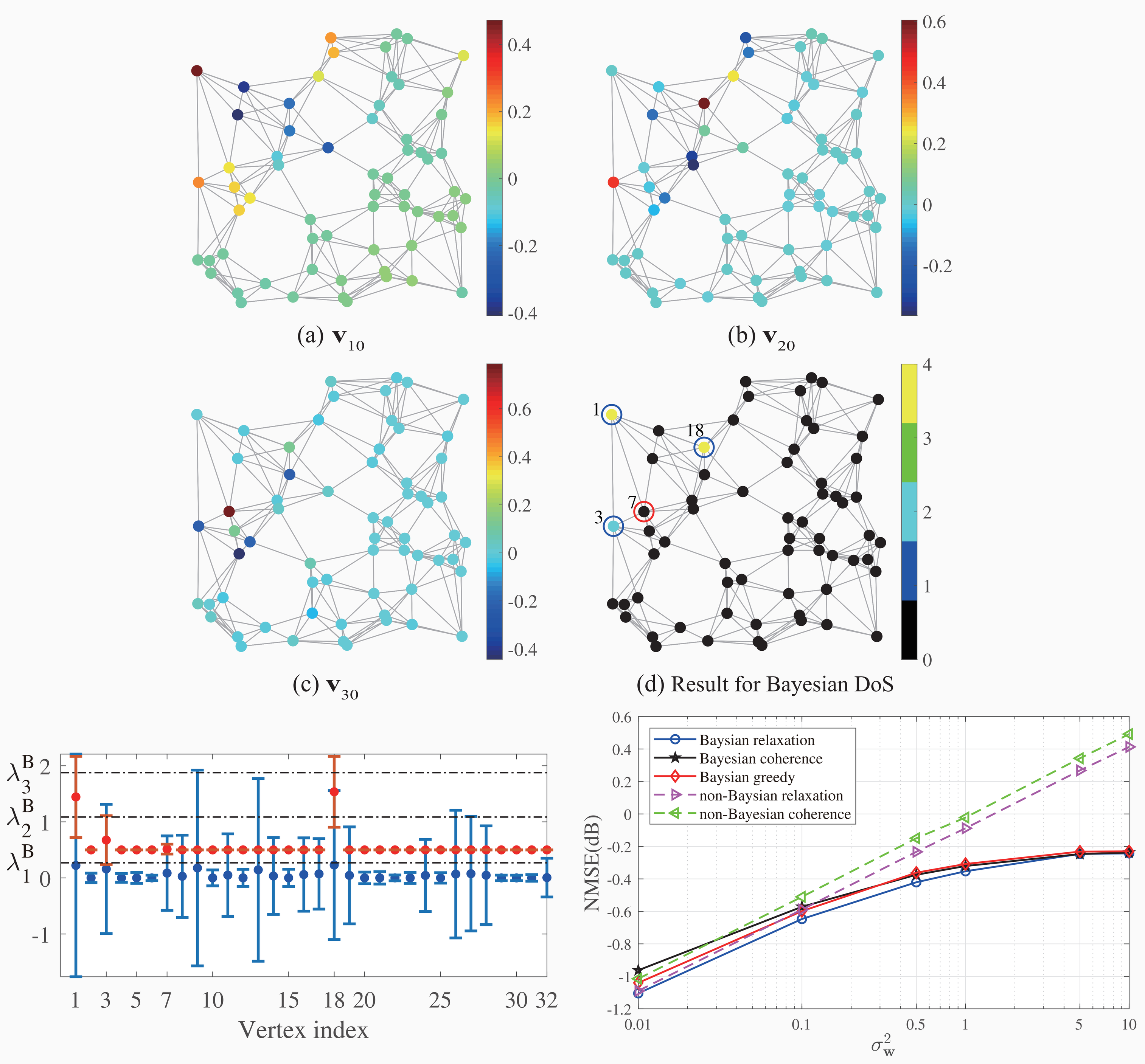
whose K largest eigenvalues are the same as $\boldsymbol{\lambda}^{\text{B}}$ and the remaining eigenvalues are constants irrelevant to $\boldsymbol{\eta}$.

Denoting the rows of $(\mathbf{V}_{\mathcal{K}}\boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}}^{1/2})$ by $\tilde{\mathbf{u}}_1^T, \cdots, \tilde{\mathbf{u}}_N^T$, we define the *Bayesian graph coherence* for vertex i as $\tilde{\mathbf{u}}_i^T\tilde{\mathbf{u}}_i = \|\boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}}^{1/2}\mathbf{V}_{\mathcal{K}}^T\boldsymbol{\delta}_i\|_2^2$. Since $\|\mathbf{V}_{\mathcal{K}}\boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}}^{1/2}\|_2^2 = \sum_{i=1}^{K}\sigma_{\mathcal{K}_i}^2$, the sampling proportion in our heuristic method is given as

$$\eta_i^* = \frac{\tilde{\mathbf{u}}_i^T\tilde{\mathbf{u}}_i}{\sum_{i=1}^{K}\sigma_{\mathcal{K}_i}^2}. \tag{8}$$

Results

The GS lies in the column space of $\mathbf{V}_{\mathcal{K}}$ with $\mathcal{K} = \{10, 20, 30\}$. The sampling budget $M = 10$ and the samples are noisy with additive *i.i.d.* Gaussian noise with $\sigma_{\mathbf{w}}^2 = 0.5$. The mean of $\hat{\mathbf{f}}_{\mathcal{K}}$ is $\mu = \mathbf{1}_{3\times 1}$ and the covariance matrix is $\boldsymbol{\Sigma}_{\hat{\mathbf{f}}_{\mathcal{K}}} = \text{diag}(1, 0.5, 0.1)$.



(e) Centers and radiuses of Gershgorin discs for $\mathbf{V}_{\mathcal{K}}\Sigma_{\hat{\mathbf{f}}_{\mathcal{K}}}\mathbf{V}_{\mathcal{K}}^T$ (blue) and $\mathbf{G}_{\text{B}}(\boldsymbol{\eta}^*)$(red)

(f) NMSE for different algorithms versus different observation time

Acknowledgement: This work was supported by the National Key Research and Development Program of China (No. 213), the Shanghai Municipal Natural Science Foundation (No. 19ZR1404700), and the NSF of China (No. 61501124).