Supplemental Material for "Locality Preserving Feature Learning"

Quanquan Gu **Marina Danilevsky** **Zhenhui Li** **Jiawei Han**
Department of Computer Science
University of Illinois at Urbana-Champaign, IL 61801, USA
{qgu3,danilev1,zli28,hanj}@illinois.edu

1 Proof of Theorem 1

Theorem 1. *Let $\mathbf{Y} \in \mathbb{R}^{n\times m}$ be a matrix where each column is an eigenvector of eigen-problem $\mathbf{Ly} = \lambda\mathbf{Dy}$. If there exists a matrix $\mathbf{A} \in \mathbb{R}^{d\times m}$ and $\mathbf{p}$ where $\mathbf{p} \in \{0,1\}^d, \mathbf{p}^T\mathbf{1} = m$ such that $\mathbf{X}^T diag(\mathbf{p})\mathbf{A} = \mathbf{Y}$, then each column of $\mathbf{A}$ is an eigenvector of eigen-problem $diag(\mathbf{p})\mathbf{XLX}^T diag(\mathbf{p})\mathbf{a} = \lambda diag(\mathbf{p})\mathbf{XDX}^T diag(\mathbf{p})\mathbf{a}$ with the same eigenvalue λ.*

Proof. Since $\mathbf{X}^T\text{diag}(\mathbf{p})\mathbf{a} = \mathbf{y}$ we have

$$\text{diag}(\mathbf{p})\mathbf{XLy} = \text{diag}(\mathbf{p})\mathbf{XLX}^T\text{diag}(\mathbf{p})\mathbf{a}$$

and

$$\lambda\text{diag}(\mathbf{p})\mathbf{XDy} = \lambda\text{diag}(\mathbf{p})\mathbf{XDX}^T\text{diag}(\mathbf{p})\mathbf{a}$$

Since $\mathbf{Ly} = \lambda\mathbf{Dy}$, by left multiplying $\text{diag}(\mathbf{p})\mathbf{X}$ on both sides, we get

$$\text{diag}(\mathbf{p})\mathbf{XLy} = \lambda\text{diag}(\mathbf{p})\mathbf{XDy}$$

That is

$$\text{diag}(\mathbf{p})\mathbf{XLX}^T\text{diag}(\mathbf{p})\mathbf{a} = \lambda\text{diag}(\mathbf{p})\mathbf{XDX}^T\text{diag}(\mathbf{p})\mathbf{a}$$

□

2 Proof of Theorem 2

Theorem 2. *The global optimal solution of Eq. (20) is*

$$\mathbf{a}^{\pi(i)*} = \begin{cases} \mathbf{c}_t^{\pi(i)}, & i \le m \\ \mathbf{0}, & otherwise. \end{cases} \quad (1)$$

where $\pi(i)$ is a sorting function such that $||\mathbf{c}_t^{\pi(1)}|| \ge ||\mathbf{c}_t^{\pi(2)}|| \ge \ldots, \ge ||\mathbf{c}_t^{\pi(d)}||$.

Proof. Eq. (20) can be rewritten as

$$\mathbf{A}_{t+1} = \arg\min_{\mathbf{A}} \frac{\mu}{\tau}\sum_{i=1}^{d} ||\mathbf{a}^i - \mathbf{c}_t^i||_2^2,$$
$$\text{s.t.}|| \begin{bmatrix} ||\mathbf{a}^1||_2 \\ ||\mathbf{a}^2||_2 \\ \vdots \\ ||\mathbf{a}^d||_2 \end{bmatrix} ||_0 \le m. \quad (2)$$

For any $\mathbf{a}^i \neq \mathbf{0}$, we have $||\mathbf{a}^i - \mathbf{c}_t^i||_0 = 1$. In such case, the optimal $\mathbf{a}^i$ which gives the smallest objective value is $\mathbf{a}^i = \mathbf{c}_t^i$. And for $\mathbf{a}^i = \mathbf{0}$, we have $||\mathbf{a}^i - \mathbf{c}_t^i||_2^2 = ||\mathbf{c}_t^i||_2^2$. Thus, in order to give the smallest objective value, we have to select the first m largest $||\mathbf{c}_t^i||_2$ and set the corresponding $\mathbf{a}^i = \mathbf{c}_t^i$, which gives the solution of Eq. (1). □

3 Proof of Theorem 3

Theorem 3. *Let $\mathbf{L}' = \mathbf{XLX}^T$, $\mathbf{D}' = \mathbf{XDX}^T$, and $\lambda_i(\mathbf{L}', \mathbf{D}'), i = 1, \ldots, d$ be the generalized value of $\mathbf{L}'$ and $\mathbf{D}'$ sorted in ascending order. The optimal objective function value J of LPFL in Eq. (8) is bounded by*

$$\sum_{i=1}^{l} \lambda_i(\mathbf{L}', \mathbf{D}') \le J \le \sum_{i=1}^{l} \lambda_{i+d-m}(\mathbf{L}', \mathbf{D}').$$

where l is the dimension of the subspace learned by $\mathbf{A}$, m is the number of selected features.

Proof. Let the pair $(\mathbf{P}, \mathbf{Q})$ be $d\times d$ symmetric matrices with generalized spectrum $\lambda_i(\mathbf{P}, \mathbf{Q})$, with $\mathbf{Q}$ a positive definite matrix. Let $(\mathbf{P}_m; \mathbf{Q}_m)$ be a corresponding pair of $m\times m$ principal submatrices where $1 \le m \le d$, with generalized eigenvalues $\lambda_i(\mathbf{P}_m; \mathbf{Q}_m)$. Then, according to generalized Courant-Fischer "Min-Max" theorem [1] in matrix computation, $\forall i, 1 \le i \le m$, we have

$$\lambda_i(\mathbf{P}, \mathbf{Q}) \le \lambda_i(\mathbf{P}_m, \mathbf{Q}_m) \le \lambda_{i+d-m}(\mathbf{P}, \mathbf{Q})$$

Applying the above result to $(\mathbf{L}', \mathbf{D}')$, we have

$$\lambda_i(\mathbf{L}', \mathbf{D}') \le \lambda_i(\mathbf{L}'_m, \mathbf{L}'_m) \le \lambda_{i+d-m}(\mathbf{L}', \mathbf{D}') \qquad (3)$$

where $\mathbf{L}'_m$ and $\mathbf{D}'_m$ are the $m \times m$ principle submatrices of $\mathbf{L}'$ and $\mathbf{D}'$, which are extracted from $\mathrm{diag}(\mathbf{p})\mathbf{L}'\mathrm{diag}(\mathbf{p})$ and $\mathrm{diag}(\mathbf{p})\mathbf{D}'\mathrm{diag}(\mathbf{p})$. Since we choose the l eigenvectors corresponding to the l smallest eigenvalues of $(\mathbf{L}'_m, \mathbf{D}'_m)$to form the linear transformation $\mathbf{A}$, we have

$$J = \sum_{i=1}^{l} \lambda_i(\mathbf{L}'_m, \mathbf{L}'_m) \qquad (4)$$

Combining Eq. (3) and Eq. (4), we obtain

$$\sum_{i=1}^{l} \lambda_i(\mathbf{L}', \mathbf{D}') \le J \le \sum_{i=1}^{l} \lambda_{i+d-m}(\mathbf{L}', \mathbf{D}').$$

This completes the proof. □

References

[1] B. Moghaddam, Y. Weiss, and S. Avidan. Generalized spectral bounds for sparse lda. In *ICML*, pages 641–648, 2006.