MATRIX COMPLETION OF NOISY GRAPH SIGNALS VIA PROXIMAL GRADIENT MINIMIZATION

Pere Giménez-Febrer, Alba Pagès-Zamora

SPCOM Group, Universitat Politècnica de Catalunya-Barcelona Tech, Spain

Abstract—This paper takes on the problem of recovering the missing entries of an incomplete matrix, which is known as matrix completion, when the columns of the matrix are signals that lie on a graph and the available observations are noisy. We solve a version of the problem regularized with the Laplacian quadratic form by means of the proximal gradient method, and derive theoretical bounds on the recovery error. Moreover, in order to speed up the convergence of the proximal gradient, we propose an initialization method that utilizes the structural information contained in the Laplacian matrix of the graph.

Index Terms—matrix completion, signal processing on graphs, proximal gradient

I. INTRODUCTION

In [1], Candès and Retch formulated the matrix completion problem as a convex minimization of the nuclear norm of the partially observed matrix, and demonstrated that an exact recovery of the original matrix is possible whenever the matrix is low rank, incoherent and uniformly sampled. The same framework is used in [2] for the recovery from noisy observations. The assumptions made imply that the matrix entries are unstructured, which is usually not the case for real data. Therefore, some works have extended the work in [1] by including extra information about hidden matrix structures into the problem formulation. For instance, in [3] an additional restriction is imposed so that the columns of the recovered matrix are a linear combination of the basis elements in a dictionary. In [4], it is observed that the temperature measurements taken by the sensors in a wireless sensor network are temporally stable in the short term, so a regularization term is added to ensure the short term stability of the recovered data. In the problem of predicting an incomplete matrix of ratings in [5], groups of users with similar background are assumed to have similar preferences, thus a penalty term is included to reduce the variability of the predicted ratings within a group.

An alternative approach to capture the interdependence between the entries in the matrix is by applying techniques from the field of signal processing on graphs. As described in [6], [7], this field extends classical signal processing tools such as filtering or domain transformations and applies them to signals on a graph, so that the vertices in the graph measure signals and the graph edges model the underlying relational structure of those signals. The connections between the vertices are represented by the weighted adjacency matrix of the graph, which can be known beforehand or inferred from the data as, for instance, in [8]. With regards to the matrix completion problem, so far the existing works under the signal processing on graphs perspective take advantage of the fact that the graph signals are known to be smooth on a given graph. For instance, in [9] the signals are assumed to be smooth on a graph when the data from connected vertices have similar values. Hence, the Laplacian quadratic form is added as a regularization function to the problem in [1] to enforce this smoothness on the recovered data. Additionally, a function named total variation is used in [10] as a regularization term to enforce the graph structure described by the adjacency matrix. Similarly, [11] takes on the collaborative filtering problem assuming that the original matrix can be factorized into two smaller matrices which lie on two different graphs.

In this paper, we focus on the recovery of partially observed graph signals that are arranged in a data matrix when the observations are noisy. We solve the problem by adding the Laplacian quadratic form as a regularization term as in [9], [11], and using the proximal gradient (PG) method, which allows us to derive theoretical bounds on the recovery error and give insight into the effect of the regularization. Moreover, we also propose an initialization method to reduce the number of iterations required to reach a solution. The remaining of this paper is organized as follows. Section II introduces the matrix completion problem, the PG algorithm, and the proposed initialization method. Section III analyzes the recovery error of the PG. Finally, Section IV includes the simulation results.

II. MATRIX COMPLETION FOR GRAPH SIGNALS

Consider the undirected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{A})$, where $\mathcal{V} = \{v_1, \ldots, v_N\}$ is the set of vertices, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges connecting the vertices, and $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ is the weighted adjacency matrix. This matrix is defined so that a non-zero entry $\boldsymbol{A}_{i,j}$ indicates that information flows from vertex j to vertex i, and the magnitude of the entries is a measure of similarity or dependence between the connected vertices. Given $\mathcal{G}$, a graph signal is defined as a map from the set of $\mathcal{V}$ vertices into a set of real numbers $y : \mathcal{V} \to \mathbb{R}$ that can be arranged in a column vector $\boldsymbol{y}$ defined as

$$\boldsymbol{y} = [y(1), \ldots, y(N)]^T,$$

where $y(i)$ is the value of the signal at vertex v_i.

Consider now a set of graph signals $\{\boldsymbol{y}_l, \forall l = 1, \ldots, L\}$ residing on the graph $\mathcal{G}$, and let $\boldsymbol{Y}$ be the $N \times L$ matrix of graph signals

$$\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_L].$$

Matrix $\boldsymbol{Y}$ is assumed low rank and cannot be observed except for a subset $\Omega \subset \{1, \ldots, N\} \times \{1, \ldots, L\}$ of entries, which are corrupted by noise so that the matrix of observed entries $\boldsymbol{M} \in \mathbb{R}^{N \times L}$ is defined as

$$\boldsymbol{M}_{i,j} = \begin{cases} \boldsymbol{Y}_{i,j} + \boldsymbol{\Delta}_{i,j}, & \text{when } (i,j) \in \Omega, \\ 0, & \text{when } (i,j) \notin \Omega, \end{cases}$$

where $\boldsymbol{\Delta}$ is a sparse noise matrix with nonzero entries $\boldsymbol{\Delta}_{i,j} \forall (i,j) \in \Omega$. For convenience, we also define the complement subset of Ω as Ω^c, the cardinality of a subset as $|\cdot|$, and the operator $(\cdot)^\Omega$ which projects a matrix $\boldsymbol{X}$ onto the subset Ω as follows

$$(\boldsymbol{X}^\Omega)_{i,j} = \begin{cases} \boldsymbol{X}_{i,j}, & \text{when } (i,j) \in \Omega, \\ 0, & \text{when } (i,j) \notin \Omega. \end{cases}$$

Thus, our objective is to recover the original matrix $\boldsymbol{Y}$ from the observed noisy entries in $\boldsymbol{M}$.

According to [2], one can recover $\boldsymbol{Y}$ with an error proportional to the noise by solving the following convex optimization problem:

$$\hat{\boldsymbol{Y}} = \underset{\boldsymbol{X}}{\operatorname{argmin}} \quad ||\boldsymbol{X}||_* \tag{1}$$
$$\text{subject to} \quad ||\boldsymbol{X}^\Omega - \boldsymbol{M}||_{\mathrm{F}} \le ||\boldsymbol{\Delta}||_{\mathrm{F}}$$

where $||\cdot||_*$ denotes the nuclear norm defined by $||\boldsymbol{X}||_* = \sum_{i=1}^r \sigma_i$, with σ_i being the singular values of $\boldsymbol{X}$ and r its rank, and $||\cdot||_{\mathrm{F}}$ is the Frobenius norm.

We can reduce the impact of the noise and take advantage of the structural information provided by the graph by incorporating its Laplacian matrix into the matrix completion problem. The quadratic Laplacian form measures the smoothness of a signal on the graph and is defined as

$$\mathcal{L}(\boldsymbol{X}) = \operatorname{Tr}(\boldsymbol{X}^T \boldsymbol{L} \boldsymbol{X}) = \sum_{l=1}^L \sum_{(i,j) \in \mathcal{E}} \boldsymbol{A}_{i,j} (\boldsymbol{X}_{i,l} - \boldsymbol{X}_{j,l})^2,$$

where $\boldsymbol{L} = \operatorname{diag}(\boldsymbol{A}\boldsymbol{1}) - \boldsymbol{A}$ is the Laplacian matrix, $\boldsymbol{1}$ is the all-ones vector, and $\operatorname{Tr}(\cdot)$ denotes the matrix trace operation. The matrix completion problem is commonly [10], [12], [13] solved by minimizing the Lagrangian form of (1) as

$$\underset{\boldsymbol{X}}{\operatorname{argmin}} \frac{1}{2} ||\boldsymbol{X}^\Omega - \boldsymbol{M}||_{\mathrm{F}}^2 + \beta ||\boldsymbol{X}||_* + \alpha \mathcal{L}(\boldsymbol{X}), \tag{2}$$

where the nuclear norm becomes a regularization term with regularization parameter β, and we have incorporated the Laplacian quadratic form as an additional term weighted by a parameter α in order to promote similarity between the edges connected on the graph.

Proximal gradient is a popular method to solve nonsmooth optimization problems such as (2) thanks to its ease of implementation, and the good results it offers [10]. As proven in [14], the solution to (2) is given by the following iterative

scheme which performs a proximal gradient minimization for $k \ge 1$:

$$\boldsymbol{X}_k = S_{t\beta}(\boldsymbol{X}_{k-1} - t(\boldsymbol{X}_{k-1}^\Omega - \boldsymbol{M} + 2\alpha \boldsymbol{L} \boldsymbol{X}_{k-1})). \tag{3}$$

The parameter t is the step size of the gradient, and $S_{t\beta}(\cdot)$ is the matrix shrinkage operator (MSO), which, given the singular value decomposition (SVD) of a matrix $\boldsymbol{X}$ as $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}$, we define as

$$S_{t\beta}(\boldsymbol{X}) = \boldsymbol{U}(\boldsymbol{\Sigma} - t\boldsymbol{B}_{\boldsymbol{X}})\boldsymbol{V}.$$

The symbol $\boldsymbol{B}_{\boldsymbol{X}}$ denotes a diagonal $N \times L$ matrix with main diagonal

$$\boldsymbol{b}_{\boldsymbol{X}} = [\beta, \ldots, \beta, \frac{\sigma_d}{t}, \ldots, \frac{\sigma_D}{t}],$$

where $D = \min(N, L)$, and $\{\sigma_d, \ldots, \sigma_D\}$ are the singular values of $\boldsymbol{X}$ smaller than $t\beta$, that is, $\sigma_i < t\beta \ \forall i \ge d$.

A. Proximal gradient initialization

Algorithms for matrix completion are usually initialized to the observed matrix $\boldsymbol{M}$, a low-rank approximation of $\boldsymbol{M}$ or a random matrix [15]. Here we propose to take advantage of the information provided by the Laplacian matrix to compute an initial point $\boldsymbol{X}_0$, hopefully closer to the optimal solution, that will reduce the number of iterations required by the PG algorithm to converge.

One of the properties of the Laplacian matrix is that its smallest eigenvalue has value 0 with multiplicity equal to the number of connected components on the graph. Let matrix $\boldsymbol{L}$ have eigenvalue decomposition $\boldsymbol{L} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$, and $\boldsymbol{Q}_0 \in \mathbb{R}^{N \times P}$ be the matrix containing the P eigenvectors with associated eigenvalue 0. Indeed, $\boldsymbol{Q}_0$ contains the graph signals that are smoothest on the graph so that $\mathcal{L}(\boldsymbol{Q}_0) = 0$. Thus, as the initial matrix $\boldsymbol{X}_0$, we propose to use matrix $\boldsymbol{M}$ and complete the zero entries using a weighted combination of the signals in $\boldsymbol{Q}_0$. Specifically, the initialization matrix is obtained as

$$\boldsymbol{X}_0 = (\boldsymbol{Q}_0 \boldsymbol{C}_0)^{\Omega^c} + \boldsymbol{M}, \tag{4}$$

where $\boldsymbol{C}_0 \in \mathbb{R}^{P \times L}$ is the matrix of coefficients that better fits $\boldsymbol{Q}_0 \boldsymbol{C}_0$ to the observed entries, that is,

$$\boldsymbol{C}_0 = \underset{\boldsymbol{C}}{\operatorname{argmin}} ||\boldsymbol{M} - (\boldsymbol{Q}_0 \boldsymbol{C})^\Omega||_F^2. \tag{5}$$

If we denote $\boldsymbol{C}_0 = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_L]$, the problem in (5) is equivalent to solving

$$\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_L\} = \underset{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_L}{\operatorname{argmin}} \sum_{l=1}^L ||\bar{\boldsymbol{m}}_l - \boldsymbol{S}_l \boldsymbol{Q}_0 \boldsymbol{c}_l||_2^2, \tag{6}$$

where $\bar{\boldsymbol{m}}_l$ is a vector of length e_l containing the observed samples of the lth column of $\boldsymbol{M}$, and $\boldsymbol{S}_l$ is a $e_l \times N$ sampling matrix. Matrix $\boldsymbol{S}_l$ has a single non-zero element per row equal to 1, and it is built so that $\bar{\boldsymbol{m}}_l = \boldsymbol{S}_l \boldsymbol{m}_l$. The solution to (6) is

$$\boldsymbol{c}_l = (\boldsymbol{Q}_0^T \boldsymbol{S}_l^T \boldsymbol{S}_l \boldsymbol{Q}_0)^{-1} \boldsymbol{Q}_0^T \boldsymbol{S}_l^T \bar{\boldsymbol{m}}_l \ \forall l = 1, \ldots, L.$$

The computational cost of this initialization is $O(N^3)$ due to the eigendecomposition of $\boldsymbol{A}$, which is comparable to the cost of one PG iteration.

III. ERROR ANALYSIS

In this section we analyze the recovery error of the PG algorithm and how the addition of the Laplacian quadratic form as a regularization term impacts on the noise. Before introducing the main results, we first make and define the following assumption and lemmas:

Assumption 1. *Similar to Assumption 7 in [13], we assume that, given small enough α and t, there exists a constant $0 \leq \gamma < 1$ such that for any matrix $\boldsymbol{X}$*

$$\left|\left|(\boldsymbol{I} - 2\alpha t \boldsymbol{L})\boldsymbol{X} - t\boldsymbol{X}^{\Omega}\right|\right|_F \leq \gamma \left|\left|\boldsymbol{X}\right|\right|_F.$$

Lemma 1. *Given a pair of matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$,*

$$\left|\left|S_{t\beta}(\boldsymbol{X}) - S_{t\beta}(\boldsymbol{Z})\right|\right|_F \leq \left|\left|\boldsymbol{X} - \boldsymbol{Z}\right|\right|_F.$$

Proof. The proof of this lemma can be found in [14]. $\quad\square$

Lemma 2. *For any matrix $\boldsymbol{X}$*

$$\left|\left|S_{t\beta}((\boldsymbol{I} - 2\alpha t \boldsymbol{L})\boldsymbol{X}) - \boldsymbol{X}\right|\right|_F \leq \sqrt{r}t\beta + 2\alpha t \left|\left|\boldsymbol{L}\boldsymbol{X}\right|\right|_F. \quad (7)$$

Proof. We begin by showing that

$$
\begin{aligned}
&\left|\left|S_{t\beta}((\boldsymbol{I} - 2\alpha t \boldsymbol{L})\boldsymbol{X}) - \boldsymbol{X}\right|\right|_{\mathrm{F}} \\
&= \left|\left|S_{t\beta}((\boldsymbol{I} - 2\alpha t \boldsymbol{L})\boldsymbol{X}) - \boldsymbol{X} + 2\alpha t \boldsymbol{L}\boldsymbol{X} - 2\alpha t \boldsymbol{L}\boldsymbol{X}\right|\right|_{\mathrm{F}} \\
&\leq \left|\left|S_{t\beta}((\boldsymbol{I} - 2\alpha t \boldsymbol{L})\boldsymbol{X}) - (\boldsymbol{I} - 2\alpha t \boldsymbol{L})\boldsymbol{X}\right|\right|_{\mathrm{F}} + 2\alpha t \left|\left|\boldsymbol{L}\boldsymbol{X}\right|\right|_{\mathrm{F}}.
\end{aligned}
$$

Next, let us define matrix $\boldsymbol{X}' = (\boldsymbol{I} - 2\alpha t \boldsymbol{L})\boldsymbol{X}$ with SVD $\boldsymbol{X}' = \boldsymbol{U}'\boldsymbol{\Sigma}'\boldsymbol{V}'$. Then,

$$
\begin{aligned}
\left|\left|S_{t\beta}((\boldsymbol{I} - 2\alpha t \boldsymbol{L})\boldsymbol{X}) - \boldsymbol{X}\right|\right|_{\mathrm{F}} &\leq \left|\left|S_{t\beta}(\boldsymbol{X}') - S_{t\beta}(\boldsymbol{X}' + t\beta\boldsymbol{I})\right|\right|_{\mathrm{F}} \\
&\quad + 2\alpha t \left|\left|\boldsymbol{L}\boldsymbol{X}\right|\right|_{\mathrm{F}}.
\end{aligned}
$$

Finally, we apply Lemma 1 and obtain (7). $\quad\square$

We now introduce the following theorem which bounds the recovery error of the original matrix when the observed entries are noiseless:

Theorem 1. *Let $\hat{\boldsymbol{Y}}$ be the rank r estimate obtained with (3) after the PG has converged, with a noiseless observed matrix $\boldsymbol{Y}^{\Omega}$. Then*

$$\left|\left|\hat{\boldsymbol{Y}} - \boldsymbol{Y}\right|\right|_F^2 \leq \frac{\sqrt{r}t\beta + 2\alpha t \left|\left|\boldsymbol{L}\boldsymbol{Y}\right|\right|_F}{1 - \gamma}.$$

Proof. Since $f(\boldsymbol{X}) = \frac{1}{2}\left|\left|\boldsymbol{X}^{\Omega} - \boldsymbol{Y}^{\Omega}\right|\right|_{\mathrm{F}}^2 + \alpha\mathcal{L}(\boldsymbol{X})$ is convex and continuously differentiable, the proximal gradient converges to a solution in the optimal solution set of (2) which is also a fixed point of (3) [12]. Hence, since $\hat{\boldsymbol{Y}}$ is a fixed point of (3), we have that

$$
\begin{aligned}
\left|\left|\hat{\boldsymbol{Y}} - \boldsymbol{Y}\right|\right|_{\mathrm{F}} &= \left|\left|S_{t\beta}(\hat{\boldsymbol{Y}} - t(\hat{\boldsymbol{Y}}^{\Omega} - \boldsymbol{Y}^{\Omega} + 2\alpha \boldsymbol{L}\hat{\boldsymbol{Y}})) - \boldsymbol{Y}\right|\right|_{\mathrm{F}} \\
&= \left|\left|S_{t\beta}((\boldsymbol{I} - 2\alpha t \boldsymbol{L})\hat{\boldsymbol{Y}} - t(\hat{\boldsymbol{Y}}^{\Omega} - \boldsymbol{Y}^{\Omega})) - S_{t\beta}((\boldsymbol{I} - 2\alpha t \boldsymbol{L})\boldsymbol{Y}) \right.\right. \\
&\quad \left.\left. + S_{t\beta}((\boldsymbol{I} - 2\alpha t \boldsymbol{L})\boldsymbol{Y}) - \boldsymbol{Y}\right|\right|_{\mathrm{F}}. \\
&\leq \left|\left|S_{t\beta}((\boldsymbol{I} - 2\alpha t \boldsymbol{L})\hat{\boldsymbol{Y}} - t(\hat{\boldsymbol{Y}}^{\Omega} - \boldsymbol{Y}^{\Omega})) - S_{t\beta}((\boldsymbol{I} - 2\alpha t \boldsymbol{L})\boldsymbol{Y})\right|\right|_{\mathrm{F}} \\
&\quad + \left|\left|S_{t\beta}((\boldsymbol{I} - 2\alpha t \boldsymbol{L})\boldsymbol{Y}) - \boldsymbol{Y}\right|\right|_{\mathrm{F}}.
\end{aligned}
$$

Applying Lemma 1:

$$
\begin{aligned}
\left|\left|\hat{\boldsymbol{Y}} - \boldsymbol{Y}\right|\right|_{\mathrm{F}} &\leq \left|\left|(\boldsymbol{I} - 2\alpha t \boldsymbol{L})(\hat{\boldsymbol{Y}} - \boldsymbol{Y}) - t(\hat{\boldsymbol{Y}} - \boldsymbol{Y})^{\Omega}\right|\right|_{\mathrm{F}} \\
&\quad + \left|\left|S_{t\beta}((\boldsymbol{I} - 2\alpha t \boldsymbol{L})\boldsymbol{Y}) - \boldsymbol{Y}\right|\right|_{\mathrm{F}}.
\end{aligned}
$$

Next, we apply Lemma 2 and Assumption 1 and obtain

$$\left|\left|\hat{\boldsymbol{Y}} - \boldsymbol{Y}\right|\right|_{\mathrm{F}} \leq \gamma \left|\left|\hat{\boldsymbol{Y}} - \boldsymbol{Y}\right|\right|_{\mathrm{F}} + \sqrt{r}t\beta + 2\alpha t \left|\left|\boldsymbol{L}\boldsymbol{Y}\right|\right|_{\mathrm{F}},$$

which leads to the error bound in Theorem 1. $\quad\square$

Intuitively, Theorem 1 shows that the recovery error with noiseless observations can be reduced by setting a lower β, as it was anticipated in [16], as well as by choosing a graph on which the graph signals in $\boldsymbol{Y}$ are smooth. Nevertheless, whether the error can be driven to 0 by having $\beta \to 0$ depends on the number of observed entries and their distribution, which is implicit through the constant γ.

In order to assess the effect of the regularization on the noise, let us switch to another perspective. In the signal processing on graphs theory [6], [7], matrix $\boldsymbol{Q}^T$ obtained from the eigendecomposition of $\boldsymbol{L}$ is referred to as the Graph Fourier Transform (GFT) matrix. The eigenvectors in $\boldsymbol{Q}$ can be viewed as frequencies on the graph, with the eigenvectors associated to the smaller eigenvalues corresponding to the lower frequencies, that is, signals that vary slowly on the graph. Given a vector $\boldsymbol{x}$, the product $\boldsymbol{L}\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T\boldsymbol{x}$ projects the vector onto the graph frequency domain of $\boldsymbol{L}$, scales the frequency coefficients by $\boldsymbol{\Lambda}$, and projects the result back to the graph domain given by $\boldsymbol{Q}$. Therefore, the result of $\boldsymbol{L}\boldsymbol{x}$ is a highpass filtered version of the vector since the lowest frequency coefficients in $\boldsymbol{Q}^T\boldsymbol{x}$ associated with the smallest eigenvalue, which is equal to 0, are eliminated and the rest are scaled according to $\boldsymbol{\Lambda}$. Then, we can introduce the following theorem:

Theorem 2. *Given $\boldsymbol{I}_0 = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\dagger}$, where $\boldsymbol{\Lambda}^{\dagger}$ is the pseudoinverse of $\boldsymbol{\Lambda}$, the frequency content of $\hat{\boldsymbol{Y}}$ on the nonzero frequencies of the graph is bounded as*

$$
\begin{aligned}
\left|\left|\boldsymbol{I}_0 \boldsymbol{Q}^T \hat{\boldsymbol{Y}}\right|\right|_F &\leq \frac{1}{2\alpha}\left(\left|\left|\boldsymbol{\Lambda}^{\dagger}\boldsymbol{Q}^T(\boldsymbol{Y}^{\Omega} - \hat{\boldsymbol{Y}}^{\Omega})\right|\right|_F \right. \quad (8) \\
&\quad \left. + \left|\left|\boldsymbol{\Lambda}^{\dagger}\boldsymbol{Q}^T \boldsymbol{\Delta}\right|\right|_F + \left|\left|\boldsymbol{\Lambda}^{\dagger}\boldsymbol{Q}^T \boldsymbol{U}' \boldsymbol{B}_{\hat{\boldsymbol{Y}}'} \boldsymbol{V}')\right|\right|_F\right),
\end{aligned}
$$

where matrices $\boldsymbol{U}', \boldsymbol{V}'$ are the left and right singular vector matrices of

$$\hat{\boldsymbol{Y}}' = \hat{\boldsymbol{Y}} - t(\hat{\boldsymbol{Y}}^{\Omega} - \boldsymbol{M}) - 2\alpha t \boldsymbol{L}\hat{\boldsymbol{Y}}. \quad (9)$$

Proof. Let matrix $\hat{\boldsymbol{Y}}'$ in (9) have SVD $\hat{\boldsymbol{Y}}' = \boldsymbol{U}'\boldsymbol{\Sigma}'\boldsymbol{V}'$. Given an estimate $\hat{\boldsymbol{Y}}$, which is a fixed point of (3), we have that $S_{t\beta}(\hat{\boldsymbol{Y}}') = \hat{\boldsymbol{Y}}$. Expanding the MSO in this last expression we obtain

$$\hat{\boldsymbol{Y}}' - t\boldsymbol{U}'\boldsymbol{B}_{\hat{\boldsymbol{Y}}'}\boldsymbol{V}' = \hat{\boldsymbol{Y}}.$$

Now, if we replace $\hat{\boldsymbol{Y}}'$ by its definition, substitute $\boldsymbol{M} = \boldsymbol{Y}^{\Omega} + \boldsymbol{\Delta}$ and rearrange the resulting equation, we have

$$2\alpha \boldsymbol{L}\hat{\boldsymbol{Y}} = \boldsymbol{Y}^{\Omega} - \hat{\boldsymbol{Y}}^{\Omega} + \boldsymbol{\Delta} - \boldsymbol{U}'\boldsymbol{B}_{\hat{\boldsymbol{Y}}'}\boldsymbol{V}'. \quad (10)$$

Note that if we multiply the matrix $\boldsymbol{L}\hat{\boldsymbol{Y}}$ by $\boldsymbol{\Lambda}^{\dagger}\boldsymbol{Q}^T$ we are left with a GFT matrix of the columns of $\hat{\boldsymbol{Y}}$, containing only

the coefficients of the nonzero frequencies. Hence, multiplying both sides of (10) by $\boldsymbol{\Lambda}^\dagger \boldsymbol{Q}^T$ we obtain the bound in (8). $\quad\square$

Theorem 2 shows that the Laplacian regularization has a lowpass filtering effect on the solution since the right-hand side of (8) vanishes with α. Moreover, the second term on the right-hand side of (8) also shows the filtering effect on the noise $\boldsymbol{\Delta}$ since the coefficients of its GFT are multiplied by $\boldsymbol{\Lambda}^\dagger$. If the noise is not smooth on the graph, that is, $\mathcal{L}(\boldsymbol{\Delta})$ is large and the GFT coefficients are concentrated around the higher frequencies, its contribution to the spectrum of $\hat{\boldsymbol{Y}}$ will be small.

IV. SIMULATION RESULTS

We have tested the PG algorithm for the Laplacian-regularized matrix completion problem on a real dataset of temperature measurements. The algorithm has been run until convergence over $N_{rea} = 20$ realizations with different percentages of observed samples, denoted by $P_s = \frac{|\Omega|}{N \cdot L} \cdot 100$. The parameters of the PG are $\beta = 200$, which is set relative to the singular values of the observed dataset, $t = 0.05$ and $\alpha = \{0, 0.5\}$. As a performance metric, we use the normalized mean square error at iteration k

$$\text{NMSE}_k = \frac{1}{N_{rea}} \sum_n^{N_{rea}} \frac{||\boldsymbol{X}_k^n - \boldsymbol{Y}||_\text{F}^2}{||\boldsymbol{Y}||_\text{F}^2},$$

where $\boldsymbol{X}_k^{(n)}$ is the estimate of the complete matrix $\boldsymbol{Y}$ at iteration k and realization n. The initial estimates $\{\boldsymbol{X}_k^{(n)} \, \forall n\}$ are calculated using (4).

The dataset $\boldsymbol{Y}$ is a 150×365 matrix of temperature readings taken by 150 stations over 365 days in 2002 in the United States. The graph signals $\boldsymbol{y}_l$ are the temperature values measured at each station. In order to obtain the weighted adjacency matrix $\boldsymbol{A}$, first a graph $\mathcal{G}'$ with unweighted adjacency matrix $\boldsymbol{P}'$ is generated as in [10] for the stations. In this graph, each station is a vertex and is connected to the 8 geographically closest stations. Next, we obtain the undirected graph $\mathcal{G}$ with symmetric adjacency matrix $\boldsymbol{P} = \text{sign}(\boldsymbol{P}'^T + \boldsymbol{P}')$. Finally, the entries of $\boldsymbol{A}$ are calculated as $\boldsymbol{A}_{i,j} = \exp(-\frac{N^2 d_{i,j}}{\sum_{i,j} d_{i,j}})$, where $d_{i,j}$ are the geodesic distances on $\mathcal{G}$. The noisy matrix $\boldsymbol{M}$ is generated by adding Gaussian noise to the observed entries, which are selected uniformly at random in each realization.

Fig. 1 shows the NMSE after convergence for different percentages of samples and parameter values. We observe that in the noiseless case the NMSE is reduced as the percentage increases. Moreover, the PG shows a larger error with $\alpha = 0.5$ than with $\alpha = 0$. This is because there are enough samples and β is small enough to allow the non-regularized PG to attain a low recovery error. In the noisy case, Fig. 1 shows that the error also decreases with P_s for $\alpha = 0.5$, although it rises for $\alpha = 0$ after $P_s = 25\%$. This is due to the absence of regularization, which causes the PG to overfit to the noisy entries thus resulting in a larger error as more observations and noise are added.

Fig. 2 shows the evolution of the NMSE for $P_s = 30\%$ for the first 200 iterations for noisy and noiseless observations with different parameter values. We observe that the starting point of the PG with noiseless observations is fairly close to
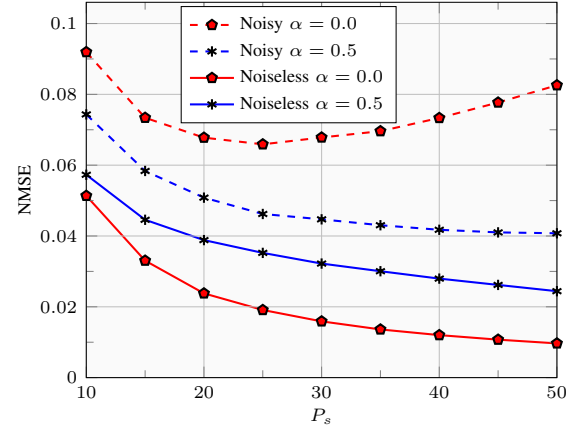

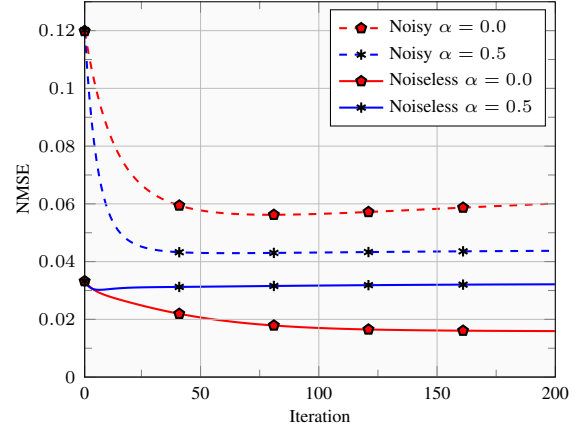Figure 1: NMSE vs. P_s for noisy and noiseless observations.

Figure 2: NMSE vs. iterations with $P_s = 30\%$ for noisy and noiseless observations.

the minimum since the initial estimate calculated with (4) has a very low NMSE equal to 0.033. For comparison, if we initialize to matrix $\boldsymbol{M}$ the NMSE is 0.699 and it takes 400 iterations to converge with $\alpha = 0$. For $\alpha = 0.5$, the PG converges to a point close to the initial error. In the noisy case, we observe a smaller error at convergence for $\alpha = 0.5$ than for $\alpha = 0$, which converges to NMSE $= 0.069$ (see Fig. 1), since the graph regularization helps filter out some of the noise. Moreover, the regularization also increases the convergence speed with noisy observations since the PG with $\alpha = 0$ needs 1000 iterations to converge at $P_s = 30\%$. Regarding the convergence speed in all the scenarios, it should be noted that there exist implementations of the PG such as the one in [14] that accelerate the convergence.

From these simulations we can conclude that, if the random sampling is able to capture the underlying structure of the dataset and β is small enough, the Laplacian regularization does not reduce the recovery error when the observations are noiseless. On the other hand, with noisy observations the regularization improves the error for datasets which are smooth on the graph since it helps filter out the noise, prevents the overfitting to the observed noisy entries, and it also reduces the iterations to convergence. Finally, the proposed initialization provides a good starting point much closer to the minimum than the observed matrix.

REFERENCES

[1] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[2] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

[3] G. Liu and P. Li, "Advancing Matrix Completion by Modeling Extra Structures beyond Low-Rankness," *arXiv preprint arXiv:1404.4646*, 2014.

[4] J. Cheng, Q. Ye, H. Jiang, D. Wang, and C. Wang, "STCDG: An efficient data gathering algorithm based on matrix completion for wireless sensor networks," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 2, pp. 850–861, 2013.

[5] A. Gogna and A. Majumdar, "Matrix completion incorporating auxiliary information for recommender system design," *Expert Systems with Applications*, vol. 42, no. 14, pp. 5789–5799, 2015.

[6] 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," *Signal Processing Magazine, IEEE*, vol. 30, no. 3, pp. 83–98, 2013.

[7] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE transactions on signal processing*, vol. 61, pp. 1644–1656, 2013.

[8] T. Jebara, J. Wang, and S.-F. Chang, "Graph construction and b-matching for semi-supervised learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 441–448.

[9] V. Kalofolias, X. Bresson, M. Bronstein, and P. Vandergheynst, "Matrix Completion on Graphs," *arXiv preprint arXiv:1408.1717*, 2014.

[10] S. Chen, A. Sandryhaila, J. M. Moura, and J. Kovačević, "Signal Recovery on Graphs," *arXiv preprint arXiv:1411.7414*, 2014.

[11] N. Rao, H.-F. Yu, P. K. Ravikumar, and I. S. Dhillon, "Collaborative filtering with graph information: Consistency and scalable methods," in *Advances in Neural Information Processing Systems*, 2015, pp. 2107–2115.

[12] K. Hou, Z. Zhou, A. M.-C. So, and Z.-Q. Luo, "On the linear convergence of the proximal gradient method for trace norm regularization," in *Advances in Neural Information Processing Systems*, 2013, pp. 710–718.

[13] K. Yi, J. Wan, T. Bao, and L. Yao, "A DCT regularized matrix completion algorithm for energy efficient data gathering in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2015, p. 96, 2015.

[14] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Mathematical Programming*, vol. 128, no. 1-2, pp. 321–353, 2011.

[15] R. Sun, "Matrix completion via nonconvex factorization: Algorithms and theory," Ph.D. dissertation, UNIVERSITY OF MINNESOTA, 2015.

[16] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.