FAST GRAPH-BASED BINARY CLASSIFICATION USING GERSHGORIN DISC ALIGNMENT OF ADJACENCY MATRIX

Gene Cheung[†] *Cheng Yang*[†]

[†] Dept. of EECS, York University, Toronto, Canada

ABSTRACT

***Index Terms*—** Graph-based classifier, semi-definite programming, Gershgorin circle theorem

1. INTRODUCTION

[1]

2. PRELIMINARIES

2.1. Graph Definitions

Suppose we are given a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = N$ nodes and edges $(i,j) \in \mathcal{E}$ connecting nodes i and j with weight $w_{ij} \in \mathbb{R}$. Denote by $\mathbf{W}$ the *adjacency matrix*, where $W_{ij} = w_{ij}$. Assuming that the edges are undirected, $\mathbf{W}$ is symmetric. Define next the diagonal *degree matrix* $\mathbf{D}$ where $D_{ii} = \sum_j W_{ij}$. The *combinatorial graph Laplacian matrix* is then defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$. To properly account for self-loops, the *generalized graph Laplacian matrix* is defined as $\mathcal{L} = \mathbf{D} - \mathbf{W} + \text{diag}(\mathbf{W})$.

2.2. GDA-based Linear Constraints

2.2.1. GCT-based Linear Constraints

Recall that the *Gershgorin Circle Theorem* (GCT) add ref states that any eigenvalue λ of a matrix $\mathbf{M}$ resides inside at least one *Gershgorin disc* i with center $c_i(\mathbf{M}) = M_{ii}$ and radius $r_i(\mathbf{M}) = \sum_{j \neq i} |M_{ij}|$, *i.e.*,

$$c_i(\mathbf{M}) - r_i(\mathbf{M}) \leq \lambda \leq c_i(\mathbf{M}) + r_i(\mathbf{M}) \tag{1}$$

This also means that smallest eigenvalue $\lambda_{\min}(\mathbf{M})$ of $\mathbf{M}$ is lower-bounded by the smallest disc left-end, *i.e.*,

$$\min_i c_i(\mathbf{M}) - r_i(\mathbf{M}) \leq \lambda_{\min}(\mathbf{M}) \tag{2}$$

Suppose now that an optimization requires that matrix variable $\mathbf{L}$ resides in the positive semi-definite (PSD) cone $\mathcal{S}$, *i.e.*, $\mathbf{L} \in \mathcal{S} = \{\mathbf{M} \,|\, \mathbf{x}^\top \mathbf{M} \mathbf{x} \geq 0, \forall \mathbf{x}\}$. Using GCT, we can write the following sufficient (but not necessary) set of N linear constraints to ensure variable $\mathbf{L}$ is PSD:

$$L_{ii} - \sum_{j \neq i} |L_{ij}| \geq 0, \quad \forall i \in \{1, \ldots, N\} \tag{3}$$

(4) states that the all disc left-ends of matrix $\mathbf{L}$ must be non-negative, which by (2) means that $0 \leq \lambda_{\min}(\mathbf{L})$, and hence $\mathbf{L}$ is PSD. Unfortunately, lower bound (2) is typically loose, meaning that $\min_i c_i - r_i \ll \lambda_{\min}$. If one imposes (4) instead of the original PSD cone constraint, one would result in a solution $\mathbf{L}^*$ that is PSD, but inferior.

Gene Cheung acknowledges the support of the NSERC grants [RGPIN-2019-06271], [RGPAS-2019-00110].

2.2.2. GDA Similarity Transform

A previous work on metric learning add ref has shown that given a generalized graph Laplacian matrix $\mathbf{L}$ corresponding to a balanced and connected signed graph $\mathcal{G}$ (with or without self-loops), a similarity transformation[1] $\mathbf{B} = \mathbf{S}\mathbf{L}\mathbf{S}^{-1}$ called GDA can be performed, where $\mathbf{S} = \text{diag}(1/v_1, \ldots, 1/v_N)$ and $\mathbf{v}$ is the first eigenvector of $\mathbf{L}$, so that the Gershgorin disc left-ends of $\mathbf{B}$ are all aligned at $\lambda_{\min}(\mathbf{B}) = \lambda_{\min}(\mathbf{L})$. In other words, there is no gap in the GCT lower bound, *i.e.*, $\min_i c_i(\mathbf{B}) - r_i(\mathbf{B}) = \lambda_{\min}(\mathbf{B})$. Instead of (4), one can now write *signal-adaptive* linear constraints to optimize variable $\mathbf{L}^t$ at iteration t to replace the PSD cone constraint:

$$L_{ii} - \sum_{j \neq i} |s_i L_{ij} / s_j| \geq 0, \quad \forall i \in \{1, \ldots, N\} \tag{4}$$

where scalars $s_i = 1/v_i^{t-1}$ and $\mathbf{v}^{t-1}$ is the first eigenvector of the previous solution $\mathbf{L}^{t-1}$ at iteration $t-1$.

3. PROBLEM FORMULATION

3.1. Graph-based Binary Classification

Given a *positive semi-definite* (PSD) graph Laplacian matrix $\mathcal{L}$ of a similarity graph $\mathcal{G}$, one can formulate a graph-based binary classification problem as follows:

$$\min_{\mathbf{x}} \mathbf{x}^\top \mathcal{L} \mathbf{x}, \quad \text{s.t.} \left\{ \begin{array}{l} x_i^2 = 1, \; \forall i \in \{1, \ldots, N\} \\ x_i = \hat{x}_i, \; i \in \mathcal{F} \end{array} \right. \tag{5}$$

The objective (5) states that the reconstructed signal $\mathbf{x}$ should be smooth w.r.t. $\mathcal{L}$. Because $\mathcal{L}$ is PSD, the objective is lower-bounded by $\mathbf{0}$. The first constraint is a binary constraint that ensures $x_i \in \{-1, 1\}$. The second constraint ensures that entries x_i in reconstructed signal $\mathbf{x}$ agrees with known labels $\hat{x}_i$ in set $\mathcal{F}$.

Optimization (5) is NP-hard because of the binary constraint on x_i's. One can define a conventional SDP relaxation add ref as follows. Define first $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$ and $\mathbf{M} = [\mathbf{X} \;\; \mathbf{x}; \; \mathbf{x}^\top \;\; 1]$. $\mathbf{M}$ is PSD because: i) 1 is PSD, and ii) Schur complement 1 of $\mathbf{M}$ is $\mathbf{X} - \mathbf{x}\mathbf{x}^\top = \mathbf{0}_{N \times N}$ is also PSD. Thus constraint $\mathbf{M} \succeq 0$ is a necessary but not sufficient condition for $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$, which together with $X_{ii} = 1, \forall i$ is equivalent to $x_i^2 = 1, \forall i$. We can now write the SDP relaxation as

$$\min_{\mathbf{x}, \mathbf{X}} \text{Tr}(\mathcal{L}\mathbf{X}) \;\; \text{s.t.} \left\{ \begin{array}{l} X_{ii} = 1, i \in \{1, \ldots, N\} \\ \mathbf{M} = \begin{bmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^\top & 1 \end{bmatrix} \succeq 0 \\ x_i = \hat{x}_i, \;\; i \in \mathcal{F} \end{array} \right. \tag{6}$$

[1]A similarity transform $\mathbf{B} = \mathbf{S}\mathbf{A}\mathbf{S}^{-1}$ of a square matrix $\mathbf{A}$, where $\mathbf{S}$ is an invertible matrix, means that $\mathbf{B}$ and $\mathbf{A}$ share the same eigenvalues.

Because (6) has linear objective and constraints with an additional PSD cone constraint, it is an SDP problem, solvable in polynomial time using a fast SDP solver such as interior point add ref.

3.2. SDP Dual

Instead of formulation (6), we derive the dual problem based on SDP duality theory add ref. To rewrite (6) in standard form, we first define

$$\mathbf{L} = \left[\begin{array}{cc} -\mathcal{L} & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{array} \right] \tag{7}$$

$$\mathbf{A}_i = \text{diag}(\mathbf{e}_{N+1}(i)), \quad \mathbf{B}_i = \left[\begin{array}{cc} \mathbf{0}_{N \times N} & \mathbf{e}_N(i) \\ \mathbf{e}_N(i)^\top & 0 \end{array} \right] \tag{8}$$

where $\mathbf{e}_K(i) \in \mathbb{R}^K$ is a canonical vector of length K with a single non-zero entry (equals to 1) at the i-th position, and $\text{diag}(\mathbf{v})$ is a diagonal matrix with diagonal entries equal to $\mathbf{v}$. Note that $\mathbf{A}_i$ and $\mathbf{B}_i$ are symmetric.

We can now rewrite optimization (6) as

$$\max_{\mathbf{M}} \; \mathbf{L} \cdot \mathbf{M}, \;\; \text{s.t.} \left\{ \begin{array}{l} \mathbf{A}_i \cdot \mathbf{M} = 1, \quad i \in \{1, \ldots, N\} \\ \mathbf{B}_i \cdot \mathbf{M} = 2\hat{x}_i, \;\; i \in \mathcal{F} \\ \mathbf{M} \succeq 0 \end{array} \right. \tag{9}$$

where the first and second constraints in (9) corresponds to the first and last constraints in (6), respectively.

Given SDP in standard form (9), we now write the corresponding dual SDP formulation as follows. First, we order known entries $\hat{x}_i$, $i \in \mathcal{F}$ into an order $o(i) \in \{1, \ldots, M\}$, $i \in \mathcal{F}$, such that a vector $\mathbf{b}$ of length $M = |\mathcal{F}|$ can be defined as

$$b_{o(i)} = 2\hat{x}_i, \quad \forall i \in \mathcal{F} \tag{10}$$

The SDP dual of (9) is

$$\min_{\mathbf{y},\mathbf{z}} \; \mathbf{1}_{N\times 1}^\top \mathbf{y} + \mathbf{b}^\top \mathbf{z} \tag{11}$$

$$\text{s.t.} \;\; \sum_{i=1}^{N+1} y_i \mathbf{A}_i + \sum_{i \in \mathcal{F}} z_{o(i)} \mathbf{B}_i - \mathbf{L} \succeq 0$$

where dual optimization variables are $\mathbf{y} \in \mathbb{R}^{N+1}$ and $\mathbf{z} \in \mathbb{R}^M$.

3.3. On Restricting $\mathbf{A}$ to Balanced Graph Laplacian Matrix

GDA-based linear constraints is that, we restrict our search space to only generalized graph Laplacian matrices of *balanced* signed graphs. We show that for the target relaxed binary classification problem (6), this is not a serious problem.

Consider the ideal solution, where $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$ and $x_i \in \{-1, 1\}$. We interpret matrix $\mathbf{A}$ in (**??**) as an generalized graph Laplacian matrix for a balanced signed graph. We observe the following:

1. For $1 \leq i, j \leq N$ and $i \neq j$: off-diagonal terms $A_{ij} = A_{ji} = -1$ if $x_i = x_j$, and $A_{ij} = A_{ji} = 1$ if $x_i = -x_j$.
2. For $1 \leq i \leq N$, $j = N+1$: last row/column off-diagonal terms $A_{ij} = A_{ji} = x_i$.

As an example, consider the label assignment $\mathbf{x} = [1 \; 1 \; -1 \; -1]^\top$. The corresponding matrix $\mathbf{A}$ is

$$\mathbf{A} = \left[\begin{array}{ccccc} 7 & -1 & 1 & 1 & 1 \\ -1 & 7 & 1 & 1 & 1 \\ 1 & 1 & 7 & -1 & -1 \\ 1 & 1 & -1 & 7 & -1 \\ 1 & 1 & -1 & -1 & 1 \end{array} \right] \tag{12}$$

We now assign color blue/red to nodes $\{1, \ldots, N+1\}$ as follows:

1. Assign color blue (red) to each node i, $1 \leq i \leq N$, with positive (negative) label x_i.
2. Assign color red to the last node $N+1$.

From Fig. draw a figure for this example, we can see that every same (different) color node pair is connected by a positive (negative) edge. We can thus conclude that $\mathbf{A}$ is a generalized graph Laplacian matrix for a balanced signed graph by the Cartwright-Harary Theorem.

4. ALGORITHM DEVELOPMENT

4.1. GDA-based Optimization

Instead of the SDP problem in (**??**), we replace the PSD cone constraints with a set of linear constraints based on GDA:

$$\min_{\mathbf{X},\mathbf{x}} \; \text{Tr}(\mathcal{L}\mathbf{X}), \;\; \text{s.t.} \left\{ \begin{array}{l} X_{ii} = 1, \; i \in \{1, \ldots, N\} \\ A_{ii} - \sum_{j \neq i} |s_i A_{ij} / s_j| \geq 0, \forall i \in \{1, \ldots, N\} \\ x_i = \hat{x}_i, \;\; i \in \mathcal{F} \end{array} \right. \tag{13}$$

where $\mathbf{A}$ is defined in (**??**). The second constraint in (13) states that the left-end of Gershgorin disc i of similar transform $\mathbf{B} = \mathbf{S}\mathbf{A}\mathbf{S}^{-1}$ of $\mathbf{A}$, where $\mathbf{S} = \text{diag}(s_1, \ldots, s_N)$, is at least 0. Satisfying these linear constraints for all disc i ensures that the resulting matrix $\mathbf{A}$ is in the PSD cone cite Cheng's PAMI arXiv paper. (13) has linear objective and constraints, and thus is a linear program (LP), solvable using any state-of-the-art LP solvers, such as Simplex or interior point red add refs. As done in Cheng's paper, (13) is solved iteratively, each iteration with scalars s_i updated appropriate, until convergence.

Let $\mathbf{H} = \sum_{i=1}^{N+1} y_i \mathbf{A}_i + \sum_{i \in \mathcal{F}} z_{o(i)} \mathbf{B}_i - \mathbf{L}$, $\tilde{y}_i = |y_i|$, $\tilde{z}_{o(i)} = |z_{o(i)}|$, then the GDA relaxed SDP dual is given by:

$$\min_{\mathbf{y},\mathbf{z}} \; \mathbf{1}_{N\times 1}^\top \mathbf{y} + \mathbf{b}^\top \mathbf{z} \tag{14}$$

$$\text{s.t.} \left\{ \begin{array}{l} y_i + L_{ii} - \sum_{j \neq i} |\frac{s_i}{s_j} \mathcal{L}_{ij}| \geq \rho, \text{if } i \notin \mathcal{F} \\ y_i + L_{ii} - \sum_{j \neq i, N+1} |\frac{s_i}{s_j} \mathcal{L}_{ij}| - |\frac{s_{o(i)}}{s_j}| \tilde{z}_{o(i)} \geq \rho, \text{if } i \in \mathcal{F} \\ y_i + L_{ii} - \sum_{i=1}^{M} |\frac{s_{o(i)}}{s_j}| \tilde{z}_{o(i)} \geq \rho, \text{if } i = N+1 \\ y_i \leq \tilde{y}_i \\ -y_i \leq \tilde{y}_i \\ z_{o(i)} \leq \tilde{z}_{o(i)} \\ -z_{o(i)} \leq \tilde{z}_{o(i)} \end{array} \right.$$

Denote $\mathbf{c} = [\mathbf{1}_{N\times 1} \; \mathbf{b} \; \mathbf{0}_{N+1+M}]$, $\mathbf{w} = [\mathbf{y} \; \mathbf{z} \; \tilde{\mathbf{y}} \; \tilde{\mathbf{z}}]^\top$. We write Eq. (14) in an LP standard form as:

$$\min_{\mathbf{w}} \; \mathbf{c}^\top \mathbf{w} \tag{15}$$

$$\text{s.t.} \left\{ \begin{array}{l} \mathbf{P}\mathbf{w} \leq \mathbf{d} \\ \{w_i\}_{i=1}^{N+1+M} \in \mathbb{R} \\ \{w_i\}_{i=N+1+M+1}^{2(N+1+M)} \geq 0 \end{array} \right.$$

where

$$\mathbf{P} = \begin{bmatrix} -1 & 0 & \dots & \dots & \dots & \\ 0 & -1 & \dots & \dots & \dots & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & -1 & |\frac{s_{o(i)}}{s_j}| & \dots & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & -1 & -1 & |\frac{s_{o(i)}}{s_j}| & \dots \\ 1 & 0 & \dots & -1 & 0 & \dots \\ -1 & 0 & \dots & -1 & 0 & \dots \\ 0 & 1 & \dots & 0 & -1 & \dots \\ 0 & -1 & \dots & 0 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

$$\mathbf{d} = \begin{bmatrix} L_{ii} - \sum_{j \neq i} |\frac{s_i}{s_j} \mathcal{L}_{ij}| - \rho \\ \vdots \\ L_{ii} - \sum_{j \neq i, N+1} |\frac{s_i}{s_j} \mathcal{L}_{ij}| - \rho \\ \vdots \\ L_{ii} - \rho \\ \mathbf{0}_{2(N+1+M)} \end{bmatrix}$$

The dual of Eq. (15) is

$$\max_{\mathbf{v}} \ \mathbf{d}^\top \mathbf{v} \tag{16}$$
$$\text{s.t.} \begin{cases} \{\mathbf{P}^\top \mathbf{v}\}_{i=1}^{N+1+M} = \mathbf{c} \\ \{\mathbf{P}^\top \mathbf{v}\}_{i=N+1+M+1}^{2(N+1+M)} \geq \mathbf{c} \\ \{v_i\}_{i=1}^{3(N+1)+2M} \geq 0 \end{cases}$$

4.2. Scalars Computation

At each iteration t, we compute appropriate scalars s_i^t as follows. Using previous solution $(\mathbf{X}^t, \mathbf{x}^t)$, we first compute corresponding matrix $\mathbf{A}^t$ using (**??**). We then approximate $\mathbf{A}^t$ with matrix $\mathbf{M}^t$ (to be discussed next), where $\mathbf{M}^t$ is a generalized graph Laplacian matrix corresponding to an irreducible *balanced* signed graph $\mathcal{G}$. We compute $\mathbf{M}^t$'s first eigenvector $\mathbf{v}^t$ via LOBPCG, and then scalars $s_i^t = 1/v_i^t, \forall i$. These computed scalars s_i^t are then used in (13) to compute a new solution $(\mathbf{X}^{t+1}, \mathbf{x}^{t+1})$.

4.3. Balancing matrix $\mathbf{A}^t$

The last remaining task to to approximate $\mathbf{A}^t$ with graph Laplacian matrix $\mathbf{M}^t$ corresponding to a balanced signed graph. try some graph balancing procedure, for example, Dinesh's ICIP'20.

5. EXPERIMENTAL RESULTS

6. CONCLUSION

7. APPENDIX

7.1. $\lambda_{\max}$ of $\mathbf{x}\mathbf{x}^\top$

By definition $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$, where $x_i \in \{-1, 1\}$, and hence $\mathbf{X}$ has diagonal entries that are all 1's. This means that Tr$(\mathbf{X}) = N$, which is also the sum of $\mathbf{X}$'s eigenvalues. Since $\mathbf{X}$ is also rank-1, we know that $\mathbf{X}$ has eigenvalue 0 with multiplicity $N-1$. Thus, we can conclude that the remaining eigenvalue must be $\lambda_{\max} = N$.

8. REFERENCES

[1] A. Ortega, P. Frossard, J. Kovacevic, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," in *Proceedings of the IEEE*, May 2018, vol. 106, no.5, pp. 808–828.