NODE VARYING REGULARIZATION FOR GRAPH SIGNALS

Maosheng Yang, Mario Coutino, Elvin Isufi and Geert Leus

Delft University of Technology, Delft, The Netherlands
E-mails: m.yang-7@student.tudelft.nl; {m.a.coutinominguez-1, e.isufi-1, g.j.t.leus}@tudelft.nl

ABSTRACT

While regularization on graphs has been successful for signal reconstruction, strategies for controlling the bias-variance trade-off of such methods have not been completely explored. In this work, we put forth a node varying regularizer for graph signal reconstruction and develop a minmax approach to design the vector of regularization parameters. The proposed design only requires as prior information an upper bound on the underlying signal energy; a reasonable assumption in practice. With such formulation, an iterative method is introduced to obtain a solution meeting global equilibrium. The approach is numerically efficient and has convergence guarantees. Numerical simulations using real data support the proposed design scheme.

Index Terms— graph signal processing, bias-variance trade-off, graph regularization, graph signal denoising, minmax problems

1. INTRODUCTION

In this work, we focus on solving the following problem

$$\boldsymbol{\omega}^* := \underset{\boldsymbol{\omega}\in\mathcal{W}}{\arg\min}\ f_{\boldsymbol{\omega}}(\boldsymbol{y};\boldsymbol{\mu}), \tag{1}$$

where $\boldsymbol{\omega} \in \mathbb{R}^d$ is a regularization parameter for the loss function $f_{\boldsymbol{\omega}}(\cdot;\cdot)$ w.r.t the data $\boldsymbol{y} \in \mathbb{R}^n$ and an underlying unknown parameter $\boldsymbol{\mu} \in \mathbb{R}^q$. The regularization parameter $\boldsymbol{\omega}$ is within a convex set $\mathcal{W}$.

Problems of the form in (1) arise naturally in applications including, hyper-parameter tuning [1], biased estimators [2], image denoising [3], and signal reconstruction [4], to name a few. Though this formulation is simple, the dependency of $f_{\boldsymbol{\omega}}$ on the underlying unknown parameter $\boldsymbol{\mu}$ impedes a straightforward solution. For the common and simple case with a scalar regularization parameter ω, this dependency problem also happens and is well-studied in the literature, for instance, the author in [5] used SURE method to estimate the term involved with the unknown parameter $\boldsymbol{\mu}$, while in [6] based on subspace information criterion. In the context of Tikhonov regularization, methods based on like, the discrepancy principle [7,8], the L-curve criterion [9] and the generalized cross-validation [10], are used to select the regularization parameter.

However, these methods are designed for scalar regularization parameter selection. Instead of solving (1), we consider a worst case scenario and focus on its *minmax* formulation

$$\boldsymbol{\omega}^* := \underset{\boldsymbol{\omega}\in\mathcal{W}}{\arg\min}\ \max_{\boldsymbol{\mu}\in\mathcal{M}}\ f_{\boldsymbol{\omega}}(\boldsymbol{y};\boldsymbol{\mu}). \tag{2}$$

This work has been submitted to EUSIPCO, 2020, currently under review. This research is supported in part by the ASPIRE project (project 14926 within the STW OTP programme), financed by the Netherlands Organization for Scientific Research (NWO). Mario Coutino is partially supported by CONACYT. Maosheng Yang's study is supported by Microelectronic-MSc track scholarship from TU Delft.

Here, set $\mathcal{M}$ defines the restrictions on the unknown parameter $\boldsymbol{\mu}$ that are often easy to get. For mathematical tractability, $\mathcal{M}$ is considered convex and compact. The study of (2) requires in general assumptions that could be violated in practice. Thus, it is always advisable to restrict this study to loss functions derived from particular problem instances. In our case, we analyze (2) for the problem of *graph signal reconstruction* [11–13]. The goal of this task is simple: *given a set of noisy observations taken over nodes of a graph, to reconstruct the graph signal using the underlying graph structure as prior knowledge*.

Different works have leveraged this problem for graph signal denoising [14–16], interpolation [17] and semi-supervised learning over graphs [18], where the most common regularizer is the so-called Tikhonov regularizer. The common assumption in all these works is that the regularization parameter is scalar. This, in turn, translates into a common factor applied by all nodes of the graph and fails to capture the signal detail in the neighborhood of a node. The specific local information of certain nodes cannot be taken into consideration.

To tackle this issue and improve the reconstruction performance, we consider $\boldsymbol{\omega}$ to be a vector of regularization parameters where each entry is associated to a node of the graph. We call it *node varying regularization* since each node is regularized differently. The proposed generalization matches naturally the form in (2) by particularizing the loss function to the mean squared error. We provide a gradient descent-based algorithm to find the optimal node varying regularization parameters and show its superior performance compared to the scalar regularization.

In the following, we formalize the problem of graph signal reconstruction in Section 2. In Section 3, we develop the node varying regularization problem, while in Section 4 we focus on the minmax design of the regularization parameter. Numerical results with synthetic and real data corroborate our theory in Section 5 and the paper conclusions are drawn in Section 6.

2. GRAPH SIGNAL RECONSTRUCTION

Let $\boldsymbol{y} \in \mathbb{R}^n$ be a vector of measurements taken over an undirected graph of n nodes, where entry $y_i := [\boldsymbol{y}]_i$ is the measurement collected on the ith node. The node measurements are of the form

$$y_i = \mu_i + \epsilon_i, \tag{3}$$

where μ_i is the mean of the ith measurement and ϵ_i is an i.i.d. random Gaussian variable distributed as $\mathcal{N}(0, \sigma_i)$. That is, the measurement vector follows the distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with covariance matrix $\boldsymbol{\Sigma} := \text{diag}(\sigma_1, \ldots, \sigma_n)$. Further, let $\boldsymbol{L} \in \mathbb{R}^{n\times n}$ be the *graph Laplacian* capturing the connectivity between nodes, defined as $\boldsymbol{D} - \boldsymbol{A}$, where $\boldsymbol{D}$ is the degree matrix and $\boldsymbol{A}$ is the adjacency matrix. The Laplacian is symmetric for undirected graphs .

The problem of graph signal reconstruction consists of estimating the noise-free signal $\boldsymbol{\mu}$ from the noisy measurements $\boldsymbol{y}$. A common

approach to solve this problem is to explore the prior information about the signal behavior over the graph [14,15]. If we assume the signal changes slowly over, we can consider the Tikhonov regularization problem

$$\hat{\boldsymbol{\mu}}_\omega := \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\ \|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \omega\cdot\boldsymbol{x}^\top\boldsymbol{L}\boldsymbol{x}, \tag{4}$$

where scalar $\omega > 0$ is the regularization parameter that controls the signal's *smooth variation* over the network topology [15]. Problem (4) is convex and has the closed-form solution

$$\hat{\boldsymbol{\mu}}_\omega = (\boldsymbol{I}+\omega\boldsymbol{L})^{-1}\boldsymbol{y} := \boldsymbol{H}_\omega\boldsymbol{y} \tag{5}$$

where we defined $\boldsymbol{H}_\omega := (\boldsymbol{I}+\omega\boldsymbol{L})^{-1}$ to ease notation. The mean square error (MSE) of the estimate in (5) is

$$\text{MSE}(\hat{\boldsymbol{\mu}}_\omega) = \text{tr}\big((\boldsymbol{I}-\boldsymbol{H}_\omega)^2\boldsymbol{\mu}\boldsymbol{\mu}^\top\big) + \text{tr}\big(\boldsymbol{H}_\omega^2\boldsymbol{\Sigma}\big). \tag{6}$$

By minimizing the MSE over ω, we can find the optimal regularization parameter ω by solving

$$\omega^* := \underset{\omega}{\arg\min}\ \text{MSE}(\hat{\boldsymbol{\mu}}_\omega). \tag{7}$$

Expression (7) is in the form (1) with $f(\boldsymbol{y};\boldsymbol{\mu}) := \text{MSE}(\hat{\boldsymbol{\mu}}_\omega)$, where now $\boldsymbol{\mu}$ is the unknown noise-free measurement to be estimated. To deal with the dependency on the unknown parameter $\boldsymbol{\mu}$ of the MSE, the work in [14] substituted the MSE cost in (7) with an upper bound. Although this approach can serve to design the regularization parameter ω through order-matching, it falls short (i) in instances where each node is weighted differently, since in (4) all nodes are weighted with a common ω; and (ii) to provide a method without spectral knowledge as its optimal design requires full eigendecomposition of the Laplacian matrix. Therefore, in the following, we propose a generalized graph-based penalizer able to capture node heterogeneity while allowing for a computationally efficient design.

3. NODE VARYING GRAPH SIGNAL RECONSTRUCTION

To address the problem of node heterogeneity, we consider a parameter vector $\boldsymbol{\omega}\in\mathbb{R}^n$, where entry ω_i is associated to node i. The node varying equivalent to (4) can be written as

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}} := \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\ \|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \boldsymbol{x}^\top\text{diag}(\boldsymbol{\omega})\boldsymbol{L}\text{diag}(\boldsymbol{\omega})\boldsymbol{x}. \tag{8}$$

The term $\boldsymbol{x}' = \text{diag}(\boldsymbol{\omega})\boldsymbol{x}$ can be seen as each node weighting accordingly its own signal before computing the signal variation $\boldsymbol{x}'^\top\boldsymbol{L}\boldsymbol{x}'$. Further, note that by setting $\boldsymbol{\omega} = \omega\mathbf{1}$, problem (8) specializes to (4).

Define $\boldsymbol{S}_{\boldsymbol{\omega}} := \text{diag}(\boldsymbol{\omega})\boldsymbol{L}\text{diag}(\boldsymbol{\omega}) = \boldsymbol{\omega}\boldsymbol{\omega}^\top\odot\boldsymbol{L}$ as an edge dependent matrix [19], which is positive semidefinite by Schur product theorem [20, p. 14, Thm. VII] and shares the same support with $\boldsymbol{L}$. Problem (8) is convex by construction since $\text{diag}(\boldsymbol{\omega})\boldsymbol{L}\text{diag}(\boldsymbol{\omega})$ is positive semidefinite. By setting the gradient of (8) to zero, the optimal closed-form solution for (8) is

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}} := (\boldsymbol{I}+\boldsymbol{S}_{\boldsymbol{\omega}})^{-1}\boldsymbol{y}. \tag{9}$$

The MSE of the estimate in (9) is now given by

$$\text{MSE}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}}) := \text{tr}\big((\boldsymbol{I}-(\boldsymbol{I}+\boldsymbol{S}_{\boldsymbol{\omega}})^{-1})^2\boldsymbol{\mu}\boldsymbol{\mu}^\top\big) + \text{tr}\big((\boldsymbol{I}+\boldsymbol{S}_{\boldsymbol{\omega}})^{-2}\boldsymbol{\Sigma}\big). \tag{10}$$

It consists of the squared norm of the bias as the first term and the variance as the second term. Likewise the scalar counter part (6), the $\text{MSE}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}})$ depends on the unknown parameter $\boldsymbol{\mu}$. To tackle this dependency and design the regularization parameter $\boldsymbol{\omega}$, we depart from approaches of the form (7) and consider a minmax formulation as in (2). Beside tackling the dependency on the underlying parameter $\boldsymbol{\mu}$, the minmax formulation also avoids working with upper bounds.

4. MINMAX PARAMETER DESIGN

The minmax formulation for the optimal regularization parameter design is

$$\hat{\boldsymbol{\omega}} := \underset{\boldsymbol{\omega}\in\mathcal{W}}{\arg\min}\ \underset{\boldsymbol{\mu}\in\mathcal{M}}{\max}\ f_{\boldsymbol{\omega}}(\boldsymbol{y};\boldsymbol{\mu}), \tag{11}$$

where $f_{\boldsymbol{\omega}}(\boldsymbol{y};\boldsymbol{\mu}) := \text{MSE}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}})$ and where $\mathcal{W}$ and $\mathcal{M}$ are two sets to be specified in the sequel. In a practical setting, no much information is available about the unknown parameter $\boldsymbol{\mu}$; however its energy (norm) is typically bounded. For instance, an energy bound on the measurements $\boldsymbol{y}$ will simply impose a (may not tight) bound on $\boldsymbol{\mu}$; or if the signal-to-noise ratio (SNR) is available, through the knowledge of the noise power, we can obtain a bound on the signal power. Hence, it is reasonable to consider that $\boldsymbol{\mu}$ lies within an ℓ_2-norm ball with radius μ_*, i.e., $\mathcal{M} := \{\boldsymbol{\mu} : \|\boldsymbol{\mu}\|_2 \le \mu_*\}$. Set $\mathcal{M}$ meets all assumptions required by minmax problems, i.e., it is convex and compact. In addition, to preserve the convexity of problem (8), we only require the regularizer parameters to be within the real set, $\mathcal{W} = \mathbb{R}^n$, so that $\text{diag}(\boldsymbol{\omega})\boldsymbol{L}\text{diag}(\boldsymbol{\omega})$ is positive semi-definite.

Before studying the details of (11), let us first analyze the $\text{MSE}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}})$ expression in (10). We observe that only the first term depends on $\boldsymbol{\mu}$. This term captures the squared norm of the bias of the estimator, $\|\text{bias}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}})\|^2$ and can be written in the quadratic form

$$\|\text{bias}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}})\|^2 = \text{tr}(\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}\boldsymbol{\mu}\boldsymbol{\mu}^\top) = \boldsymbol{\mu}^\top\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}\boldsymbol{\mu}, \tag{12}$$

where $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}} = (\boldsymbol{I}-(\boldsymbol{I}+\boldsymbol{S}_{\boldsymbol{\omega}})^{-1})^2$ is a positive definite matrix that depends on $\boldsymbol{\omega}$. For a fixed $\boldsymbol{\omega}$, the bias term accepts a simple maximization when $\boldsymbol{\mu}$ is restricted to set $\mathcal{M}$. That is, it suffices to find the eigenvector $\boldsymbol{\mu}^*$ corresponding to the largest eigenvalue of $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}$, more specifically

$$\boldsymbol{\mu}^* := \underset{\boldsymbol{\mu}\in\mathcal{M}}{\arg\max}\ \boldsymbol{\mu}^\top\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}\boldsymbol{\mu} = \mu_*\lambda_{\max}(\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}), \tag{13}$$

where $\lambda_{\max}(\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}})$ denotes the largest eigenvalue of $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}$ and μ_* is the energy upper bound. The following result ensures that the solution of (13) can be obtained efficiently.

Proposition 1. *Let $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}$ be given as above, then the maximizer of* (13) *is the eigenvector of $\boldsymbol{S}_{\boldsymbol{\omega}}$ related with $\lambda_{\max}(\boldsymbol{S}_{\boldsymbol{\omega}})$.*

Proof. From their definition, both $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}$ and $\boldsymbol{S}_{\boldsymbol{\omega}}$ are positive semidefinite, thus $\boldsymbol{S}_{\boldsymbol{\omega}}+\boldsymbol{I}\succeq\boldsymbol{I}$. As the ith eigenvalue of $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}$ is given as $\lambda_i(\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}) := (1-(1+\lambda_i(\boldsymbol{S}_{\boldsymbol{\omega}}))^{-1})^2$, we conclude that $\max_i\lambda_i(\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}) = \max_i\lambda_i(\boldsymbol{S}_{\boldsymbol{\omega}})$. □

Hence, the solution of (13) can be computed efficiently through power iteration using the matrix $\boldsymbol{S}_{\boldsymbol{\omega}}$ which enjoys the sparsity of the graph Laplacian, instead of $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}$.

Since the inner maximization step can be solved exactly, the outer minimization can be performed with an iterative first-order method [21]. A first-order based method for solving the minmax problem (11) has the update

$$\boldsymbol{\omega}_{t+1} = \boldsymbol{\omega}_t - \eta_t\nabla_{\boldsymbol{\omega}}f_{\boldsymbol{\omega}_t}(\boldsymbol{y};\boldsymbol{\mu}_t^*), \tag{14}$$

where η_t is the step size at iteration t and $\nabla_{\boldsymbol{\omega}}f_{\boldsymbol{\omega}_t}$ is the gradient of f w.r.t $\boldsymbol{\omega}$ evaluated at $\boldsymbol{\omega}_t$. Here, $\boldsymbol{\mu}_t^* := \mu_*\boldsymbol{v}_t$ with $\boldsymbol{v}_t$ being the normalized eigenvector related with the largest eigenvalue of $\boldsymbol{S}_{\boldsymbol{\omega}_t} = \text{diag}(\boldsymbol{\omega}_t)\boldsymbol{L}\text{diag}(\boldsymbol{\omega}_t)$. With derivations in appendix[1], the

[1] Available online at this_link

closed-form expression of the gradient $\nabla_{\boldsymbol{\omega}} f_{\omega_t}(\boldsymbol{y}; \boldsymbol{\mu}_t^*)$ is

$$\nabla_{\boldsymbol{\omega}} f_{\omega_t}(\boldsymbol{y}; \boldsymbol{\mu}_t^*) = \text{diag}^{-1}\Big([-4(\boldsymbol{I} + \boldsymbol{S}_{\omega_t})^{-2}(\boldsymbol{\mu}_t^* \boldsymbol{\mu}_t^{*\top} + \boldsymbol{\Sigma}) + 4(\boldsymbol{I} + \boldsymbol{S}_{\omega_t})^{-1} \boldsymbol{\mu}_t^* \boldsymbol{\mu}_t^{*\top}](\boldsymbol{I} + \boldsymbol{S}_{\omega_t})^{-1} \text{diag}(\boldsymbol{\omega}_t) \boldsymbol{L} \Big). \quad (15)$$

Although it seems that evaluating (15) requires the inversion of some matrices, these operations can be implemented efficiently as the solution to symmetric diagonal dominant (SDD) systems [22, 23]. Algorithm 1 summarizes the minmax procedure for solving (11).

To study the theoretical guarantees of this Algorithm, we introduce the following definition.

Definition 1 (FNE). *A point $(\boldsymbol{\omega}^*; \boldsymbol{\mu}^*)$ is a first-order Nash equilibrium (FNE) of the game* (11) *if*

$$\langle \nabla_{\boldsymbol{\omega}} f(\boldsymbol{\omega}^*; \boldsymbol{\mu}^*), \boldsymbol{\omega} - \boldsymbol{\omega}^* \rangle \geq 0, \ \forall \ \boldsymbol{\omega} \in \mathcal{W} \quad (16)$$

and

$$\langle \nabla_{\boldsymbol{\mu}} f(\boldsymbol{\omega}^*; \boldsymbol{\mu}^*), \boldsymbol{\mu} - \boldsymbol{\mu}^* \rangle \leq 0, \ \forall \ \boldsymbol{\mu} \in \mathcal{M} \quad (17)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product.

This definition guarantees first-order necessary optimality conditions for the objective function (for each player). Hence, they are necessary conditions to guarantee the so-called first-order Nash equilibrium [24], i.e.,

$$f(\boldsymbol{\omega}^*; \boldsymbol{\mu}) \leq f(\boldsymbol{\omega}^*; \boldsymbol{\mu}^*) \leq f(\boldsymbol{\omega}; \boldsymbol{\mu}^*), \ \forall \ \boldsymbol{\omega} \in \mathcal{W}, \ \forall \boldsymbol{\mu} \in \mathcal{M}. \quad (18)$$

As Algorithm 1 is an iterative method and in a practical setting it always has a numerical tolerance, in the following, the notion of approximate-FNE is introduced.

Definition 2 (Approximate FNE). *A point $(\boldsymbol{\omega}^*; \boldsymbol{\mu}^*)$ is an ϵ-first-order Nash equilibrium (ϵ-FNE) of the game* (11) *if*

$$\mathcal{X}(\boldsymbol{\omega}^*; \boldsymbol{\mu}^*) \leq \epsilon \text{ and } \mathcal{Y}(\boldsymbol{\omega}^*; \boldsymbol{\mu}^*) \leq \epsilon, \quad (19)$$

where

$$\mathcal{X}(\boldsymbol{\omega}^*; \boldsymbol{\mu}^*) := -\min_{\boldsymbol{\omega}} \langle \nabla_{\boldsymbol{\omega}} f(\boldsymbol{\omega}^*; \boldsymbol{\mu}^*), \boldsymbol{\omega} - \boldsymbol{\omega}^* \rangle \quad \text{s.t. } \boldsymbol{\omega} \in \mathcal{W}, \|\boldsymbol{\omega} - \boldsymbol{\omega}^*\| \leq 1, \quad (20)$$

and

$$\mathcal{Y}(\boldsymbol{\omega}^*; \boldsymbol{\mu}^*) := \max_{\boldsymbol{\mu}} \langle \nabla_{\boldsymbol{\mu}} f(\boldsymbol{\omega}^*; \boldsymbol{\mu}^*), \boldsymbol{\mu} - \boldsymbol{\mu}^* \rangle \quad \text{s.t. } \boldsymbol{\mu} \in \mathcal{M}, \|\boldsymbol{\mu} - \boldsymbol{\mu}^*\| \leq 1. \quad (21)$$

This definition is based on the first-order optimality measure of the objective of each variable. Such a condition guarantees that each variable cannot improve their objective function by using first-order information, providing a both theoretical and numerically meaningful stopping criteria i.e., convergence criterion. Now, we are ready to state the following result regarding the convergence of Algorithm 1.

Proposition 2 (Convergence). *Let the problem* (13) *have a unique solution for each $\boldsymbol{\omega}_t$. Then, Algorithm 1 is guaranteed to converge to an ϵ-FNE of the game* (11) *for $T, K \to \infty$.*

Algorithm 1 Iterative First-Order Method for MinMax Game (11)

Input: μ^*: signal energy bound; T: number of gradient descent iterations; K: number of power iterations; η: step size; $\boldsymbol{\Sigma}$: noise covariance matrix; $\boldsymbol{L}$: graph Laplacian matrix;
Output: optimized regularization parameter $\boldsymbol{\omega}$

Initialization : $\boldsymbol{\omega}_0 = \omega_0 \mathbf{1}$, $\boldsymbol{\mu}_0 = \mathbf{1}$
for $t = 0$ to $T - 1$ **do**
  $\boldsymbol{S}_{\omega_t} = \text{diag}(\boldsymbol{\omega}_t) \boldsymbol{L} \text{diag}(\boldsymbol{\omega}_t)$
  for $k = 0$ to $K - 1$ **do**
    $\mathbf{z}_k = \boldsymbol{S}_{\boldsymbol{\omega}_t} \boldsymbol{\mu}_k$
    $\boldsymbol{\mu}_{k+1} = \mu^* \mathbf{z}_k / \|\mathbf{z}_k\|_2$
  end for
  $\boldsymbol{\mu}_t^* = \boldsymbol{\mu}_K$, the largest eigenvector of $\boldsymbol{S}_{\omega_t}$
  Compute $\nabla_{\boldsymbol{\omega}} f_{\omega_t}$ as (15) by substituting $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}_t^*$
  $\boldsymbol{\omega}_{t+1} = \boldsymbol{\omega}_t - \eta \nabla_{\boldsymbol{\omega}} f_{\omega_t}(\boldsymbol{y}; \boldsymbol{\mu}_t^*)$
end for
return $\boldsymbol{\omega} = \boldsymbol{\omega}_T$

Proof. (Sketch.) When the stated condition holds, then Danskin's theorem [25] holds, i.e.,

$$\nabla_{\boldsymbol{\omega}} \max_{\boldsymbol{\mu} \in \mathcal{M}} f_{\boldsymbol{\omega}}(\boldsymbol{y}; \boldsymbol{\mu}) = \nabla_{\boldsymbol{\omega}} f_{\boldsymbol{\omega}}(\boldsymbol{y}; \boldsymbol{\mu}^*) \quad (22)$$

with $\boldsymbol{\mu}^* = \arg\max \ f_{\boldsymbol{\omega}}(\boldsymbol{y}; \boldsymbol{\mu})$ and for $K \to \infty$, the power method finds the exact solution to (13). The rest of the proof specializes the result in [21, Thm. 3.4] □

Although the uniqueness condition for problem (13) might seem restrictive, this behaviour is typically observed in practice. Furthermore, even when this is not the case, we can consider a proximal term $\|\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}\|_2$ in (13) to guarantee the convergence to an ϵ-FNE; see [21] for further technical details. Next, we use the proposed method to design a robust worst-case regularizer for node varying graph signal reconstruction.

5. NUMERICAL EXPERIMENTS

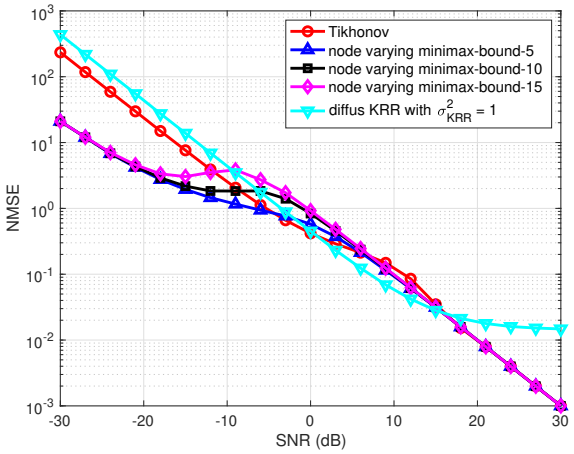
In this section, we corroborate the proposed method with synthetic data on Erdős–Rényi graphs and with real data from the Molene weather dataset[2]. We first obtained an optimal regularization parameter $\boldsymbol{\omega}^*$ by solving the minmax problem (11) with Algorithm 1. Then, we used the found regularization parameter to reconstruct the signal as in (9). We compared our method with two other state-of-the-art approaches:

i) the standard Tikhonov based denoising (4). Based on the bias-variance trade-off study and scaling law in [14], we optimally set the regularization parameter $\omega = \mathcal{O}(\sqrt{\theta/(\lambda_2 \lambda_n)})$, where $\theta = \sqrt{1/\text{SNR}}$, and λ_2 and λ_n are the smallest and the largest non-trivial eigenvalues of the graph Laplacian, respectively.

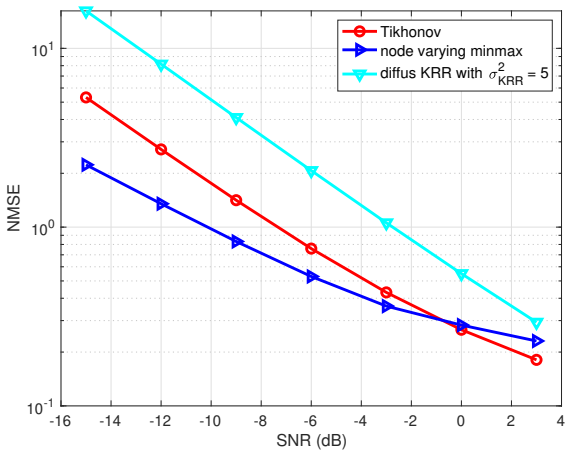
ii) Diffusion kernel-based ridge regression with parameter $\sigma_{KRR}^2 = 1$ (diffusion kernel parameter) and regularization parameter $c = 10^{-4}$, which is studied well in kernel-based graph signal reconstruction [13].

We measured the performance through the normalized mean squared error (NMSE), which is defined as $\text{NMSE} = \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2 / \|\boldsymbol{\mu}\|^2$. In our experiments, we analyzed under different signal-to-noise ratios (SNRs). The true signal is corrupted with white Gaussian noise to

[2] https://donneespubliques.meteofrance.fr/donneeslibres/Hackathon/RADOMEH.tar.gz



(a) Synthetic data.



(b) Molene weather dataset.

Fig. 1: Normalize mean squared error comparison of different methods as a function of the signal-to-noise ratio.

yield an SNR given by SNR $= 10\log_{10}(\|\boldsymbol{\mu}\|^2/(n\sigma^2))$ with n being the number of graph nodes and σ^2 the noise variance.

Synthetic data. We built an Erdős–Rényi graph of $n = 50$ nodes with a connection probability of 0.5. We generated a deterministic smooth graph signal which has an ideal low-pass graph frequency content with bandwidth 20 [15]. We observed the ℓ_2-norm $\|\boldsymbol{\mu}\|_2 = 4.47$. For the gradient descent based method in Algorithm 1, we set the number of iterations $T = 100$, $K = 30$, and the step size $\eta = 0.002$. We then initialized $\boldsymbol{\omega} = \omega_0\mathbf{1}$ with ω_0 being the optimal Tikhonov regularization parameter, and $\boldsymbol{\mu} = \mathbf{1}$. To evaluate the recovery performance in different noisy situations, we considered an SNR in the range $[-30, 30]$ dB. To evaluate the effect of the energy bound on the reconstruction performance, we considered the energy bound μ_* to have three values $\{5, 10, 15\}$, which are all above the true energy. Our results are averaged over 100 noise realizations and 50 graph realizations for a total of 5000 Monte-Carlo runs.

From Fig. 1a, the performance of our minmax formulation stands out. Specifically, with any energy bound –whether a stricter one ($\mu_* = 5$) or a looser one ($\mu_* = 15$)– the proposed method gives better results compared with the other contenders in the low-SNR regimes. In the medium-SNR range, our method with a loose energy bound will behave worse than Tikhonov, but this difference becomes negligible when the energy bound gets tighter. Finally, at high-SNR regime all methods reach a similar performance except for the diffusion kernel-based method that may have a bias. This trend shows that our method generalizes the Tikhonov regularization and indicates that local node detail is more important in low SNR settings.

Molene weather data. This dataset consists of 744 hourly temperature recordings collected in January 2014 over 32 cities in the region of Brest, France. We built the graph from the coordinates of the stations by connecting all the neighbours in a given radius with a weight $\mathbf{W}_G(i,j) = \exp\{-kd^2(i,j)\}$, where $d(i,j)$ is the Euclidean distance between stations i and j, and parameter k is five. We removed the average value of weather data over time and space. For this experiment, we artificially added noise and considered an SNR in the interval -15 to 3 dB. We set the energy bound to the true one plus a trivial deviation, which is a reasonable assumption based on historical same-day recordings. For the diffusion kernel method, we here modified the parameter σ^2_{KRR} to be 5 for a better performance. The other parameters remain the same as in the former experiment.

Fig. 1b shows the performance of the three different methods. This result shows that the proposed node varying approach should be considered in harsher settings. When the SNR improves (i.e., the data matches better the true one) regularization is less needed as it biases the results. However, likewise in the synthetic dataset, our method yields a superior performance in low SNR regimes.

6. CONCLUSIONS

In this paper, we proposed a node varying regularizer for reconstructing graph signals. The method considers a vector of regularizer parameters where each entry is associated to a specific node. As such, it generalizes state-of-the-art regularizers which consider the same scalar parameter for all nodes. To design the regularization that minimizes the MSE, we develop a minmax approach that tackles the dependency issue on the unknown parameter. By levering results from the first-order Nash equilibrium, we provide a gradient-descent algorithm to obtain the optimal regularization parameter vector with convergence guarantees. Numerical results with synthetic and real data corroborate our findings and show the proposed method outperforms the state-of-the-art methods, especially when the signal-to-noise ratio is low.

7. REFERENCES

[1] D. J. Toal, N. W. Bressloff, and A. J. Keane, "Kriging hyperparameter tuning strategies," *AIAA journal*, vol. 46, no. 5, pp. 1240–1252, 2008.

[2] P. Stoica and R. L. Moses, "On biased estimators and the unbiased cramér-rao lower bound," *Signal Processing*, vol. 21, no. 4, pp. 349–350, 1990.

[3] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[4] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.

[5] S. Kay and Y. C. Eldar, "Rethinking biased estimation [lecture notes]," *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 133–136, 2008.

[6] M. Sugiyama and H. Ogawa, "Optimal design of regularization term and regularization parameter by subspace information criterion," *Neural networks : the official journal of the International Neural Network Society*, vol. 15, no. 3, p. 349â361, April 2002. [Online]. Available: https://doi.org/10.1016/s0893-6080(02)00022-9

[7] O. Scherzer, "The use of morozov's discrepancy principle for tikhonov regularization for solving nonlinear ill-posed problems," *Computing*, vol. 51, no. 1, pp. 45–60, 1993. [Online]. Available: https://doi.org/10.1007/BF02243828

[8] S. W. Anzengruber and R. Ramlau, "Morozovs discrepancy principle for tikhonov-type functionals with nonlinear operators," *Inverse Problems*, vol. 26, no. 2, p. 025001, Dec 2009.

[9] P. C. Hansen and D. P. O'Leary, "The use of the l-curve in the regularization of discrete ill-posed problems," *SIAM Journal on Scientific Computing*, vol. 14, no. 6, pp. 1487–1503, 1993. [Online]. Available: https://doi.org/10.1137/0914086

[10] G. H.Golub and U. von Matt, "Generalized cross-validation for large-scale problems," *Journal of Computational and Graphical Statistics*, vol. 6, no. 1, pp. 1–34, 1997. [Online]. Available: https://amstat.tandfonline.com/doi/abs/10.1080/10618600.1997.10474725

[11] V. Cevher, P. Indyk, L. Carin, and R. G. Baraniuk, "Sparse signal recovery and acquisition with graphical models," *IEEE Signal Processing Magazine*, vol. 27, no. 6, pp. 92–103, 2010.

[12] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *IEEE Transactions on Signal Processing*, vol. 64, no. 16, pp. 4363–4378, 2016.

[13] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 764–778, Feb 2017.

[14] P. Chen and S. Liu, "Bias-variance tradeoff of graph laplacian regularizer," *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1118–1122, Aug 2017.

[15] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, May 2013.

[16] J. Pang, G. Cheung, A. Ortega, and O. C. Au, "Optimal graph laplacian regularization for natural image denoising," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 2294–2298.

[17] S. K. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 5445–5449.

[18] A. Gadde, A. Anis, and A. Ortega, "Active semi-supervised learning using sampling theory for graph signals," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 492–501.

[19] M. Coutino, E. Isufi, and G. Leus, "Advances in distributed graph filtering," *IEEE Transactions on Signal Processing*, vol. 67, no. 9, pp. 2320–2333, May 2019.

[20] J. Schur, "Bemerkungen zur theorie der beschränkten bilinearformen mit unendlich vielen veränderlichen." *Journal für die reine und angewandte Mathematik*, vol. 140, pp. 1–28, 1911. [Online]. Available: http://eudml.org/doc/149352

[21] M. Nouiehed, M. Sanjabi, T. Huang, J. D. Lee, and M. Razaviyayn, "Solving a class of non-convex min-max games using iterative first order methods," in *Advances in Neural Information Processing Systems*, 2019, pp. 14 905–14 916.

[22] J. A. Kelner, L. Orecchia, A. Sidford, and Z. A. Zhu, "A simple, combinatorial algorithm for solving sdd systems in nearly-linear time," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, 2013, pp. 911–920.

[23] D. A. Spielman and S.-H. Teng, "Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems," in *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, 2004, pp. 81–90.

[24] C. Jin, P. Netrapalli, and M. I. Jordan, "Minmax optimization: Stable limit points of gradient descent ascent are locally optimal," *arXiv preprint arXiv:1902.00618*, 2019.

[25] J. M. Danskin, "The theory of max-min, with applications," *SIAM Journal on Applied Mathematics*, vol. 14, no. 4, pp. 641–664, 1966. [Online]. Available: https://doi.org/10.1137/0114053