Unrolled Multimodal Signal Restoration with Signed Twofold Graph Learning

Haruki Yokota, Hiroshi Higashi, and Yuichi Tanaka
Graduate School of Engineering, The University of Osaka, Japan
Email: h.yokota@sip.comm.eng.osaka-u.ac.jp, {higashi, ytanaka}@comm.eng.osaka-u.ac.jp

***Abstract*—In this paper, we propose an unrolling-based algorithm that jointly learns a twofold graph, which encodes spatial and modality characteristics simultaneously, from noisy signals while restoring multimodal signals at the same time. To learn the twofold graph, we formulate a convex optimization problem that considers smoothness on the spatial graph and similarity/dissimilarity on the modality graph. For multimodal signal restoration, we formulate an optimization problem based on the maximum a posteriori estimator of a matrix normal distribution. By unrolling the iterative optimization steps for both graph learning and signal restoration, an end-to-end interpretable neural network is obtained where hyperparameters in the iterative algorithm can be learned. Experimental results on synthetic data demonstrate that the proposed algorithm outperforms conventional multimodal graph signal restoration algorithms.**

I. Introduction

In sensor networks, sensors often capture a rich variety of data simultaneously, like video/audio streams and temperature/pressure readings. Such data are called multimodal signals. In practice, real-world multimodal signals are imperfect, suffering from noise or missing values. Signal restoration for those multimodal signals, including denoising and interpolation, is therefore crucial for a wide range of applications, such as autonomous driving, human-computer interaction, and medical diagnostics [1], [2], [3], [4].

A key principle in signal restoration on sensor networks is utilizing underlying structure(s) of data, whose structures are mathematically represented as graphs. Therefore, *graph signal processing* has been extensively studied over the past few decades [5], [6], [7].

Multimodal signals on a network can be assumed to have an underlying spatial structure as a spatial graph, and inter-modality relations are given by a modality graph. This assumption is known as *twofold graph assumption* (TGA) [8].

Previous works for signal restoration with TGA face two key challenges. First, many approaches either assume graphs are known a priori or require heavily parameterized deep neural network architectures to learn them [9], [10], [11], which makes it difficult in data-scarce scenarios. Second, all methods are designed for *unsigned* graphs, i.e., graphs only have non-negative edge weights. In other words, existing methods only model the pairwise *similarities* [12]. This clearly neglects potential *dissimilarities* or *negative correlations* that are especially important to analyze signals across modalities.

A representative example of inter-modality negative correlation can be found in environmental monitoring systems, where positive/negative correlations between meteorological measurements are encoded as positive/negative graph edges, respectively. In such cases, signed graphs—a graph with both positive and negative edge weights—are expected to improve signal restoration performance [13], [14].

In this paper, we address these limitations on multimodal graph signal restoration by proposing a method that: 1) jointly learns a spatial unsigned graph and a modality signed graph from data and 2) restores multimodal signals on the learned twofold graph via algorithm unrolling.

In the proposed method, we first formulate an optimization problem for simultaneous signal restoration and twofold graph learning. To solve this, we propose an alternating minimization framework that decouples the problem into two convex subproblems: signal restoration and graph learning. Each subproblem is then solved using iterative algorithms based on a maximum a posteriori (MAP) estimator of matrix normal distributions [15] and primal-dual splitting [16]. Finally, we unroll these iterative solvers to construct a lightweight deep learning architecture, allowing us to train internal hyperparameters end-to-end [17]. This unrolled network architecture estimates a solution in a fixed number of layers which is significantly smaller than that of the original iterative approach.

We evaluated our method for restoration of synthetic multimodal graph signals. Experimental results demonstrate that the proposed method significantly outperforms competing baselines in both signal denoising and interpolation in multiple scenarios.

Notation: We use bold uppercase letters for matrices ($\mathbf{A}$), and bold lowercase for vectors ($\mathbf{a}$). The corresponding non-bold letters with subscripts denote their elements (A_{ij} and a_i) The jth column of matrix $\mathbf{A}$ is denoted as the vector $\mathbf{a}_j$. The identity matrix is $\mathbf{I}$, and vectors of all ones and zeros are $\mathbf{1} = [1, \dots, 1]^\top$ and $\mathbf{0} = [0, \dots, 0]^\top$, respectively. The ℓ-p norm is $\|\cdot\|_p$ and the Frobenius norm is denoted as $\|\cdot\|_F$.

An undirected graph is denoted as $\mathcal{G}$, and we denote the set of N nodes in $\mathcal{G}$ as $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$. The topology and edge weights of $\mathcal{G}$ are defined by a symmetric weighted adjacency matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$, where W_{ij} represents the weight of the edge between node v_i and v_j. Throughout the paper, we assume $W_{ii} = 0$, i.e., the graphs have no self-loops.

The degree matrix is the diagonal matrix denoted as $\mathbf{D} = \operatorname{diag}(d_1, \dots, d_N)$, where $d_i = \sum_j W_{ij}$. The combinatorial graph Laplacian $\mathbf{L}$ for unsigned graphs is given by $\mathbf{L} = \mathbf{D} -$

979-8-3315-7206-8/25/$31.00

$\mathbf{W}$, and those for signed graphs are defined as $\bar{\mathbf{L}} = \bar{\mathbf{D}} - \mathbf{W}$, where $\bar{\mathbf{D}}$ is an absolute degree matrix whose ith diagonal element is defined as $\bar{d}_{ii} = \sum_j |W_{ij}|$.

II. Related Works

Graph learning is often required when no graph is given a priori, but we can still assume that underlying networks exist [18], [19]. In this section, we review unsigned/signed graph learning methods.

A. Unsigned Graph Learning

Let us consider an observation model of a graph signal $\mathbf{y} \in \mathbb{R}^N$ on a single modal graph $\mathcal{G}$ with N nodes as $\mathbf{y} = \mathbf{x} + \boldsymbol{\epsilon}$ where $\mathbf{x}$ is the unknown true signal and $\boldsymbol{\epsilon}$ is additive white Gaussian noise (AWGN) with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_N)$.

Based on the graph factor analysis model [18], $\mathbf{x}$ can be seen as a sample drawn from a multivariate Gaussian distribution with the precision matrix $\mathbf{L}$: $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_x, \mathbf{L}^\dagger + \sigma_\epsilon^2 \mathbf{I}_N)$ where $\boldsymbol{\mu}_x$ is the mean of $\mathbf{x}$ and $(\cdot)^\dagger$ represents the Moore-Penrose pseudo inverse.

The graph learning and signal restoration under this model with $\boldsymbol{\mu}_x = \mathbf{0}$ is discussed in [18], where they formulate an optimization problem to find $\mathbf{x}$ and $\mathbf{L}$ from $\mathbf{y}$ as:

$$\min_{\mathbf{x},\mathbf{L}} \|\mathbf{x} - \mathbf{y}\|_2^2 + \alpha \mathbf{x}^\top \mathbf{L} \mathbf{x} + \beta \|\mathbf{L}\|_F^2, \tag{1}$$

where α and β are regularization parameters and $\|\cdot\|_F$ denotes the Frobenius norm. The first term measures the data fidelity, and the second term measures the signal variation on the given $\mathbf{L}$ where $\mathbf{x}^\top \mathbf{L} \mathbf{x} = \sum_{i<j}^N W_{ij}(x_i - x_j)^2$, and the third term is a regularization on edge weights. (1) is non-convex due to the coupling of variables $\mathbf{x}$ and $\mathbf{L}$ in the second term.

In [18], the following alternating optimization is applied.

1) $\mathbf{L}$ is optimized with the fixed $\mathbf{x}$ by solving the following constrained quadratic optimization problem:

$$\min_{\mathbf{L} \in \mathcal{L}} \alpha \mathbf{x}^\top \mathbf{L} \mathbf{x} + \beta \|\mathbf{L}\|_F^2 \tag{2}$$

where $\mathcal{L}$ is a set of graph Laplacians given by

$$\mathcal{L} = \{\mathbf{L} | \mathrm{tr}(\mathbf{L}) = N,\ L_{ij} = L_{ji} \le 0,\ i \ne j, \mathbf{L}\mathbf{1} = 0\}.$$

2) $\mathbf{x}$ is obtained with the fixed $\mathbf{L}$ by solving

$$\min_{\mathbf{x}} \|\mathbf{x} - \mathbf{y}\|_2^2 + \alpha \mathbf{x}^\top \mathbf{L} \mathbf{x}. \tag{3}$$

This has a closed-form solution $\mathbf{x} = (\mathbf{I}_N + \alpha \mathbf{L})^{-1} \mathbf{y}$ as long as $(\mathbf{I}_N + \alpha \mathbf{L})$ is invertible.

These two steps are iterated until the difference from the previous iteration reaches a pre-defined tolerance value.

B. Signed Graph Learning

Most graph learning methods, including the aforementioned one, are designed for unsigned graphs, i.e., $W_{ij} \ge 0$ for all i, j. Unsigned graphs encode similarities. However, it is natural that we often encounter dissimilarities as well: In this case, signed graph learning is required [20], [21], [14].

A signed graph uses both positive and negative edge weights, which are stored in a *signed adjacency matrix* $\bar{\mathbf{W}} \in \mathbb{R}^{N \times N}$. The goal is to find a $\bar{\mathbf{W}}$ that reflects the relationships in a set of observed signals $\mathbf{X} \in \mathbb{R}^{P \times N}$, where P is the number of observations or features and N is the number of nodes.

An edge weight of a signed graph can be represented as $\bar{W}_{ij} = S_{ij} |\bar{W}_{ij}|$ where $S_{ij} = \mathrm{sign}(\bar{W}_{ij})$ is the edge sign matrix. Therefore, signed graph learning is formulated as follows [22]:

$$\min_{\mathbf{S} \in \mathcal{S}, \bar{\mathbf{W}} \in \mathcal{W}} \alpha \sum_{i,j} |\bar{W}_{ij}| \|\mathbf{x}_i - S_{ij} \mathbf{x}_j\|_2^2 + \beta \|\bar{\mathbf{W}}\|_F^2 \tag{4}$$

where $\mathcal{W}$ is a set of symmetric matrices with zero diagonals and $\mathcal{S} = \{-1, 1\}^{N \times N}$. Optimizing $\mathbf{S}$ together with $\bar{\mathbf{W}}$ is non-convex and hence, a two-step approach is considered in [22].

First, compute the distance matrix $\bar{\mathbf{Z}}$ together with the sign matrix $\mathbf{S}$ as:

$$\bar{Z}_{ij} = \min\{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2, \|\mathbf{x}_i + \mathbf{x}_j\|_2^2\} \tag{5}$$

$$S_{ij} = \begin{cases} 1 & \text{if } \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \le \|\mathbf{x}_i + \mathbf{x}_j\|_2^2 \\ -1 & \text{otherwise.} \end{cases} \tag{6}$$

Second, given these precomputed signs $\mathbf{S}$ and distances $\bar{\mathbf{Z}}$, optimize the weight magnitude $\mathbf{W} \ge \mathbf{0}$ by solving a convex optimization problem;

$$\min_{\mathbf{W} \in \mathcal{W}} \alpha \sum_{i,j} W_{ij} \bar{Z}_{ij} + \frac{\beta}{2} \|\mathbf{W}\|_F^2. \tag{7}$$

The signed adjacency matrix is then given by $\bar{\mathbf{W}}^\star = \mathbf{S} \circ \mathbf{W}$, where $\circ$ is the element-wise product. The minimization in (7) can be efficiently solved using proximal gradient-based algorithms [16].

III. Twofold Signed Graph Learning and Signal Restoration for Multimodal Signals

In this section, we present our proposed method for joint signal restoration and twofold signed graph learning for multimodal graph signals.

A. Formulation

Analogous to the multimodal graph signal model used in the other works [8], [12], we consider the following observation model of a multimodal signal $\mathbf{Y} \in \mathbb{R}^{N \times M}$ on a twofold graph:

$$\mathbf{Y} = \mathbf{X} + \mathbf{N} \tag{8}$$

where $\mathbf{X}$ is the unknown true multimodal graph signal whose column vectors are signals on a spatial graph $\mathcal{G}_s$, and row vectors are signals on a modality graph $\mathcal{G}_m$. $\mathbf{N}$ is a matrix whose entry is AWGN. We aim to recover $\mathbf{X}$ and its corresponding twofold graph Laplacians $\mathbf{L}_s$ and $\mathbf{L}_m$ from $\mathbf{Y}$.

First of all, in contrast to the other works on TGA [8], [12] without explicit signal models, we assume that $\mathbf{X}$ follows a matrix normal distribution:

$$\mathbf{X} \sim \mathcal{N}_{N \times M}(\mathbf{M}_\mathbf{X}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c) \tag{9}$$

where $\mathbf{M}_\mathbf{X}$ is the mean matrix in which $[\mathbf{M}_\mathbf{X}]_{ij} = \mathbb{E}[X_{ij}]$, $\boldsymbol{\Sigma}_r \in \mathbb{R}^{N \times N} = \mathbf{L}_s^\dagger$, and $\boldsymbol{\Sigma}_c \in \mathbb{R}^{M \times M} = \mathbf{L}_m^\dagger$ are a row-wise covariance matrix and a column-wise covariance matrix, respectively. Similar to the normal distribution (of vectors), the probability density function of $\mathbf{X}$ is given by

$$p(\mathbf{X}) = \frac{\exp\left(-\frac{1}{2}\mathrm{tr}\left[\mathbf{\Sigma}_c^{-1}\left(\mathbf{X}-\mathbf{M}_{\mathbf{X}}\right)^\top\mathbf{\Sigma}_r^{-1}\left(\mathbf{X}-\mathbf{M}_{\mathbf{X}}\right)\right]\right)}{2\pi^{\frac{NM}{2}}\det(\mathbf{\Sigma}_r)^{\frac{M}{2}}\det(\mathbf{\Sigma}_c)^{\frac{N}{2}}}. \tag{10}$$

Here, we consider the maximum a posteriori (MAP) estimator $\mathbf{X}^\star$ of $\mathbf{X}$ given $\mathbf{Y}$. Without loss of generality, we assume $\mathbf{X}$ is centered such that $\mathbf{M}_{\mathbf{X}} = \mathbf{0}_{N\times M}$. Then, the MAP estimate of $\mathbf{X}$ can be written as follows:

$$\begin{aligned}\mathbf{X}^\star &:= \arg\max_{\mathbf{X}} p(\mathbf{X}|\mathbf{Y})\\ &= \arg\max_{\mathbf{X}} p(\mathbf{Y}|\mathbf{X})p(\mathbf{X})\\ &= \arg\min_{\mathbf{X}} \left(-\log p(\mathbf{Y}|\mathbf{X}) - \log p(\mathbf{X})\right).\end{aligned} \tag{11}$$

Based on the observation model and the matrix normal distribution, we can derive

$$\begin{aligned}&p(\mathbf{N}) = p(\mathbf{Y}|\mathbf{X}) = \exp(-(\mathbf{Y}-\mathbf{X})^\top(\mathbf{Y}-\mathbf{X}))\\ &p(\mathbf{X}) \propto \exp(-\mathrm{tr}(\mathbf{\Sigma}_c^{-1}\mathbf{X}^\top\mathbf{\Sigma}_r^{-1}\mathbf{X})).\end{aligned} \tag{12}$$

By substituting (12) into (11), $\mathbf{X}^\star$ is given by

$$\mathbf{X}^\star := \arg\min_{\mathbf{X}} \|\mathbf{Y}-\mathbf{X}\|_F^2 + \alpha\mathrm{tr}(\mathbf{L}_m\mathbf{X}^\top\mathbf{L}_s\mathbf{X}). \tag{13}$$

Note that, solving this problem requires both $\mathbf{L}_s$ and $\mathbf{L}_m$ to be known while they are unknown in our setting. Thus, we formulate the following objective function to jointly estimate $\mathbf{X}^\star$ and the sparse graph Laplacians:

$$\begin{aligned}\min_{\mathbf{X},\mathbf{L}_s,\mathbf{L}_m} &\frac{1}{2}\|\mathbf{Y}-\mathbf{X}\|_F^2 + \alpha\mathrm{tr}(\mathbf{L}_m\mathbf{X}^\top\mathbf{L}_s\mathbf{X})\\ &+ \frac{\beta_s}{2}\|\mathbf{L}_s\|_F^2 + \frac{\beta_m}{2}\|\mathbf{L}_m\|_F^2.\end{aligned} \tag{14}$$

B. Optimization

The objective function (14) is non-convex in general. Therefore, we solve (14) with alternating optimization, where we fix either $\mathbf{X}$ or $\mathbf{L}_e$ ($e \in \{s, m\}$) to optimize the other one.

1) Signal Restoration: We first consider solving for $\mathbf{X}$ with given $\mathbf{L}_s$ and $\mathbf{L}_m$. In this case, the problem in (14) reduces to the following.

$$\min_{\mathbf{X}}\ J(\mathbf{X}) = \frac{1}{2}\|\mathbf{Y}-\mathbf{X}\|_F^2 + \frac{\mu}{2}\mathrm{tr}(\mathbf{L}_m\mathbf{X}^\top\mathbf{L}_s\mathbf{X}) \tag{15}$$

where μ is a regularization parameter.

Taking the derivative with respect to $\mathbf{X}$ on the objective function and setting it to zero yields the following first-order optimality condition:

$$\nabla_{\mathbf{X}}J(\mathbf{X}) = \mathbf{Y}-\mathbf{X} + \frac{\mu}{2}(\mathbf{L}_s\mathbf{X}\mathbf{L}_m + \mathbf{L}_m^\top\mathbf{X}^\top\mathbf{L}_s^\top) = \mathbf{0}. \tag{16}$$

(16) can be rewritten as

$$\mu\mathbf{L}_s\mathbf{X}\mathbf{L}_m + \mathbf{X} = \mathbf{Y}. \tag{17}$$

(17) is the standard form of *Sylvester equation*: $\mathbf{AXB}+\mathbf{X} = \mathbf{C}$. We solve (17) with a preconditioned iterative gradient algorithm proposed in [23]. We state the algorithm to solve (17) in Algorithm 1: κ is a step size, $\mathbf{P}_s \in \mathbb{R}^{N\times N}$ and $\mathbf{P}_m \in \mathbb{R}^{M\times M}$ are preconditioning matrices. The choice of preconditioning matrices is investigated in [23]. Here, we can simply set $\mathbf{P}_s$ and $\mathbf{P}_m$ as the diagonal entries of $\mathbf{L}_s$ and $\mathbf{L}_m$, respectively.

Algorithm 1 Iterative algorithm for solving (17)

Require: $\hat{\mathbf{X}}^{(0)}, \tilde{\mathbf{X}}^{(0)}$
Ensure: $\mathbf{X}^{(i)}$
  for $i = 0 \to K-1$ **do**
    $\mathbf{X}^{(i)} = (\hat{\mathbf{X}}^{(i-1)} + \tilde{\mathbf{X}}^{(i-1)})/2$
    $\hat{\mathbf{X}}^{(i)} = \mathbf{X}^{(i-1)} + \kappa\mathbf{P}_s^{-1}(\mathbf{Y} - \mu\mathbf{L}_s\mathbf{X}^{(i-1)}\mathbf{L}_m - \mathbf{X}^{(i-1)})$
    $\tilde{\mathbf{X}}^{(i)} = \mathbf{X}^{(i-1)} + \kappa(\mathbf{Y} - \mu\mathbf{L}_s\mathbf{X}^{(i-1)}\mathbf{L}_m - \mathbf{X}^{(i-1)})\mathbf{P}_m^{-1}$
  end for
  $\mathbf{X}^\star = (\hat{\mathbf{X}}^{(K-1)} + \tilde{\mathbf{X}}^{(K-1)})/2$

2) Twofold Signed Graph Learning: Next, we consider a twofold signed graph learning problem with the fixed $\mathbf{X}$:

$$\min_{\mathbf{L}_s,\ \mathbf{L}_m\in\mathcal{L}} \alpha\mathrm{tr}(\mathbf{L}_m\mathbf{X}^\top\mathbf{L}_s\mathbf{X}) + \frac{\beta_s}{2}\|\mathbf{L}_s\|_F^2 + \frac{\beta_m}{2}\|\mathbf{L}_m\|_F^2. \tag{18}$$

The first term in (18) promotes a graph signal smoothness on the twofold graph, which results in the row-wise smoothness with respect to $\mathbf{L}_s$ and column-wise smoothness with respect to $\mathbf{L}_m$. Note that this term is non-convex. Thus, we relax it by decoupling the smoothness term into the following: $\mathrm{tr}(\mathbf{X}^\top\mathbf{L}_s\mathbf{X}) + \mathrm{tr}(\mathbf{X}\mathbf{L}_m\mathbf{X}^\top)$. As a result, the following graph learning problem is derived:

$$\begin{aligned}\min_{\mathbf{L}_s,\ \mathbf{L}_m\in\mathcal{L}} &\alpha_s\mathrm{tr}(\mathbf{X}^\top\mathbf{L}_s\mathbf{X}) + \alpha_m\mathrm{tr}(\mathbf{X}\mathbf{L}_m\mathbf{X}^\top)\\ &+\frac{\beta_s}{2}\|\mathbf{L}_s\|_F^2 + \frac{\beta_m}{2}\|\mathbf{L}_m\|_F^2.\end{aligned} \tag{19}$$

To create a tractable problem, we further reformulate (18) into the following unconstrained optimization problem:

$$\begin{aligned}\min_{\mathbf{W}_s,\mathbf{W}_m} &\alpha_s\|\mathbf{W}_s\circ\mathbf{Z}_s\|_1 + \alpha_m\|\mathbf{W}_m\circ\mathbf{Z}_m\|_1\\ &+\beta_s\|\mathbf{W}_s\|_2^2 + \beta_m\|\mathbf{W}_m\|\\ &-\eta_s\mathbf{1}^\top\log(\mathbf{W}_s\mathbf{1}) - \eta_m\mathbf{1}^\top\log(\mathbf{W}_m\mathbf{1})\end{aligned} \tag{20}$$

where $\mathbf{Z}_s \in \mathbb{R}^{N\times N}$ and $\mathbf{Z}_m \in \mathbb{R}^{M\times M}$ are pairwise distance matrices. Recall that graph Laplacians can be derived from adjacency matrices as $\mathbf{L} = \mathrm{diag}(\mathbf{W1}) - \mathbf{W}$. Here, we assume that the spatial graph is unsigned, but the modality graph is signed. Thus, the distance matrices are computed as

$$[\mathbf{Z}_s]_{ij} = \sum_{m=1}^{M}|x_i - x_j|^2$$

$$[\mathbf{Z}_m]_{sr} = \min\{\sum_{n=1}^{N}|x_s - x_r|^2, \sum_{n=1}^{N}|x_s + x_r|^2\}$$

in which i, j denote the row indices of $\mathbf{X}$, and s, r denote the column indices of $\mathbf{X}$. We compute the edge signs $\mathbf{S}_m$ together with $\mathbf{Z}_m$ and use it to obtain signed adjacency for the modality graph as $\bar{\mathbf{W}}_m = \mathbf{S}_m \circ \mathbf{W}_m$ after solving (20).

Algorithm 2 Proposed PDS Algorithm for solving (21)

Require: $\mathbf{w}^{(0)}, \mathbf{v}_1^{(0)}$
Ensure: $\mathbf{w}^{(i)}$
while $\|\mathbf{w}^{(i+1)} - \mathbf{w}^{(i)}\| / \|\mathbf{w}^{(i)}\| > \epsilon$ **do**
$\mathbf{y}^{(i)} = \mathbf{w}^{(i)} - \tau(2\mathbf{Q}_2\mathbf{w}^{(i)} + \mathbf{S}^\top \mathbf{v}_1^{(i)})$
$\bar{\mathbf{y}}_1^{(i)} = \mathbf{v}_1^{(i)} + \gamma(\mathbf{S}\mathbf{w}^{(i)})$
$\mathbf{p}^{(i)} = \mathrm{prox}_{\gamma(\iota_{\mathcal{W}_v})}(\mathbf{y}^{(i)} - \gamma\mathbf{Q}_1\mathbf{z})$
$\bar{\mathbf{p}}_1^{(i)} = \bar{\mathbf{y}}_1^{(i)} - \gamma \mathrm{prox}_{\frac{1}{\gamma}(-\mathbf{1}^\top \log(\cdot))}\left(\frac{\bar{\mathbf{y}}_1^{(i)}}{\gamma}\right)$
$\mathbf{q}^{(i)} = \mathbf{p}^{(i)} - \gamma(2\mathbf{Q}_2\mathbf{p}^{(i)} + \mathbf{S}^\top \bar{\mathbf{p}}_1^{(i)})$
$\bar{\mathbf{q}}_1^{(i)} = \bar{\mathbf{p}}_1^{(i)} + \gamma(\mathbf{S}\mathbf{p}^{(i)})$
$\mathbf{w}^{(i+1)} = \mathbf{w}^{(i)} - \mathbf{y}^{(i)} + \mathbf{q}^{(i)}$
$\mathbf{v}_1^{(i+1)} = \mathbf{v}_1^{(i)} - \bar{\mathbf{y}}_1^{(i)} + \bar{\mathbf{q}}_1^{(i)}$
end while

We further reformulate (20) with a vectorized form. Let us define $\mathbf{w} = [\mathrm{vec}(\mathbf{W}_s)^\top, \mathrm{vec}(\mathbf{W}_m)^\top]^\top \in \mathbb{R}^{E_s+E_m}$ and $\mathbf{z} = [\mathrm{vec}(\mathbf{z}_s)^\top, \mathrm{vec}(\mathbf{z}_m)^\top]^\top \in \mathbb{R}^{E_s+E_m}$, where $E_s = N(N-1)/2$ and $E_m = M(M-1)/2$. The vectorized problem is then written as

$$\min_{\mathbf{w}} \mathbf{w}^\top \mathbf{Q}_1 \mathbf{z} + \mathbf{w}^\top \mathbf{Q}_2 \mathbf{w} - \mathbf{c}^\top \log(\mathbf{S}\mathbf{w}) + \iota_{>0}(\mathbf{w}) \tag{21}$$

where $\mathbf{S} : \mathbf{S}\mathbf{w} \to \mathbf{d} \in \mathbb{R}^{N+M}$ is a linear operator to transform a vector of edge weights into a vector of degrees $\mathbf{d} = [\mathbf{d}_s^\top, \mathbf{d}_m^\top]^\top$ where $\mathbf{d}_s = \mathbf{W}_s\mathbf{1}$ and $\mathbf{d}_m = \mathbf{W}_m\mathbf{1}$.

$$\mathbf{Q}_1 = \begin{pmatrix} \alpha_s \mathbf{I}_{E_s} & \mathbf{0}_{E_s,E_m} \\ \mathbf{0}_{E_m,E_s} & \alpha_m \mathbf{I}_{E_m} \end{pmatrix}$$
$$\mathbf{Q}_2 = \begin{pmatrix} \beta_s \mathbf{I}_{E_s} & \mathbf{0}_{E_s,E_m} \\ \mathbf{0}_{E_m,E_s} & \beta_m \mathbf{I}_{E_m} \end{pmatrix}$$
$$\mathbf{c} = [\eta_s \mathbf{1}^\top, \eta_m \mathbf{1}^\top]^\top.$$

Moreover, $\iota_{\geq 0}(\cdot)$ is an indicator function defined as:

$$(\iota_{\geq 0}(\mathbf{w}))_i = \begin{cases} 0 & w_i \geq 0 \\ \infty & \text{otherwise.} \end{cases} \tag{22}$$

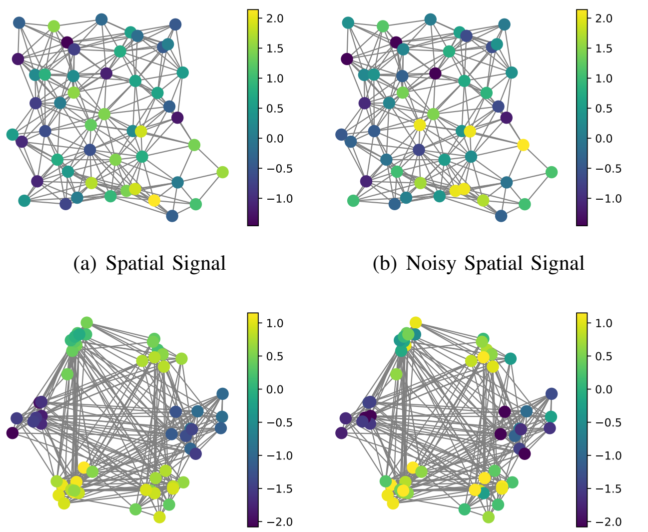
(21) can be solved with the primal-dual splitting (PDS) algorithm [16]. We give the iterative algorithm for solving (21) in Algorithm 2: τ and γ are step sizes for the primal and dual updates, respectively.

C. Algorithm Unrolling

The proposed algorithm consists of two parts: 1) solving (21) to learn twofold graphs from a set of noisy signals, and 2) solving (17) to get the clean signal. We *unfold* iterations in Algorithms 1 and 2 for learning hyperparameters from training data.

The graph learning module consists of eight parameters $\{\alpha_e, \beta_e, \eta_e, \tau, \gamma\}_{e\in s,m}$, and the signal restoration module consists of two parameters $\{\kappa, \mu\}$. By unfolding the iterations, they can be set trainable as

$$\begin{aligned} \mathbf{\Theta}_{PDS} &:= \{\alpha_s^{(t)}, \alpha_m^{(t)}, \beta_s^{(t)}, \beta_m^{(t)}, \eta_s^{(t)}, \eta_m^{(t)}, \tau^{(t)}, \gamma^{(t)}\}_{t=1}^T \\ \mathbf{\Theta}_{SL} &:= \{\kappa^{(k)}, \mu^{(k)}\}_{k=1}^K, \end{aligned} \tag{23}$$



(c) Modality Signal (d) Noisy Modality Signal

Fig. 1. Visualization of synthetic signal ($\sigma^2 = 0.5$). Top: Signal from the 10th modality on the spatial graph. Bottom: Signal from the 10th node on the modality graph.

where T and K are the numbers of inner layers of graph learning and signal restoration, respectively. As a result, sets of trainable parameters are represented as $\mathbf{\Theta} := \{\mathbf{\Theta}_{PDS}^{(l)}, \mathbf{\Theta}_{SL}^{(l)}\}_{l=1}^L$, where L is the number of outer layers. Each of the parameters is trained to minimize the mean squared error (MSE):

$$\mathbb{E}(\mathbf{\Theta}) = \frac{1}{NM}\|\mathbf{X}^\star - \mathbf{X}_{\text{ground truth}}\|_F^2. \tag{24}$$

Note that all components are (sub-)differentiable, and the parameters can be trained with backpropagation.

IV. Experiments

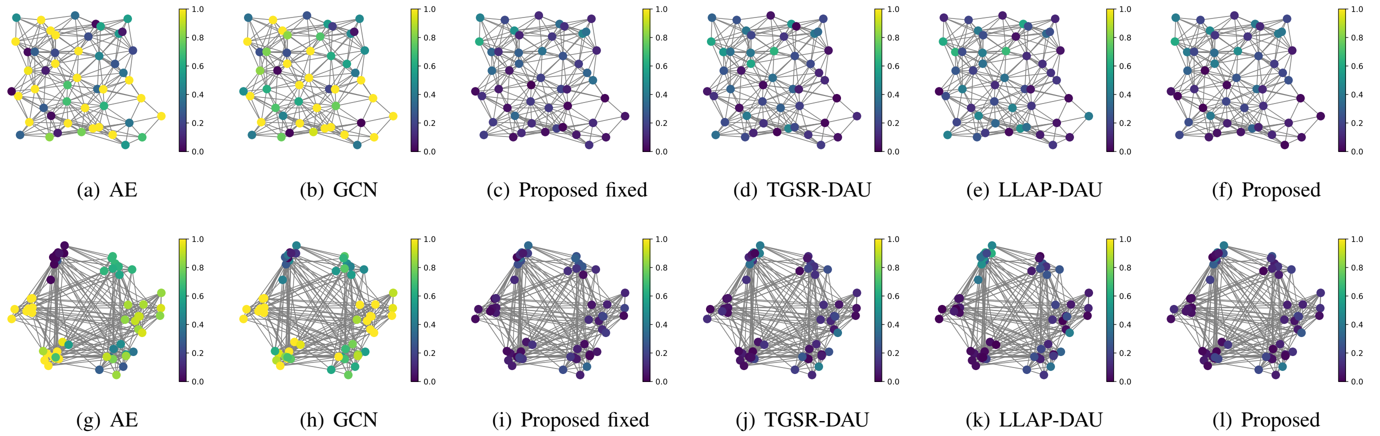
In this section, we evaluate our proposed method on multimodal graph signal denoising/interpolation and compare its performance against competing baselines.

Synthetic Data Generation:

We designed synthetic data to model an environmental monitoring sensor network, where signals exhibit spatial smoothness and have both positive and negative correlations across different modalities.

We generate the synthetic data in two stages: an underlying twofold graph is first created, and then signals consistent with the twofold graph are yielded.

First, we create a twofold graph. The spatial graph $\mathcal{G}_s$ is a six-nearest neighbor graph of $N = 50$ nodes randomly placed in a 2-D space $[0,1] \times [0,1]$. The modality graph $\mathcal{G}_m$ is a community graph with $M = 60$ nodes randomly assigned to six communities. Each community in the modality graph corresponds to a single modality. Here, we consider the stochastic block model [24] for the connectivity of the modality graph: Inner-community edges and inter-community



(a) AE (b) GCN (c) Proposed fixed (d) TGSR-DAU (e) LLAP-DAU (f) Proposed

(g) AE (h) GCN (i) Proposed fixed (j) TGSR-DAU (k) LLAP-DAU (l) Proposed

Fig. 2. Denoisng results ($L = 5$, $\sigma^2 = 0.5$). The absolute error of the 10th modality and the 10th node is plotted on the spatial graph (top) and the modality graph (bottom), respectively.

edges are generated with probability $p = 0.9$ and $q = 0.1$, respectively.

Second, we generate a synthetic multimodal graph signal $\mathbf{X} \in \mathbb{R}^{N \times M}$. The columns of $\mathbf{X}$ are partitioned according to the communities as $\mathbf{X} = [\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_5]$. For each community j containing c_j nodes, we generate a corresponding signal subblock $\mathbf{X}_j \in \mathbb{R}^{N \times c_j}$ as follows:

1) Generate a spatial signal $\mathbf{s}_j \in \mathbb{R}^N$ that is smooth on $\mathcal{G}_s$. This vector is drawn from an N-dimensional multivariate normal distribution: $\mathbf{s}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{L}_s^\dagger)$.
2) Generate a jth modality signal $\mathbf{p}_j \in \mathbb{R}^{c_j}$ that is smooth within the community. This vector is defined by:

$$\mathbf{p}_j = \frac{1}{4} f(\mathbf{v}_j), \ \mathbf{v}_j = \left(\frac{\pi}{2} \cdot \frac{i}{c_j - 1} \right)_{i=0}^{c_j - 1}, \tag{25}$$

where $f(\cdot)$ is a trigonometric function.
3) Combine two components to form $\mathbf{X}_j$ for each community j as:

$$\mathbf{X}_j = \mathbf{s}_j \mathbf{1}_{c_j}^\top + \mathbf{1}_N \mathbf{p}_j^\top. \tag{26}$$

We set $f(\cdot)$ to either a sine or cosine function to introduce positive/negative correlation across communities. The full signal $\mathbf{X}$ is constructed by stacking these subblocks. As a result, we yield a multimodal graph signal $\mathbf{X}$ where each column is smooth on the spatial graph $\mathcal{G}_s$ and each row is piecewise smooth [25] on the modality graph $\mathcal{G}_m$. A snapshot of the generated data is shown in Fig. 1.

For denoising, we generate 50 samples with five noise levels $\sigma = \{0.1, 0.2, 0.3, 0.4, 0.5\}$. For interpolation, we generate 50 samples with three missing rates $r = \{0.1, 0.2, 0.3\}$. The dataset was split into 80% for training and 20% for testing.

Experiment Setup:

We fix the number of graph learning layers to $T = 30$ and the number of signal restoration layers to $K = 30$ in all experiments. We compare the denoising performance against autoencoder (AE) [26], graph convolutional network (GCN) [27], graph signal denoising via twofold graph smoothness regularization with deep algorithm unrolling (TGSR-DAU) [8], and multimodal graph signal denoising with simultaneously learning Laplacian matrix using deep algorithm rolling (LLAP-DAU) [28]. TGSR-DAU and LLAP-DAU are algorithm-unrolling-based methods. For baselines requiring a known graph (GCN and TGSR-DAU), we constructed a six-nearest neighbor graph from the features of the noisy observation matrix $\mathbf{Y}$.

TABLE I
MEAN SQUARED ERRORS FOR DENOISING

Methods$\backslash\sigma^2$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
AE	0.959	0.959	0.959	0.959	0.959
GCN	0.973	0.973	0.972	0.971	0.971
Proposed (w/ fixed param.)	0.016	0.019	0.024	0.032	0.042
TGSR-DAU	0.010	0.016	0.026	0.040	0.060
LLAP-DAU	0.008	0.015	0.024	0.038	0.055
Proposed	**0.006**	**0.010**	**0.017**	**0.027**	**0.040**

TABLE II
MEAN SQUARED ERRORS FOR INTERPOLATION

Methods$\backslash r^2$	**0.1**	**0.2**	**0.3**
AE	0.117	0.117	0.117
GCN	0.969	0.968	0.968
Proposed (w/ fixed param.)	0.122	0.147	0.178
TGSR-DAU	0.060	0.068	0.079
LLAP-DAU	0.080	0.094	0.113
Proposed	**0.040**	**0.044**	**0.053**

All methods are trained for 30 epochs using Adam optimizer and a learning rate $= 0.01$. For AE, TGSR-DAU, LLAP-DAU, and the proposed method, we selected the best number of layers L from the following set ($L = \{1, 3, 5, 7, 9\}$) based on performance on a validation set. The number of layers for GCN is fixed to two as in [12]. We also compare results with the proposed solver with fixed parameters and 500 iterations for each subproblem. The parameters are hand-tuned to achieve the smallest MSE for denoising on data with $\sigma^2 = 0.5$.

Results:

We list the mean squared errors across varying levels of noise/missing rates in Tables I and II, respectively. We also visualize the absolute error between the ground truth and denoised signals in Fig. 2. The results in Tables I and II demonstrate that the proposed method consistently outperforms all competing baselines. The proposed solver with fixed parameters performs well on the specific noise level since it was tuned for $\sigma^2 = 0.5$, but it does not perform well in the other cases. In contrast, the unrolled version of the proposed method generalizes effectively and achieves the lowest error in all cases. Additionally, MSE differences between the proposed method and TGSR-DAU or LLAP-DAU become large as degradation increases, which may reflect the effectiveness of signed graphs.

V. Conclusions

In this paper, we propose an interpretable algorithm-unrolling-based method for jointly learning a twofold signed graph and restoring multimodal signals. Our approach is built on the assumption that signals are smooth on a spatial graph and exhibit positive or negative correlations on a modality graph. We formulated the joint problem using a Matrix Normal prior and a MAP estimator based on the MAP estimator of matrix normal distribution, leading to two alternating subproblems: 1) a signal restoration step, which is reduced to a Sylvester equation, and 2) a graph learning step, which is solved efficiently using a primal-dual splitting algorithm. We derive iterative solvers for each problem and present a framework based on algorithm unrolling to train the hyperparameters from data. Experiments on synthetic data demonstrate that our method outperforms conventional methods on signal denoising and interpolation tasks.

Acknowledgment

This work is supported in part by JSPS KAKENHI with grant number 25KJ1755 and JST AdCORP under Grant JP-MJKB2307.

References

[1] Y. Xiao, F. Codevilla, A. Gurram, O. Urfalioglu, and A. M. López, "Multimodal end-to-end autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 537–547, 2022.

[2] H. Liu, T. Lou, Y. Zhang, Y. Wu, Y. Xiao, C. S. Jensen, and D. Zhang, "EEG-Based Multimodal Emotion Recognition: A Machine Learning Perspective," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–29, 2024.

[3] W. Liu, J. Chen, H. Wang, Z. Fu, W. J. G. M. Peijnenburg, and H. Hong, "Perspectives on Advancing Multimodal Learning in Environmental Science and Engineering Studies," *Environmental Science & Technology*, vol. 58, no. 38, pp. 16 690–16 703, 2024.

[4] J. C. Arceo, L. Yu, and S. Bai, "Robust Sensor Fusion and Biomimetic Control of a Lower-Limb Exoskeleton With Multimodal Sensors," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 5425–5435, 2025.

[5] 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, 2013.

[6] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph Signal Processing: Overview, Challenges, and Applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[7] Y. Tanaka, Y. C. Eldar, A. Ortega, and G. Cheung, "Sampling Signals on Graphs: From Theory to Applications," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 14–30, 2020.

[8] M. Nagahama and Y. Tanaka, "Multimodal graph signal denoising via twofold graph smoothness regularization with deep algorithm unrolling," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 5862–5866.

[9] N. Shahid, N. Perraudin, V. Kalofolias, G. Puy, and P. Vandergheynst, "Fast robust pca on graphs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 740–756, 2016.

[10] F. Wang, Y. Wang, G. Cheung, and C. Yang, "Graph sampling for matrix completion using recurrent gershgorin disc shift," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2814–2829, 2020.

[11] S. Zheng, Z. Zhu, Z. Liu, Z. Guo, Y. Liu, Y. Yang, and Y. Zhao, "Multi-modal graph learning for disease prediction," *IEEE Transactions on Medical Imaging*, vol. 41, no. 9, pp. 2207–2216, 2022.

[12] K. Takanami, Y. Bandoh, S. Takamura, and Y. Tanaka, "Multimodadl graph signal denoising with simultaneous graph learning using deep algorithm unrolling," in *2023 IEEE International Conference on Image Processing (ICIP)*, 2023, pp. 1555–1559.

[13] C. Dinesh, G. Cheung, S. Bagheri, and I. V. Bajić, "Efficient signed graph sampling via balancing & gershgorin disc perfect alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 4, pp. 2330–2348, 2025.

[14] H. Yokota, H. Higashi, Y. Tanaka, and G. Cheung, "Efficient learning of balanced signed graphs via iterative linear programming," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.

[15] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. Chapman and Hall/CRC, 2000.

[16] L. Condat, "A Primal—Dual Splitting Method for Convex Optimization Involving Lipschitzian, Proximable and Linear Composite Terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.

[17] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021.

[18] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning Laplacian Matrix in Smooth Graph Signal Representations," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.

[19] G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the Dots: Identifying Network Structure via Graph Signal Processing," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 16–43, 2019.

[20] C. Yang, G. Cheung, and W. Hu, "Signed Graph Metric Learning via Gershgorin Disc Perfect Alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 7219–7234, 2022.

[21] C. Dinesh, S. Bagheri, G. Cheung, and I. V. Bajić, "Linear-Time Sampling on Signed Graphs Via Gershgorin Disc Perfect Alignment," in *ICASSP 2022 - 2022 IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP*, 2022, pp. 5942–5946.

[22] G. Matz and T. Dittrich, "Learning signed graphs from data," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5570–5574.

[23] J. Zhou, R. Wang, and Q. Niu, "A Preconditioned Iteration Method for Solving Sylvester Equations," *Journal of Applied Mathematics*, vol. 2012, no. 1, p. 401059, 2012.

[24] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Networks*, vol. 5, no. 2, pp. 109–137, 1983.

[25] S. Chen, R. Varma, A. Singh, and J. Kovačević, "Representations of piecewise smooth signals on graphs," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 6370–6374.

[26] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. 110, pp. 3371–3408, 2010.

[27] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations*, 2017.

[28] H. Kojima, K. Takanami, J. Hara, Y. Bandoh, S. Takamura, H. Higashi, and Y. Tanaka, "Algorithm unrolling-based denoising of multimodal graph signals," 2025.