
KerGM: Kernelized Graph Matching: Supplementary Material

Zhen Zhang¹, Yijian Xiang¹, Lingfei Wu², Bing Xue¹, Arye Nehorai¹

¹Washington University in St. Louis

²IBM Research

¹{zhen.zhang, yijian.xiang, xuebing, nehorai}@wustl.edu

²lwu@email.wm.edu

Abstract

The supplementary material consists of three parts. In the first part, we prove all the mathematical results of the kernelized graph matching formulation. We also provide additional technical discussions. In the second part, we prove all the mathematical results of the entropy-regularized Frank-Wolfe algorithm. In the third part, we describe the implementation details and give more experimental results.

1 Proofs of the mathematical results of kernelized graph matching

1.1 Proving Corollary 1

Corollary 1. $\forall \mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n \times n}$, $\Psi \odot \mathbf{X} \odot \mathbf{Y} = \Psi \odot (\mathbf{XY})$, and $\mathbf{Y} \odot (\mathbf{X} \odot \Psi) = (\mathbf{YX}) \odot \Psi$.

Proof.

(1).

$\forall i, j = 1, 2, \dots, n$, the (i, j) element of $\Psi \odot \mathbf{X} \odot \mathbf{Y}$ is

$$\begin{aligned} [\Psi \odot \mathbf{X} \odot \mathbf{Y}]_{ij} &= \sum_{k=1}^n [\Psi \mathbf{X}]_{ik} \mathbf{Y}_{kj} = \sum_{k=1}^n \mathbf{Y}_{kj} \sum_{\alpha=1}^n \Psi_{i\alpha} \mathbf{X}_{\alpha k} \\ &= \sum_{\alpha=1}^n \Psi_{i\alpha} \left(\sum_{k=1}^n \mathbf{X}_{\alpha k} \mathbf{Y}_{kj} \right) = \sum_{\alpha=1}^n \Psi_{i\alpha} [\mathbf{XY}]_{\alpha j} = [\Psi \odot (\mathbf{XY})]_{ij}. \end{aligned} \quad (1)$$

Therefore, $\Psi \odot \mathbf{X} \odot \mathbf{Y} = \Psi \odot (\mathbf{XY})$.

(2).

$\forall i, j = 1, 2, \dots, n$, the (i, j) element of $\mathbf{Y} \odot (\mathbf{X} \odot \Psi)$ is

$$[\mathbf{Y} \odot (\mathbf{X} \odot \Psi)]_{ij} = \sum_{k=1}^n \mathbf{Y}_{ik} [\mathbf{X} \Psi]_{kj} = \sum_{k=1}^n \mathbf{Y}_{ik} \sum_{\alpha=1}^n \mathbf{X}_{k\alpha} \Psi_{\alpha j} \quad (2)$$

$$= \sum_{\alpha=1}^n \left(\sum_{k=1}^n \mathbf{Y}_{ik} \mathbf{X}_{k\alpha} \right) \Psi_{\alpha j} = \sum_{\alpha=1}^n [\mathbf{YX}]_{i\alpha} \Psi_{\alpha j} = [(\mathbf{YX}) \odot \Psi]_{ij}. \quad (3)$$

Therefore, $\mathbf{Y} \odot (\mathbf{X} \odot \Psi) = (\mathbf{YX}) \odot \Psi$. □

1.2 Proving Proposition 1

Proposition 1. Define the function $\langle \cdot, \cdot \rangle_{\mathcal{F}_{\mathcal{H}}} : \mathcal{H}^{n \times n} \times \mathcal{H}^{n \times n} \rightarrow \mathbb{R}$ such that $\langle \Psi, \Xi \rangle_{\mathcal{F}_{\mathcal{H}}} \triangleq \text{tr}(\Psi^T * \Xi) = \sum_{i,j=1}^n \langle \Psi_{ij}, \Xi_{ij} \rangle_{\mathcal{H}}, \forall \Psi, \Xi \in \mathcal{H}^{n \times n}$. Then $\langle \cdot, \cdot \rangle_{\mathcal{F}_{\mathcal{H}}}$ induces an inner product on $\mathcal{H}^{n \times n}$.

Proof. It is sufficient to show that the function $\langle \cdot, \cdot \rangle_{\mathcal{F}_{\mathcal{H}}}$ satisfies the following properties.

1. [Conjugate symmetry]:

$$\overline{\langle \Psi, \Xi \rangle_{\mathcal{F}_{\mathcal{H}}}} = \sum_{i,j=1}^n \overline{\langle \Psi_{ij}, \Xi_{ij} \rangle_{\mathcal{H}}} = \sum_{i,j=1}^n \langle \Xi_{ij}, \Psi_{ij} \rangle_{\mathcal{H}} = \langle \Xi, \Psi \rangle_{\mathcal{F}_{\mathcal{H}}}, \quad (4)$$

2. [Linearity in the first argument]:

$$\langle a\Psi, \Xi \rangle_{\mathcal{F}_{\mathcal{H}}} = \sum_{i,j=1}^n \langle a\Psi_{ij}, \Xi_{ij} \rangle_{\mathcal{H}} = \sum_{i,j=1}^n a \langle \Psi_{ij}, \Xi_{ij} \rangle_{\mathcal{H}} = a \langle \Psi, \Xi \rangle_{\mathcal{F}_{\mathcal{H}}} \quad (5)$$

$$\begin{aligned} \langle \Psi^{(1)} + \Psi^{(2)}, \Xi \rangle_{\mathcal{F}_{\mathcal{H}}} &= \sum_{i,j=1}^n \langle \Psi_{ij}^{(1)} + \Psi_{ij}^{(2)}, \Xi_{ij} \rangle_{\mathcal{H}} = \sum_{i,j=1}^n \langle \Psi_{ij}^{(1)}, \Xi_{ij} \rangle_{\mathcal{H}} + \sum_{i,j=1}^n \langle \Psi_{ij}^{(2)}, \Xi_{ij} \rangle_{\mathcal{H}}. \\ &= \langle \Psi^{(1)}, \Xi \rangle_{\mathcal{F}_{\mathcal{H}}} + \langle \Psi^{(2)}, \Xi \rangle_{\mathcal{F}_{\mathcal{H}}} \end{aligned} \quad (6)$$

$$= \langle \Psi^{(1)}, \Xi \rangle_{\mathcal{F}_{\mathcal{H}}} + \langle \Psi^{(2)}, \Xi \rangle_{\mathcal{F}_{\mathcal{H}}} \quad (7)$$

3. [Positive-definiteness]:

$$\langle \Psi, \Psi \rangle_{\mathcal{F}_{\mathcal{H}}} = \sum_{i,j=1}^n \langle \Psi_{ij}, \Psi_{ij} \rangle_{\mathcal{H}} \geq 0. \quad (8)$$

$$\langle \Psi, \Psi \rangle_{\mathcal{F}_{\mathcal{H}}} = 0 \iff \forall i, j = 1, 2, \dots, n, \Psi_{ij} = 0 \iff \Psi = O. \quad (9)$$

□

1.3 Proving Corollary 2

Corollary 2. $\langle \Psi \odot X, \Xi \rangle_{\mathcal{F}_{\mathcal{H}}} = \langle \Psi, \Xi \odot X^T \rangle_{\mathcal{F}_{\mathcal{H}}}$ and $\langle X \odot \Psi, \Xi \rangle_{\mathcal{F}_{\mathcal{H}}} = \langle \Psi, X^T \odot \Xi \rangle_{\mathcal{F}_{\mathcal{H}}}$.

Proof.

(1).

$$\begin{aligned} \langle \Psi \odot X, \Xi \rangle_{\mathcal{F}_{\mathcal{H}}} &= \sum_{i=1}^n \sum_{j=1}^n \langle [\Psi \odot X]_{ij}, \Xi_{ij} \rangle_{\mathcal{H}} = \sum_{i=1}^n \sum_{j=1}^n \langle \sum_{k=1}^n \Psi_{ik} X_{kj}, \Xi_{ij} \rangle_{\mathcal{H}} \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \langle \Psi_{ik}, X_{kj} \Xi_{ij} \rangle_{\mathcal{H}} = \sum_{i=1}^n \sum_{k=1}^n \langle \Psi_{ik}, \sum_{j=1}^n X_{kj} \Xi_{ij} \rangle_{\mathcal{H}} \\ &= \sum_{i=1}^n \sum_{k=1}^n \langle \Psi_{ik}, [\Xi \odot X^T]_{ik} \rangle_{\mathcal{H}} = \langle \Psi, \Xi \odot X^T \rangle_{\mathcal{F}_{\mathcal{H}}}. \end{aligned} \quad (10)$$

(2).

$$\langle X \odot \Psi, \Xi \rangle_{\mathcal{F}_{\mathcal{H}}} = \sum_{i=1}^n \sum_{j=1}^n \langle [X \odot \Psi]_{ij}, \Xi_{ij} \rangle_{\mathcal{H}} = \sum_{i=1}^n \sum_{j=1}^n \langle \sum_{k=1}^n X_{ik} \Psi_{kj}, \Xi_{ij} \rangle_{\mathcal{H}} \quad (11)$$

$$= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \langle \Psi_{kj}, X_{ik} \Xi_{ij} \rangle_{\mathcal{H}} = \sum_{j=1}^n \sum_{k=1}^n \langle \Psi_{kj}, \sum_{i=1}^n X_{ik} \Xi_{ij} \rangle_{\mathcal{H}} \quad (12)$$

$$= \sum_{j=1}^n \sum_{k=1}^n \langle \Psi_{kj}, [X^T \odot \Xi]_{kj} \rangle_{\mathcal{H}} = \langle \Psi, X^T \odot \Xi \rangle_{\mathcal{F}_{\mathcal{H}}}. \quad (13)$$

□

1.4 Computing the gradient

$$\nabla J_\alpha(\mathbf{X}) = (1 - 2\alpha)[(\Psi^{(1)} * \Psi^{(1)})\mathbf{X} + \mathbf{X}(\Psi^{(2)} * \Psi^{(2)})] - 2(\Psi^{(1)} \odot \mathbf{X}) * \Psi^{(2)} - \mathbf{K}^N, \quad (14)$$

where $\forall i, j = 1, 2, \dots, n$, $[\Psi^{(1)} * \Psi^{(1)}]_{ij} = \sum_{e_{ik}^1, e_{kj}^1 \in \mathcal{E}_1} k^E(\bar{q}_{ik}^1, \bar{q}_{kj}^1)$; $\forall a, b = 1, 2, \dots, n$, $[\Psi^{(2)} * \Psi^{(2)}]_{ab} = \sum_{e_{ac}^2, e_{cb}^2 \in \mathcal{E}_2} k^E(\bar{q}_{ac}^2, \bar{q}_{cb}^2)$; and $\forall i, a = 1, 2, \dots, n$, $[(\Psi^{(1)} \odot \mathbf{X}) * \Psi^{(2)}]_{ia} = \sum_{e_{ik}^1 \in \mathcal{E}_1, e_{ca}^2 \in \mathcal{E}_2} \mathbf{X}_{kc} k^E(\bar{q}_{ik}^1, \bar{q}_{ca}^2)$.

We first present two useful lemmas.

Lemma 1. $\langle \Psi, \Xi \odot \mathbf{X} \rangle_{\mathcal{F}_\mathcal{H}} = \langle \Xi^T * \Psi, \mathbf{X} \rangle_{\mathcal{F}}$, and $\langle \Psi, \mathbf{X} \odot \Xi \rangle_{\mathcal{F}_\mathcal{H}} = \langle \Psi * \Xi^T, \mathbf{X} \rangle_{\mathcal{F}}$.

Remark 1. In our paper, we only consider the kernel values. That is, the inner product values are real numbers. So $\langle \psi, \varphi \rangle_{\mathcal{H}} = \langle \varphi, \psi \rangle_{\mathcal{H}}$, $\forall \psi, \varphi \in \mathcal{H}$.

Proof.

(1).

$$\begin{aligned} \langle \Psi, \Xi \odot \mathbf{X} \rangle_{\mathcal{F}_\mathcal{H}} &= \sum_{i=1}^n \sum_{j=1}^n \langle \Psi_{ij}, [\Xi \odot \mathbf{X}]_{ij} \rangle_{\mathcal{H}} = \sum_{i=1}^n \sum_{j=1}^n \langle \Psi_{ij}, \sum_{k=1}^n \Xi_{ik} \mathbf{X}_{kj} \rangle_{\mathcal{H}} \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \langle \Psi_{ij}, \Xi_{ik} \mathbf{X}_{kj} \rangle_{\mathcal{H}} = \sum_{k=1}^n \sum_{j=1}^n \mathbf{X}_{kj} \sum_{i=1}^n \langle \Psi_{ij}, \Xi_{ik} \rangle_{\mathcal{H}} \\ &= \sum_{k=1}^n \sum_{j=1}^n [\Xi^T * \Psi]_{kj} \mathbf{X}_{kj} = \langle \Xi^T * \Psi, \mathbf{X} \rangle_{\mathcal{F}}. \end{aligned} \quad (15)$$

(2).

$$\begin{aligned} \langle \Psi, \mathbf{X} \odot \Xi \rangle_{\mathcal{F}_\mathcal{H}} &= \sum_{i=1}^n \sum_{j=1}^n \langle \Psi_{ij}, [\mathbf{X} \odot \Xi]_{ij} \rangle_{\mathcal{H}} = \sum_{i=1}^n \sum_{j=1}^n \langle \Psi_{ij}, \sum_{k=1}^n \mathbf{X}_{ik} \Xi_{kj} \rangle_{\mathcal{H}} \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \langle \Psi_{ij}, \mathbf{X}_{ik} \Xi_{kj} \rangle_{\mathcal{H}} = \sum_{i=1}^n \sum_{k=1}^n \mathbf{X}_{ik} \sum_{j=1}^n \langle \Psi_{ij}, \Xi_{kj} \rangle_{\mathcal{H}} \\ &= \sum_{i=1}^n \sum_{k=1}^n [\Psi * \Xi^T]_{ik} \mathbf{X}_{ik} = \langle \Psi * \Xi^T, \mathbf{X} \rangle_{\mathcal{F}}. \end{aligned} \quad (16)$$

□

Lemma 2. $\Psi * (\Xi \odot \mathbf{X}) = (\Psi * \Xi)\mathbf{X}$, $(\mathbf{X} \odot \Psi) * \Xi = \mathbf{X}(\Psi * \Xi)$, and $\Psi * (\mathbf{X} \odot \Xi) = (\Psi \odot \mathbf{X}) * \Xi$.

Proof. The proof procedure is very similar with Corollary 1. □

Now we prove the equality (14).

Proof.

(1).

We first rewrite the function $J_\alpha(\mathbf{X})$ as

$$J_\alpha(\mathbf{X}) = -\langle \mathbf{K}^N, \mathbf{X} \rangle_{\mathcal{F}} + (1 - \alpha)J_1(\mathbf{X}) - \alpha J_2(\mathbf{X}),$$

where $J_1(\mathbf{X}) = \frac{1}{2} \|\Psi^{(1)} \odot \mathbf{X} - \mathbf{X} \odot \Psi^{(2)}\|_{\mathcal{F}_\mathcal{H}}^2$ and $J_2(\mathbf{X}) = \frac{1}{2} \|\Psi^{(1)} \odot \mathbf{X} + \mathbf{X} \odot \Psi^{(2)}\|_{\mathcal{F}_\mathcal{H}}^2$

We first compute the gradient of $J_1(\mathbf{X})$. We employ the following fact:

$$\forall \mathbf{E} \in \mathbb{R}^{n \times n}, \langle \nabla J_1(\mathbf{X}), \mathbf{E} \rangle_{\mathcal{F}} = \lim_{t \rightarrow 0} \frac{J_1(\mathbf{X} + t\mathbf{E}) - J_1(\mathbf{X})}{t}. \quad (17)$$

$$\begin{aligned}
& J_1(\mathbf{X} + t\mathbf{E}) - J_1(\mathbf{X}) \\
&= \frac{1}{2} \|\Psi^{(1)} \odot (\mathbf{X} + t\mathbf{E}) - (\mathbf{X} + t\mathbf{E}) \odot \Psi^{(2)}\|_{\mathcal{F}_\mathcal{H}}^2 - \frac{1}{2} \|\Psi^{(1)} \odot \mathbf{X} - \mathbf{X} \odot \Psi^{(2)}\|_{\mathcal{F}_\mathcal{H}}^2 \\
&= \frac{1}{2} t^2 \|\Psi^{(1)} \odot \mathbf{E} - \mathbf{E} \odot \Psi^{(2)}\|_{\mathcal{F}_\mathcal{H}}^2 + t \langle \Psi^{(1)} \odot \mathbf{X} - \mathbf{X} \odot \Psi^{(2)}, \Psi^{(1)} \odot \mathbf{E} - \mathbf{E} \odot \Psi^{(2)} \rangle_{\mathcal{F}_\mathcal{H}}
\end{aligned} \tag{18}$$

Immediately, $\langle \nabla J_1(\mathbf{X}), \mathbf{E} \rangle_{\mathcal{F}} = \langle \Psi^{(1)} \odot \mathbf{X} - \mathbf{X} \odot \Psi^{(2)}, \Psi^{(1)} \odot \mathbf{E} - \mathbf{E} \odot \Psi^{(2)} \rangle_{\mathcal{F}_\mathcal{H}}$.

We can rewrite the above formula as

$$\begin{aligned}
& \langle \Psi^{(1)} \odot \mathbf{X} - \mathbf{X} \odot \Psi^{(2)}, \Psi^{(1)} \odot \mathbf{E} - \mathbf{E} \odot \Psi^{(2)} \rangle_{\mathcal{F}_\mathcal{H}} \\
&= \langle \Psi^{(1)} \odot \mathbf{X} - \mathbf{X} \odot \Psi^{(2)}, \Psi^{(1)} \odot \mathbf{E} \rangle_{\mathcal{F}_\mathcal{H}} - \langle \Psi^{(1)} \odot \mathbf{X} - \mathbf{X} \odot \Psi^{(2)}, \mathbf{E} \odot \Psi^{(2)} \rangle_{\mathcal{F}_\mathcal{H}} \\
&= \langle \Psi^{(1)} * (\Psi^{(1)} \odot \mathbf{X}) - \Psi^{(1)} * (\mathbf{X} \odot \Psi^{(2)}), \mathbf{E} \rangle_{\mathcal{F}} - \langle (\Psi^{(1)} \odot \mathbf{X}) * \Psi^{(2)} - (\mathbf{X} \odot \Psi^{(2)}) * \Psi^{(2)}, \mathbf{E} \rangle_{\mathcal{F}_\mathcal{H}} \\
&= \langle (\Psi^{(1)} * \Psi^{(1)})\mathbf{X} - 2(\Psi^{(1)} \odot \mathbf{X}) * \Psi^{(2)} + \mathbf{X}(\Psi^{(2)} * \Psi^{(2)}), \mathbf{E} \rangle_{\mathcal{F}},
\end{aligned} \tag{19}$$

where the 3rd equality holds because of Lemma 1 and both $\Psi^{(1)}$ and $\Psi^{(2)}$ are symmetric, and the last equality holds because of Lemma 2. Therefore,

$$\nabla J_1(\mathbf{X}) = (\Psi^{(1)} * \Psi^{(1)})\mathbf{X} - 2(\Psi^{(1)} \odot \mathbf{X}) * \Psi^{(2)} + \mathbf{X}(\Psi^{(2)} * \Psi^{(2)}).$$

Similarly, we have

$$\nabla J_2(\mathbf{X}) = (\Psi^{(1)} * \Psi^{(1)})\mathbf{X} + 2(\Psi^{(1)} \odot \mathbf{X}) * \Psi^{(2)} + \mathbf{X}(\Psi^{(2)} * \Psi^{(2)}).$$

Finally, substituting $\nabla J_1(\mathbf{X})$ and $\nabla J_2(\mathbf{X})$ into $\nabla J_\alpha(\mathbf{X}) = -\mathbf{K}^N + (1 - \alpha)\nabla J_1(\mathbf{X}) - \alpha\nabla J_2(\mathbf{X})$, we obtain the result (14).

(2).

For the first term $\Psi^{(1)} * \Psi^{(1)}$, we have

$$[\Psi^{(1)} * \Psi^{(1)}]_{ij} = \sum_{k=1}^n \langle \Psi_{ik}^{(1)}, \Psi_{kj}^{(1)} \rangle_{\mathcal{H}_\mathcal{K}} = \sum_{e_{ik}^1, e_{kj}^1 \in \mathcal{E}_1} \langle \psi(\bar{q}_{ik}^1), \psi(\bar{q}_{kj}^1) \rangle_{\mathcal{H}_\mathcal{K}} = \sum_{e_{ik}^1, e_{kj}^1 \in \mathcal{E}_1} k^E(\bar{q}_{ik}^1, \bar{q}_{kj}^1). \tag{20}$$

For the second term $\Psi^{(2)} * \Psi^{(2)}$, we have similar explanations.

For the third term $(\Psi^{(1)} \odot \mathbf{X}) * \Psi^{(2)}$, we have

$$\begin{aligned}
[(\Psi^{(1)} \odot \mathbf{X}) * \Psi^{(2)}]_{ia} &= \sum_{c=1}^n \langle [\Psi^{(1)} \odot \mathbf{X}]_{ic}, \Psi_{ca}^{(2)} \rangle_{\mathcal{H}_\mathcal{K}} = \sum_{c=1}^n \langle \sum_{k=1}^n \Psi_{ik}^{(1)} \mathbf{X}_{kc}, \Psi_{ca}^{(2)} \rangle_{\mathcal{H}_\mathcal{K}} \\
&= \sum_{c=1}^n \sum_{k=1}^n \langle \Psi_{ik}^{(1)} \mathbf{X}_{kc}, \Psi_{ca}^{(2)} \rangle_{\mathcal{H}_\mathcal{K}} = \sum_{e_{ik}^1 \in \mathcal{E}_1, e_{ca}^2 \in \mathcal{E}_2} \mathbf{X}_{kc} \langle \psi(\bar{q}_{ik}^1), \psi(\bar{q}_{ca}^2) \rangle_{\mathcal{H}_\mathcal{K}} \\
&= \sum_{e_{ik}^1 \in \mathcal{E}_1, e_{ca}^2 \in \mathcal{E}_2} \mathbf{X}_{kc} k^E(\bar{q}_{ik}^1, \bar{q}_{ca}^2).
\end{aligned} \tag{21}$$

□

1.4.1 Gradients in compact matrix multiplication forms

In this section, we rewrite the terms of (14) in compact matrix multiplication forms, providing a convenient way to compute gradients. We first give some necessary definitions.

1. Given a graph $\mathcal{G} = \{\mathbf{A}, \mathcal{V}, \mathbf{P}, \mathcal{E}, \mathbf{Q}\}$ of n nodes and m edges. We define the Head-incidence matrix $\mathbf{G} \in \{0, 1\}^{n \times m}$ and the Tail-incidence matrix $\mathbf{H} \in \{0, 1\}^{n \times m}$. For any edge $e_{ij} \in \mathcal{E}$, we arbitrarily assign a direction on e_{ij} , e.g., $v_i \rightarrow v_j$ or $v_j \rightarrow v_i$. Suppose that the artificially assigned direction of e_{ij} is $v_j \rightarrow v_i$, then the items $\mathbf{G}(v_j, e_{ij}) = 1$ and $\mathbf{H}(v_i, e_{ij}) = 1$. A toy example is shown in Fig. 1.

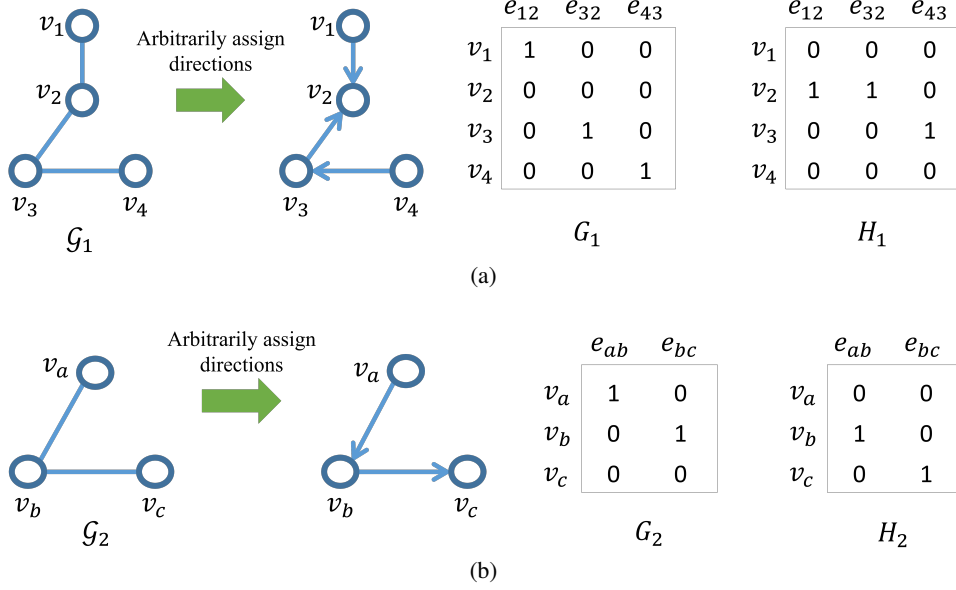


Figure 1: (a) A toy Graph \mathcal{G}_1 , and its Head-incidence matrix G_1 and Tail-incidence matrix H_1 ; (b) A toy Graph \mathcal{G}_2 , and its Head-incidence matrix G_2 and Tail-incidence matrix H_2 .

2. Given two graphs $\mathcal{G}_1 = \{\mathbf{A}_1, \mathcal{V}_1, \mathbf{P}_1, \mathcal{E}_1, \mathbf{Q}_1\}$ of n_1 nodes and m_1 edges, and $\mathcal{G}_2 = \{\mathbf{A}_2, \mathcal{V}_2, \mathbf{P}_2, \mathcal{E}_2, \mathbf{Q}_2\}$ of n_2 nodes and m_2 edges, let $\mathbf{K}_{11}^E \in \mathbb{R}^{m_1 \times m_1}$, $\mathbf{K}_{22}^E \in \mathbb{R}^{m_2 \times m_2}$, and $\mathbf{K}_{12}^E \in \mathbb{R}^{m_1 \times m_2}$ be three kernel matrices induced by the kernel k^E (the edge affinity function). They are defined such that

$$[\mathbf{K}_{11}^E](e_{i_1 j_1}^1, e_{i_2 j_2}^1) = k^E(\vec{q}_{i_1 j_1}^1, \vec{q}_{i_2 j_2}^1), \quad \text{if } e_{i_1 j_1}^1, e_{i_2 j_2}^1 \in \mathcal{E}_1, \quad (22)$$

$$[\mathbf{K}_{22}^E](e_{a_1 b_1}^2, e_{a_2 b_2}^2) = k^E(\vec{q}_{a_1 b_1}^2, \vec{q}_{a_2 b_2}^2), \quad \text{if } e_{a_1 b_1}^2, e_{a_2 b_2}^2 \in \mathcal{E}_2, \quad (23)$$

$$[\mathbf{K}_{12}^E](e_{ij}^1, e_{ab}^2) = k^E(\vec{q}_{ij}^1, \vec{q}_{ab}^2), \quad \text{if } e_{ij}^1 \in \mathcal{E}_1, e_{ab}^2 \in \mathcal{E}_2, \quad (24)$$

Let G_1 and H_1 , and G_2 and H_2 be the Head-incidence matrix and the Tail-incidence matrix of graph \mathcal{G}_1 and \mathcal{G}_2 (see Fig. 1), respectively. Then the terms in (14) can be written as

$$\begin{aligned} \Psi^{(1)} * \Psi^{(1)} &= H_1(G_1^T G_2 \circ K_{11}^E) H_2^T + H_1(G_1^T H_2 \circ K_{11}^E) G_2^T \\ &\quad + G_1(H_1^T G_2 \circ K_{11}^E) H_2^T + G_1(H_1^T H_2 \circ K_{11}^E) G_2^T, \end{aligned} \quad (25)$$

$$\begin{aligned} \Psi^{(2)} * \Psi^{(2)} &= H_1(G_1^T G_2 \circ K_{22}^E) H_2^T + H_1(G_1^T H_2 \circ K_{22}^E) G_2^T \\ &\quad + G_1(H_1^T G_2 \circ K_{22}^E) H_2^T + G_1(H_1^T H_2 \circ K_{22}^E) G_2^T, \end{aligned} \quad (26)$$

$$\begin{aligned} \text{and } (\Psi^{(1)} \odot \mathbf{X}) * \Psi^{(2)} &= H_1(G_1^T \mathbf{X} G_2 \circ K_{12}^E) H_2^T + H_1(G_1^T \mathbf{X} H_2 \circ K_{12}^E) G_2^T \\ &\quad + G_1(H_1^T \mathbf{X} G_2 \circ K_{12}^E) H_2^T + G_1(H_1^T \mathbf{X} H_2 \circ K_{12}^E) G_2^T, \end{aligned} \quad (27)$$

where \circ denotes the Hadamard product between matrices.

Proof.

(1).

We first prove the equality (27). It suffices to show that the (i, a) term in the left part equals to the (i, a) term in the right part.

Let $e_{ik_1}^1, e_{ik_2}^1, \dots, e_{ik_{D^i}}^1$ be the edges incident to the node v_i in graph \mathcal{G}_1 , where D^i is the degree of the node v_i . Without loss of generality, we assume the assigned directions of these edges are, for some $1 \leq s \leq D^i$,

$$v_i \rightarrow v_{k_1}, v_i \rightarrow v_{k_2}, \dots, v_i \rightarrow v_{k_s} \quad \text{and} \quad v_{k_{s+1}} \rightarrow v_i, v_{k_{s+2}} \rightarrow v_i, \dots, v_{k_{D^i}} \rightarrow v_i. \quad (28)$$

Considering the i th row, $\mathbf{G}_1(v_i, :)$ and $\mathbf{H}_1(v_i, :)$, of the matrix \mathbf{G}_1 and \mathbf{H}_1 , we have

$$\mathbf{G}_1(v_i, e_{ik_1}^1) = \mathbf{G}_1(v_i, e_{ik_2}^1) = \dots = \mathbf{G}_1(v_i, e_{ik_s}^1) = 1, \quad (29)$$

$$\text{and } \mathbf{H}_1(v_i, e_{ik_{s+1}}^1) = \mathbf{H}_1(v_i, e_{ik_{s+2}}^1) = \dots = \mathbf{H}_1(v_i, e_{ik_{D^i}}^1) = 1. \quad (30)$$

Any other item in the i th row of the matrix \mathbf{G}_1 and \mathbf{H}_1 is zero.

Let $e_{C_1a}^2, e_{C_2a}^2, \dots, e_{C_{D^a}a}^2$ be the edges incident to the node v_a in graph \mathcal{G}_2 , where D^a is the degree of the node v_a . Without loss of generality, we assume the assigned directions of these edges are, for some $1 \leq t \leq D^a$,

$$v_{C_1} \rightarrow v_a, v_{C_2} \rightarrow v_a, \dots, v_{C_t} \rightarrow v_a \quad \text{and} \quad v_a \rightarrow v_{C_{t+1}}, v_a \rightarrow v_{C_{t+2}}, \dots, v_a \rightarrow v_{C_{D^a}}. \quad (31)$$

Considering the a th row, $\mathbf{G}_2(v_a, :)$ and $\mathbf{H}_2(v_a, :)$, of the matrix \mathbf{G}_2 and \mathbf{H}_2 , we have

$$\mathbf{G}_2(v_a, e_{C_{t+1}a}^2) = \mathbf{G}_2(v_a, e_{C_{t+2}a}^2) = \dots = \mathbf{G}_2(v_a, e_{C_{D^a}a}^2) = 1, \quad (32)$$

$$\text{and } \mathbf{H}_2(v_a, e_{C_1a}^2) = \mathbf{H}_2(v_a, e_{C_2a}^2) = \dots = \mathbf{H}_2(v_a, e_{C_ta}^2) = 1. \quad (33)$$

Any other item in the a th row of the matrix \mathbf{G}_2 and \mathbf{H}_2 is zero.

We first consider the first term $\mathbf{H}_1(\mathbf{G}_1^T \mathbf{X} \mathbf{G}_2 \circ \mathbf{K}_{12}^E) \mathbf{H}_2^T$ in $(\Psi^{(1)} \odot \mathbf{X}) * \Psi^{(2)}$ (see (27)),

$$[\mathbf{H}_1(\mathbf{G}_1^T \mathbf{X} \mathbf{G}_2 \circ \mathbf{K}_{12}^E) \mathbf{H}_2^T]_{ia} \quad (34)$$

$$= [\mathbf{H}_1(v_i, :)] (\mathbf{G}_1^T \mathbf{X} \mathbf{G}_2 \circ \mathbf{K}_{12}^E) [\mathbf{H}_2(v_a, :)]^T \quad (35)$$

$$= \sum_{\alpha=1}^{D^i} \sum_{\beta=1}^{D^a} \mathbf{H}_1(v_i, e_{ik_\alpha}^1) \mathbf{H}_2(v_a, e_{C_\beta a}^2) [\mathbf{G}_1^T \mathbf{X} \mathbf{G}_2 \circ \mathbf{K}_{12}^E](e_{ik_\alpha}^1, e_{C_\beta a}^2) \quad (36)$$

$$= \sum_{\alpha=s+1}^{D^i} \sum_{\beta=1}^t [\mathbf{G}_1^T \mathbf{X} \mathbf{G}_2 \circ \mathbf{K}_{12}^E](e_{ik_\alpha}^1, e_{C_\beta a}^2) \quad \text{By the definition of } \mathbf{H}_1(30), \mathbf{H}_2(33) \quad (37)$$

$$= \sum_{\alpha=s+1}^{D^i} \sum_{\beta=1}^t [\mathbf{G}_1^T \mathbf{X} \mathbf{G}_2](e_{ik_\alpha}^1, e_{C_\beta a}^2) \times [\mathbf{K}_{12}^E](e_{ik_\alpha}^1, e_{C_\beta a}^2) \quad (38)$$

$$= \sum_{\alpha=s+1}^{D^i} \sum_{\beta=1}^t [\mathbf{K}_{12}^E](e_{ik_\alpha}^1, e_{C_\beta a}^2) [\mathbf{G}_1(:, e_{ik_\alpha}^1)^T \mathbf{X} \mathbf{G}_2(:, e_{C_\beta a}^2)] \quad (39)$$

$$= \sum_{\alpha=s+1}^{D^i} \sum_{\beta=1}^t \mathbf{X}_{k_\alpha C_\beta} [\mathbf{K}_{12}^E](e_{ik_\alpha}^1, e_{C_\beta a}^2), \quad (40)$$

$$= \sum_{\alpha=1}^s \sum_{\beta=t+1}^{D^a} \mathbf{X}_{k_\alpha C_\beta} k^E(\vec{q}_{ik_\alpha}^1, \vec{q}_{C_\beta a}^2). \quad (41)$$

where $\mathbf{G}_1(:, e_{ik_\alpha}^1)$ is a column of \mathbf{G}_1 , corresponding to the edge $e_{ik_\alpha}^1$, and $\mathbf{G}_2(:, e_{C_\beta a}^2)$ is a column of \mathbf{G}_2 , corresponding to the edge $e_{C_\beta a}^2$. By the definition of \mathbf{G}_1 (29) and \mathbf{G}_2 (32), $\mathbf{G}_1(v_{k_\alpha}, e_{ik_\alpha}^1) = 1$, since the direction of edge $e_{ik_\alpha}^1$ is assigned as $v_{k_\alpha} \rightarrow v_i$, for $\alpha = s+1, s+2, \dots, D^i$. We also have $\mathbf{G}_2(v_{C_\beta}, e_{C_\beta a}^2) = 1$, since the direction of edge $e_{C_\beta a}^2$ is assigned as $v_{C_\beta} \rightarrow v_a$, for $\beta = 1, 2, \dots, t$. All the other terms in $\mathbf{G}_1(:, e_{ik_\alpha}^1)$ and $\mathbf{G}_2(:, e_{C_\beta a}^2)$ are zero. The above discussion explains why the equality (40) holds.

Similarly, we can prove that

$$[H_1(G_1^T X H_2 \circ K_{12}^E) G_2^T]_{ia} = \sum_{\alpha=s+1}^{D^i} \sum_{\beta=t+1}^{D^a} X_{k_\alpha C_\beta} k^E(\bar{q}_{ik_\alpha}^1, \bar{q}_{C_\beta a}^2), \quad (42)$$

$$[G_1(H_1^T X G_2 \circ K_{12}^E) H_2^T]_{ia} = \sum_{\alpha=1}^s \sum_{\beta=1}^t X_{k_\alpha C_\beta} k^E(\bar{q}_{ik_\alpha}^1, \bar{q}_{C_\beta a}^2), \quad (43)$$

$$[G_1(H_1^T X H_2 \circ K_{12}^E) G_2^T]_{ia} = \sum_{\alpha=1}^s \sum_{\beta=t+1}^{D^a} X_{k_\alpha C_\beta} k^E(\bar{q}_{ik_\alpha}^1, \bar{q}_{C_\beta a}^2). \quad (44)$$

Adding (34), (42), (43), and (44), we can obtain

$$\begin{aligned} & [H_1(G_1^T X G_2 \circ K_{12}^E) H_2^T + H_1(G_1^T X H_2 \circ K_{12}^E) G_2^T + G_1(H_1^T X G_2 \circ K_{12}^E) H_2^T + \\ & G_1(H_1^T X H_2 \circ K_{12}^E) G_2^T]_{ia} = \sum_{e_{ik}^1 \in \mathcal{E}_1, e_{ca}^2 \in \mathcal{E}_2} X_{kc} k^E(\bar{q}_{ik}^1, \bar{q}_{ca}^2) = [(\Psi^{(1)} \odot X) * \Psi^{(2)}]_{ia}, \end{aligned} \quad (45)$$

where the equality holds because of the fact that $e_{ik_\alpha}^1 \in \mathcal{E}_1$, $\alpha = 1, 2, \dots, D^i$, and $e_{C_\beta a}^2 \in \mathcal{E}_2$, $\beta = 1, 2, \dots, D^a$.

We finish the proof of the equality (27)!

(2).

For proving the equality (25) and (26), we can write

$$\Psi^{(1)} * \Psi^{(1)} = (\Psi^{(1)} \odot I) * \Psi^{(1)} \quad \text{and} \quad \Psi^{(2)} * \Psi^{(2)} = (\Psi^{(2)} \odot I) * \Psi^{(2)}, \quad (46)$$

and directly apply the proved equality (27). \square

2 Proofs of the theoretical results of the EnFW algorithm

The EnFW algorithm is used to solve the following optimization problem

$$\min_{\mathbf{X}} F_\alpha(\mathbf{X}) = J_\alpha(\mathbf{X}) + \lambda H(\mathbf{X}) \quad \mathbf{X} \in \mathcal{D}_n, \quad (47)$$

where $J_\alpha(\mathbf{X})$ is a quadratic function with respect to \mathbf{X} , $H(\mathbf{X}) = \sum_{i=1}^n \sum_{j=1}^n X_{ij} \log(X_{ij})$, $\lambda > 0$, and $\mathcal{D}_n = \{\mathbf{X} \in \mathbb{R}_+^{n \times n} | \mathbf{X} \mathbf{1} = \frac{1}{n} \mathbf{1}, \mathbf{X}^T \mathbf{1} = \frac{1}{n} \mathbf{1}\}$. Note that $H(\mathbf{X})$ is convex with respect to \mathbf{X} .

Write the quadratic function $J_\alpha(\mathbf{X} + s(\mathbf{Y} - \mathbf{X}))$ as

$$J_\alpha(\mathbf{X} + s(\mathbf{Y} - \mathbf{X})) = J_\alpha(\mathbf{X}) + s \langle \nabla J_\alpha(\mathbf{X}), \mathbf{Y} - \mathbf{X} \rangle_F + \frac{1}{2} \text{vec}(\mathbf{Y} - \mathbf{X})^T \nabla^2 J_\alpha(\mathbf{X}) \text{vec}(\mathbf{Y} - \mathbf{X}) s^2,$$

Define the coefficient of the quadratic term with respect to s as

$$Q(\mathbf{X}, \mathbf{Y}) \triangleq \frac{1}{2} \text{vec}(\mathbf{Y} - \mathbf{X})^T \nabla^2 J_\alpha(\mathbf{X}) \text{vec}(\mathbf{Y} - \mathbf{X}) = \frac{1}{2} \langle \nabla J_\alpha(\mathbf{Y} - \mathbf{X}), \mathbf{Y} - \mathbf{X} \rangle_F. \quad (48)$$

Define the gap function $g(\mathbf{X})$ as

$$g(\mathbf{X}) = \langle \nabla J_\alpha(\mathbf{X}), \mathbf{X} \rangle_F + \lambda H(\mathbf{X}) - \min_{\mathbf{Y} \in \mathcal{D}_n} \langle \nabla J_\alpha(\mathbf{X}), \mathbf{Y} \rangle_F + \lambda H(\mathbf{Y}). \quad (49)$$

The algorithm is detailed as

2.1 Proving proposition 2

Proposition 2. If \mathbf{X}^* is an optimal solution of (47), then $g(\mathbf{X}^*) = 0$.

Algorithm 1 The EnFW optimization algorithm for minimizing F_α (47)

- 1: Initialize $\mathbf{X}_0 \in \mathcal{D}_n$
 - 2: **while** not converge **do**
 - 3: Compute the gradient $\nabla J_\alpha(\mathbf{X}_t)$ (14),
 - 4: Obtain the optimal direction \mathbf{Y}_t , i.e., $\mathbf{Y}_t = \operatorname{argmin}_{\mathbf{Y} \in \mathcal{D}_n} \langle \nabla J_\alpha(\mathbf{X}_t), \mathbf{Y} \rangle_F + \lambda H(\mathbf{Y})$,
 - 5: Compute $G_t = g(\mathbf{X}_t)$ and $Q_t = Q(\mathbf{X}_t, \mathbf{Y}_t)$,
 - 6: Determine the stepsize s_t : If $Q_t \leq 0$; $s_t = 1$, else $s_t = \min\{G_t/(2Q_t), 1\}$,
 - 7: Update $\mathbf{X}_{t+1} = \mathbf{X}_t + s_t(\mathbf{Y}_t - \mathbf{X}_t)$.
 - 8: **end**
 - 9: Output the solution \mathbf{X}_α^* .
-

Proof. It is sufficient to show that

$$\langle \nabla J_\alpha(\mathbf{X}^*), \mathbf{X}^* \rangle_F + \lambda H(\mathbf{X}^*) = \min_{\mathbf{Y} \in \mathcal{D}_n} \langle \nabla J_\alpha(\mathbf{X}^*), \mathbf{Y} \rangle_F + \lambda H(\mathbf{Y}).$$

For any $\mathbf{Y} \in \mathcal{D}_n$, we have that $\mathbf{X}^* + s(\mathbf{Y} - \mathbf{X}^*) \in \mathcal{D}_n$, $\forall s \in [0, 1]$, because \mathcal{D}_n is a convex set. Since \mathbf{X}^* is an optimal solution of (47), we have

$$J_\alpha(\mathbf{X}^*) + \lambda H(\mathbf{X}^*) \leq J_\alpha(\mathbf{X}^* + s(\mathbf{Y} - \mathbf{X}^*)) + \lambda H(\mathbf{X}^* + s(\mathbf{Y} - \mathbf{X}^*)) \quad (50)$$

$$\leq J_\alpha(\mathbf{X}^* + s(\mathbf{Y} - \mathbf{X}^*)) + \lambda[(1-s)H(\mathbf{X}^*) + sH(\mathbf{Y})], \quad (51)$$

where the inequality holds because $H(\mathbf{X})$ is a convex function. We reorder the above inequality, and get

$$\frac{J_\alpha(\mathbf{X}^* + s(\mathbf{Y} - \mathbf{X}^*)) - J_\alpha(\mathbf{X}^*)}{s} \geq \lambda H(\mathbf{X}^*) - \lambda H(\mathbf{Y}). \quad (52)$$

Taking the limit $s \rightarrow 0$ of (52) yields

$$\langle \nabla J_\alpha(\mathbf{X}^*), \mathbf{Y} - \mathbf{X}^* \rangle_F \geq \lambda H(\mathbf{X}^*) - \lambda H(\mathbf{Y}). \quad (53)$$

Reordering the above equality yields

$$\forall \mathbf{Y} \in \mathcal{D}_n, \langle \nabla J_\alpha(\mathbf{X}^*), \mathbf{X}^* \rangle_F + \lambda H(\mathbf{X}^*) \leq \langle \nabla J_\alpha(\mathbf{X}^*), \mathbf{Y} \rangle_F + \lambda H(\mathbf{Y}) \quad (54)$$

$$\iff \langle \nabla J_\alpha(\mathbf{X}^*), \mathbf{X}^* \rangle_F + \lambda H(\mathbf{X}^*) = \min_{\mathbf{Y} \in \mathcal{D}_n} \langle \nabla J_\alpha(\mathbf{X}^*), \mathbf{Y} \rangle_F + \lambda H(\mathbf{Y}). \quad (55)$$

□

2.2 Proving Theorem 1

Theorem 1. *The generated objective function value sequence, $\{F_\alpha(\mathbf{X}_t)\}_{t=0}$, will decreasingly converge. The generated points sequence, $\{\mathbf{X}_t\}_{t=0} \subseteq \mathcal{D}_n \subseteq \mathbb{R}^{n \times n}$, will weakly converge to the first-order stationary point, at the rate $O(\frac{1}{\sqrt{t+1}})$, i.e,*

$$\min_{1 \leq t \leq T} g(\mathbf{X}_t) \leq \frac{2 \max\{\Delta_0, \sqrt{L\Delta_0/n}\}}{\sqrt{T+1}}, \quad (56)$$

where $\Delta_0 = F_\alpha(\mathbf{X}_0) - \min_{\mathbf{X} \in \mathcal{D}_n} F_\alpha(\mathbf{X})$, and L is the largest absolute eigenvalue of $\nabla^2 J_\alpha(\mathbf{X})$.

Before we prove Theorem 1, we introduce a lemma.

Lemma 3. *The generated objective function value sequence, $\{F_\alpha(\mathbf{X}_t)\}_{t=0}$, satisfies*

$$F_\alpha(\mathbf{X}_{t+1}) \leq F_\alpha(\mathbf{X}_t) - (G_t - \frac{L}{n}s)s, \quad \forall s \in [0, 1]. \quad (57)$$

Proof. Since $H(\mathbf{X})$ is convex, we have

$$\forall s \in [0, 1], \lambda H(\mathbf{X}_t + s(\mathbf{Y}_t - \mathbf{X}_t)) \leq \lambda H(\mathbf{X}_t) + s(\lambda H(\mathbf{Y}_t) - \lambda H(\mathbf{X}_t)). \quad (58)$$

Write the quadratic function $J_\alpha(\mathbf{X}_t + s(\mathbf{Y}_t - \mathbf{X}_t))$ as

$$J_\alpha(\mathbf{X}_t + s(\mathbf{Y}_t - \mathbf{X}_t)) = J_\alpha(\mathbf{X}_t) + s\langle \nabla J_\alpha(\mathbf{X}_t), \mathbf{Y}_t - \mathbf{X}_t \rangle_F + Q(\mathbf{X}_t, \mathbf{Y}_t)s^2. \quad (59)$$

Adding (58) and (59), and reordering the resulting inequality, we have

$$F_\alpha(\mathbf{X}_t + s(\mathbf{Y}_t - \mathbf{X}_t)) \leq F_\alpha(\mathbf{X}_t) + s[\lambda H(\mathbf{Y}_t) - \lambda H(\mathbf{X}_t) + \langle \nabla J_\alpha(\mathbf{X}_t), \mathbf{Y}_t - \mathbf{X}_t \rangle_F] + Q(\mathbf{X}_t, \mathbf{Y}_t)s^2. \quad (60)$$

Since $\lambda H(\mathbf{Y}_t) + \langle \nabla J_\alpha(\mathbf{X}_t), \mathbf{Y}_t \rangle_F = \min_{\mathbf{Y} \in \mathcal{D}_n} \langle \nabla J_\alpha(\mathbf{X}_t), \mathbf{Y} \rangle_F + \lambda H(\mathbf{Y})$ (See the 4th line in Algorithm 1), we have

$$G_t = g(\mathbf{X}_t) = -[\lambda H(\mathbf{Y}_t) - \lambda H(\mathbf{X}_t) + \langle \nabla J_\alpha(\mathbf{X}_t), \mathbf{Y}_t - \mathbf{X}_t \rangle_F], \quad (61)$$

which is based on the definition of $g(\mathbf{X})$ (49) and G_t (See the 5th line in Algorithm 1). Substituting (61) into (60), we have

$$\forall s \in [0, 1], F_\alpha(\mathbf{X}_t + s(\mathbf{Y}_t - \mathbf{X}_t)) \leq F_\alpha(\mathbf{X}_t) - (G_t - Q_t s)s. \quad (62)$$

Set $s = s_t$ (See the 6th and 7th line in Algorithm 1), we have

$$F_\alpha(\mathbf{X}_{t+1}) \leq F_\alpha(\mathbf{X}_t) - (G_t - Q_t s_t)s_t.$$

Now we consider the function

$$A_t(s) = (G_t - Q_t s)s, s \in [0, 1].$$

We discuss the maximizer of $A_t(s)$ for $s \in [0, 1]$. Considering that G_t is nonnegative (this is because of the definition of $g(\mathbf{X})$), we have

1. If $Q_t \leq 0$, then $A_t(s)$ achieve its maximum at $s = 1$,
2. If $Q_t > 0$, then $A_t(s)$ achieve its maximum at $s = \min\{G_t/(2Q_t), 1\}$.

Therefore, our stepsize s_t is just the maximizer of $A_t(s)$. That is

$$\begin{aligned} F_\alpha(\mathbf{X}_{t+1}) &\leq F_\alpha(\mathbf{X}_{t+1}) \leq F_\alpha(\mathbf{X}_t) - (G_t - Q_t s_t)s_t = F_\alpha(\mathbf{X}_t) - \max_{s \in [0, 1]} (G_t - Q_t s)s \\ &\leq F_\alpha(\mathbf{X}_t) - (G_t - Q_t s)s \quad \forall s \in [0, 1]. \end{aligned} \quad (63)$$

Since

$$Q_t = Q(\mathbf{X}_t, \mathbf{Y}_t) = \frac{1}{2} \text{vec}(\mathbf{Y}_t - \mathbf{X}_t)^T \nabla^2 J_\alpha(\mathbf{X}_t) \text{vec}(\mathbf{Y}_t - \mathbf{X}_t) \leq \frac{L}{2} \|\mathbf{X}_t - \mathbf{Y}_t\|_F^2,$$

and

$$\|\mathbf{X} - \mathbf{Y}\|_F^2 \leq \frac{2}{n}, \forall \mathbf{X}, \mathbf{Y} \in \mathcal{D}_n,$$

we have $Q_t \leq \frac{L}{n}$.

Combining (63) and the fact $Q_t \leq \frac{L}{n}$, we obtain the desired result. \square

Now we prove Theorem 1

Proof. We consider the inequality (57) in Lemma 3.

If $G_t > \frac{2L}{n}$, then $(G_t - \frac{L}{n}s)s$ is maximized at $s = 1$. So

$$F_\alpha(\mathbf{X}_{t+1}) \leq F_\alpha(\mathbf{X}_t) - (G_t - \frac{L}{n}) \leq F_\alpha(\mathbf{X}_t) - \frac{G_t}{2},$$

where the last equality holds because $\frac{L}{n} \leq \frac{G_t}{2}$.

If $G_t \leq \frac{2L}{n}$, then $(G_t - \frac{L}{n}s)s$ is maximized at $s = \frac{nG_t}{2L}$. So

$$F_\alpha(\mathbf{X}_{t+1}) \leq F_\alpha(\mathbf{X}_t) - \left(\frac{G_t}{2}\right) \frac{nG_t}{2L}.$$

In summary,

$$F_\alpha(\mathbf{X}_{t+1}) \leq F_\alpha(\mathbf{X}_t) - \frac{G_t}{2} \min\{1, \frac{nG_t}{2L}\} \quad (64)$$

(I.) Since $G_t \geq 0$ by definition (49), we have $F_\alpha(\mathbf{X}_{t+1}) \leq F_\alpha(\mathbf{X}_t)$, i.e., $\{F_\alpha(\mathbf{X}_t)\}_{t=0}$ is a decreasing sequence. $F_\alpha(\mathbf{X})$ is continuous on the compact region \mathcal{D}_n , which implies that $F_\alpha(\mathbf{X})$ is bounded below. So the sequence $\{F_\alpha(\mathbf{X}_t)\}_{t=0}$ will converge [3].

(II.) Taking the sum of (64) over $t = 0, 1, 2, \dots, T$, we obtain,

$$F_\alpha(\mathbf{X}_{T+1}) - F_\alpha(\mathbf{X}_0) \leq - \sum_{t=0}^T \frac{G_t}{2} \min\{1, \frac{G_t n}{2L}\}. \quad (65)$$

Let

$$G_T^* = \min_{0 \leq t \leq T} G_t = \min_{0 \leq t \leq T} G(\mathbf{X}_t).$$

Considering the additional fact that $-\Delta_0 \triangleq F_\alpha(\mathbf{X}^*) - F_\alpha(\mathbf{X}_0) \leq F_\alpha(\mathbf{X}_{T+1}) - F_\alpha(\mathbf{X}_0)$, we have

$$\Delta_0 \geq \sum_{t=0}^T \frac{G_t}{2} \min\{1, \frac{G_t n}{2L}\} \geq (T+1) \frac{G_T^*}{2} \min\{1, \frac{G_T^* n}{2L}\}.$$

(a.) If $\frac{nG_T^*}{2L} \geq 1$, then $\Delta_0 \geq (T+1) \frac{G_T^*}{2} \iff G_T^* \leq \frac{2\Delta_0}{T+1} \leq \frac{2\Delta_0}{\sqrt{T+1}}$;

(b.) If $\frac{nG_T^*}{2L} < 1$, then $\Delta_0 \geq (T+1) (\frac{G_T^*}{2}) \frac{nG_T^*}{2L} \iff G_T^* \leq \sqrt{\frac{4L\Delta_0}{n(T+1)}}$.

In summary, we have $G_T^* \leq \frac{2 \max\{\Delta_0, \sqrt{L\Delta_0/n}\}}{\sqrt{T+1}}$. □

2.3 Proving Theorem 2

Theorem 2. If $J_\alpha(\mathbf{X})$ is convex, we have $F_\alpha(\mathbf{X}_T) - F_\alpha(\mathbf{X}^*) \leq \frac{4L}{n(T+1)}$.

Before we prove Theorem 2, we introduce a lemma.

Lemma 4. If $J_\alpha(\mathbf{X})$ is convex, and let \mathbf{X}^* is a global minimizer of problem (47), then $g(\mathbf{X}_t) \geq F_\alpha(\mathbf{X}_t) - F_\alpha(\mathbf{X}^*)$.

Proof.

$$g(\mathbf{X}_t) = \langle \nabla J_\alpha(\mathbf{X}_t), \mathbf{X}_t \rangle_F + \lambda H(\mathbf{X}_t) - \min_{\mathbf{Y} \in \mathcal{D}_n} \langle \nabla J_\alpha(\mathbf{X}_t), \mathbf{Y} \rangle_F + \lambda H(\mathbf{Y}). \quad (66)$$

$$= \max_{\mathbf{Y} \in \mathcal{D}_n} \left\{ \langle \nabla J_\alpha(\mathbf{X}_t), \mathbf{X}_t - \mathbf{Y} \rangle_F + \lambda H(\mathbf{X}_t) - \lambda H(\mathbf{Y}) \right\} \quad (67)$$

$$\geq \langle \nabla J_\alpha(\mathbf{X}_t), \mathbf{X}_t - \mathbf{X}^* \rangle_F + \lambda H(\mathbf{X}_t) - \lambda H(\mathbf{X}^*) \quad (68)$$

$$\geq J_\alpha(\mathbf{X}_t) - J_\alpha(\mathbf{X}^*) + \lambda H(\mathbf{X}_t) - \lambda H(\mathbf{X}^*) \quad (69)$$

$$= F_\alpha(\mathbf{X}_t) - F_\alpha(\mathbf{X}^*), \quad (70)$$

where the inequality (69) holds because $J_\alpha(\mathbf{X})$ is convex. □

Now we prove Theorem 2.

Proof. We still need the inequality (57) in Lemma 3.

$$\begin{aligned} F_\alpha(\mathbf{X}_{t+1}) &\leq F_\alpha(\mathbf{X}_t) - (G_t - \frac{L}{n}s)s \\ &\leq F_\alpha(\mathbf{X}_t) - [F_\alpha(\mathbf{X}_t) - F_\alpha(\mathbf{X}^*)]s + \frac{L}{n}s^2, \quad \forall s \in [0, 1], \end{aligned} \quad (71)$$

where the inequality (71) holds because of Lemma 4. Reordering (71) yields

$$[F_\alpha(\mathbf{X}_{t+1}) - F_\alpha(\mathbf{X}^*)] \leq (1-s)[F_\alpha(\mathbf{X}_t) - F_\alpha(\mathbf{X}^*)] + \frac{L}{n}s^2. \quad (72)$$

We set $s = \frac{2}{t+1}$, and multiply $\frac{t(t+1)}{2}$ on both sides. Then we have

$$\begin{aligned} \frac{t(t+1)}{2} [F_\alpha(\mathbf{X}_{t+1}) - F_\alpha(\mathbf{X}^*)] &\leq \frac{t(t-1)}{2} [F_\alpha(\mathbf{X}_t) - F_\alpha(\mathbf{X}^*)] + \frac{L}{n} \frac{2t}{t+1} \\ &\leq \frac{t(t-1)}{2} [F_\alpha(\mathbf{X}_t) - F_\alpha(\mathbf{X}^*)] + \frac{2L}{n} \end{aligned} \quad (73)$$

Taking the sum of (73) over $t = 1, 2, \dots, T$, we have

$$\frac{T(T+1)}{2} [F_\alpha(\mathbf{X}_{T+1}) - F_\alpha(\mathbf{X}^*)] \leq \frac{2TL}{n} \iff F_\alpha(\mathbf{X}_{T+1}) - F_\alpha(\mathbf{X}^*) \leq \frac{4L}{n(T+1)}. \quad (74)$$

We obtain the desired result. \square

3 Experiments

In the first part, we present the implementation details of the experiments in the paper, including the description of Sinkhorn–Knopp algorithm, and convergence criterions for EnFW.

3.1 Implementation details

The Sinkhorn–Knopp algorithm [4, 1] is used to solve the problem

$$\min_{\mathbf{X}} \langle \mathbf{D}, \mathbf{X} \rangle_{\text{F}} + \lambda H(\mathbf{X}) \quad \text{s.t.} \quad \mathbf{X} \geq 0, \quad \mathbf{X} \vec{\mathbf{1}}_n = \vec{\mathbf{a}}, \text{ and } \mathbf{X}^T \vec{\mathbf{1}}_n = \vec{\mathbf{b}}, \quad (75)$$

where $\vec{\mathbf{a}}$ and $\vec{\mathbf{b}}$ satisfy $\sum_{i=1}^n \vec{\mathbf{a}}_i = \sum_{j=1}^n \vec{\mathbf{b}}_j = 1$. In our case $\vec{\mathbf{a}} = \frac{1}{n} \vec{\mathbf{1}}_n$ and $\vec{\mathbf{b}} = \frac{1}{n} \vec{\mathbf{1}}_n$.

The algorithm is shown in Algorithm 2. The detailed derivations can be found in [1]. In our

Algorithm 2 The Sinkhorn–Knopp optimization algorithm for minimizing (75)

- 1: Initialize $\vec{\mathbf{r}} = \frac{1}{n} \vec{\mathbf{1}}_n$, $t = 0$, and RelativeError = 1, and write $\mathbf{C} = \exp(-\frac{\mathbf{D}}{\lambda})$ (pointwise).
 - 2: **while** $t < \text{MaxNum}_{\text{Sinkhorn}}$ and RelativeError > Tolerance **do**
 - 3: $\vec{\mathbf{r}}_{\text{new}} = (\frac{1}{n} \vec{\mathbf{1}}_n) ./ \left\{ \mathbf{C} \left[(\frac{1}{n} \vec{\mathbf{1}}_n) ./ (\mathbf{C}^T \vec{\mathbf{r}}) \right] \right\}$.
 - 4: RelativeError = $\frac{\|\vec{\mathbf{r}}_{\text{new}} - \vec{\mathbf{r}}\|_2}{\|\vec{\mathbf{r}}\|_2}$.
 - 5: $t = t + 1$
 - 6: **end**
 - 7: $\vec{\mathbf{s}} = (\frac{1}{n} \vec{\mathbf{1}}_n) ./ (\mathbf{C}^T \vec{\mathbf{r}})$, $\mathbf{X}^* = \text{diag}(\vec{\mathbf{r}}) \mathbf{C} \text{diag}(\vec{\mathbf{s}})$.
 - 8: Output the solution \mathbf{X}^* .
-

experiments, we set $\text{MaxNum}_{\text{Sinkhorn}} = 10000$, and set Tolerance = 10^{-6} .

Convergence

fied for continuing the iterations in Algorithm 1.

1. The maximal number of iterations is set to be 1000, i.e., $t < 1000$;
2. The tolerance of relative error between \mathbf{X}_{t+1} and \mathbf{X}_t is set to be 10^{-8} , i.e., $\frac{\|\mathbf{X}_{t+1} - \mathbf{X}_t\|_{\text{F}}}{\|\mathbf{X}_t\|_{\text{F}}} > 10^{-8}$.
3. The gap $G_t = g(\mathbf{X}_t)$ should be greater than *eps* to avoid numerical error, where *eps* is the spacing of floating point numbers of Matlab.

3.2 Experimental results

We re-implement the experiments in Section 6.1 in the paper, for comparing the matching performance of different algorithms. Here, we test the performance of KerGM_I on synthetic datasets. That is, we use the exact edge affinity kernel¹ and compute the gradients by using the formulas in (25), (26), and (27). We still consider the same parameter settings:

¹Note that in the paper, we test the performance of KerGM_{II} on synthetic datasets. We in fact use the Fourier random features [2] to approximate the kernel.

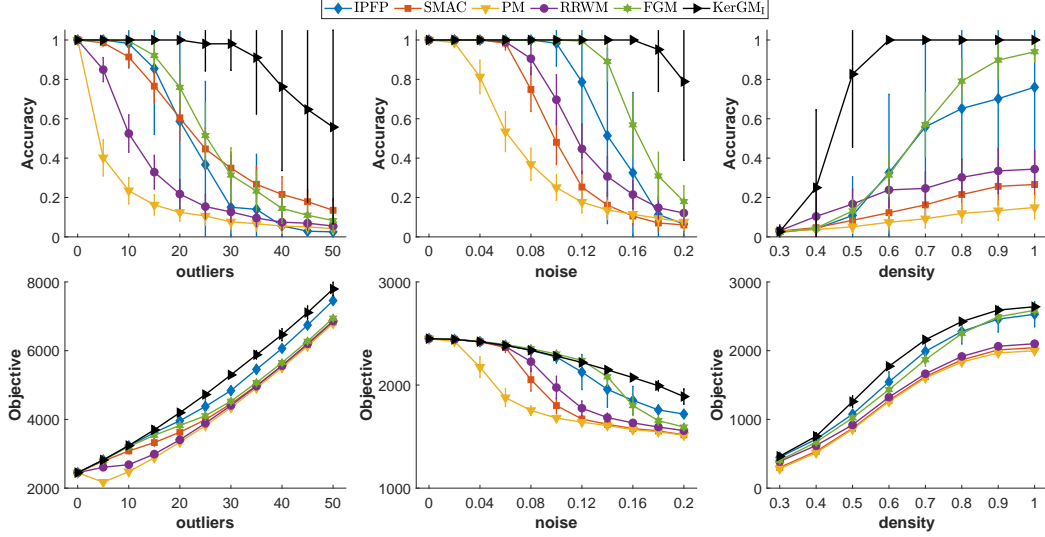


Figure 2: Comparison of graph matching algorithms on synthetic datasets

1. We change the number of outlier nodes, n_{out} , from 0 to 50 while fixing the noise, $\sigma = 0$, and the edge density, $\rho = 1$.
2. We change σ from 0 to 0.2 while fixing $n_{\text{out}} = 0$ and $\rho = 1$.
3. We change ρ from 0.3 to 1 while fixing $n_{\text{out}} = 5$ and $\sigma = 0.1$.

We repeat the experiments 100 times, and report the average matching accuracies and objective function values. In Fig. 2, we show the results. It can be seen that our KerGM_I still consistently outperforms the baselines.

References

- [1] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013.
- [2] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in neural information processing systems*, pages 1177–1184, 2008.
- [3] Halsey Lawrence Royden. *Real analysis*. Krishna Prakashan Media, 1968.
- [4] Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.