Adaptive Softassign via Hadamard-Equipped Sinkhorn

Supplementary Material

A. More details of the projected fixed point method

Consider the objective function

$$\mathcal{Z}(M) = \frac{1}{2}\operatorname{tr}\left(M^T A M \widetilde{A}\right) + \lambda \operatorname{tr}\left(M^T K\right). \quad (31)$$

Given an initial condition $M^{(k)}$, we can linearize the objective function at $M^{(k)}$ via the Taylor series approximation:

$$\mathcal{Z}(M) \approx \mathcal{Z}(M^{(k)}) + \operatorname{tr}\left\{\nabla \mathcal{Z}(M^{(k)})^T (M - M^{(k)})\right\} \quad (32)$$

One can find an approximate solution to (31) by maximizing a sequence of the linearization of $\mathcal{Z}(M)$ in (32). Since $M^{(k)}$ is a constant, the maximization of the linear approximation is equivalent to the following linear assignment problem

$$\max_{M \in \Sigma_{n\times n}} \operatorname{tr}(\nabla \mathcal{Z}(M^{(k)})^T M) = \max_{M \in \Sigma_{n\times n}} \langle \nabla \mathcal{Z}(M^{(k)}), M \rangle, \quad (33)$$

Therefore, the quadratic assignment problem can be transformed into a series of linear assignment problems. This idea is first proposed in [8] with a different objective function. The solution to each linear assignment problem in (33) can be shown as

$$M^{(k)} = \mathcal{P}(\nabla \mathcal{Z}(M^{(k-1)})). \quad (34)$$

where P is a projection that includes the doubly stochastic projection used in [23] and the discrete projection used in [19]. A generation of such an iterative formula is

$$M^{(k)} = (1-\alpha) M^{(k-1)} + \alpha \mathcal{P}(\nabla \mathcal{Z}(M^{(k-1)})). \quad (35)$$

Such a formula can cover all the points between two doubly stochastic matrices. The resulting algorithm is called the *projected fixed-point method* [19].

B. Proofs of Proposition 1 and Proposition 2

Proposition 1 *For a square matrix X and $\beta > 0$, we have*

$$|\langle S_X^\beta, X\rangle - \langle S_X^\infty, X\rangle| \le \|S_X^\beta - S_X^\infty\| \|X\| \le \frac{c}{\mu}(e^{(-\mu\beta)})\|X\|, \quad (36)$$

where c and $\mu > 0$ are constants independent of β.

Proposition 2 *For a square matrix X and β, $\Delta\beta > 0$, we have*

$$\|S_X^\beta - S_X^{\beta+\Delta\beta}\| \le (1 - e^{(-\mu\Delta\beta)})\frac{c}{\mu} e^{(-\mu\beta)}, \quad (37)$$

where c and $\mu > 0$ are constants independent of β.

Proof *We first transform the problem (12) into vector form*

$$\langle M, X\rangle + \frac{1}{\beta}\mathcal{H}(M) = \mathbf{m}^T \mathbf{x} - \frac{1}{\beta}\sum \mathbf{m}_i \ln(\mathbf{m}_i), \quad (38)$$

where $\mathbf{m} = \operatorname{vec} M$. Since $\sum_i^{n^2} \mathbf{m}_i = n$, then (12) is equivalent to the well studied problem [4]

$$\mathbf{s}_{\mathbf{x}}^\beta = \operatorname*{argmin}_{\mathbf{m}} \mathbf{m}^T(-\mathbf{x}) + \frac{1}{\beta}\sum_i \mathbf{m}_i(\ln(\mathbf{m}_i) - 1). \quad (39)$$

Let $\dot{S}_X^\beta$ be derivative of S_X^β with respect to β, according to the proof of [4, Proposition 5.1], $\dot{S}_X^\beta$ converges towards zero exponentially i.e., there exist a $c_0 > 0$ and $\mu > 0$ such that

$$|(\dot{S}_X^\beta)_{ij}| \le c_0 e^{-\mu\beta}.$$

According to the fundamental theory of Calculus,

$$\begin{aligned} |(S_X^\infty)_{ij} - (S_X^\beta)_{ij}| &= |\int_\beta^\infty (\dot{S}_X^\tau)_{ij} d\tau| \le \int_\beta^\infty |(\dot{S}_X^\tau)_{ij}| d\tau \\ &\le \frac{c_0}{\mu}(e^{(-\mu\beta)}). \end{aligned} \quad (40)$$

Similarly, we have

$$\begin{aligned} |(S_X^\beta)_{ij} - (S_X^{\beta+\Delta\beta})_{ij}| &= |\int_\beta^{\beta+\Delta\beta} (\dot{S}_X^\tau)_{ij} d\tau| \\ &\le \frac{c_0}{\mu}(e^{(-\mu\beta)} - e^{(-\mu(\beta+\Delta\beta))}). \end{aligned} \quad (41)$$

The rest of the proof for the two propositions follows easily from this.

C. Proof of Proposition 3

Proposition 3 *Hadamard-Equipped Sinkhorn*
Let $X \in \mathbb{R}_+^{n\times n}$, then

$$\mathcal{P}_{sk}(X) = X \circ SK^{(X)} = X \circ (\mathbf{r}^T \otimes \mathbf{c}) \quad (42)$$

where $SK^{(X)} \in \mathbb{R}^{n\times n}$ is unique, $\mathbf{r}$ and $\mathbf{c} \in \mathbb{R}_+^n$ are balancing vectors so that $D_{(\mathbf{r})} X D_{(\mathbf{c})}$ is a doubly stochastic matrix.

Proof

$$\begin{aligned} \mathcal{P}_{sk}(X) &= D_{(\mathbf{r})} X D_{(\mathbf{c})} \\ &= X \circ \underbrace{(\mathbf{r}^T \otimes \mathbf{c})}_{SK^{(X)}}. \end{aligned} \quad (43)$$

D. Proofs of Lemma 1, Lemma 2, and Lemma 3

Lemma 1 *Let $X \in \mathbb{R}_{+}^{n\times n}$, $\mathbf{u}$ and $\mathbf{v} \in \mathbb{R}_{+}^{n}$, then*

$$\mathcal{P}_{sk}(X) = \mathcal{P}_{sk}(X \circ (\mathbf{u}^T \otimes \mathbf{v})). \tag{44}$$

Proof *Let $Y = X \circ (\mathbf{u}^T \otimes \mathbf{v})$, we have*

$$\mathcal{P}_{sk}(Y) = Y \circ (\mathbf{r}_Y^T \otimes \mathbf{c}_Y). \tag{45}$$

Then

$$\begin{aligned}
\mathcal{P}_{sk}(X \circ (\mathbf{u}^T \otimes \mathbf{v})) &= X \circ (\mathbf{u}^T \otimes \mathbf{v}) \circ (\mathbf{r}_Y^T \otimes \mathbf{c}_Y) \\
&= X \circ ((\underbrace{\mathbf{u} \circ \mathbf{r}_Y}_{\mathbf{r}_1})^T \otimes (\underbrace{\mathbf{v} \circ \mathbf{c}_Y}_{\mathbf{c}_1})) \\
&= X \circ (\mathbf{r}_1^T \otimes \mathbf{c}_1) \\
&= \mathcal{P}_{sk}(X).
\end{aligned} \tag{46}$$

Since $X \circ (\mathbf{r}_1^T \otimes \mathbf{c}_1)$ is a doubly stochastic matrix, $\mathbf{r}_1^T \otimes \mathbf{c}_1 = SK^{(X)}$ according to the Proposition 3.

Lemma 2 *Sinkhorn-Hadamard product*
Let $X_1, X_2 \in \mathbb{R}_{+}^{n\times n}$, then $\mathcal{P}_{sk}(X_1 \circ X_2) = \mathcal{P}_{sk}(\mathcal{P}_{sk}(X_1) \circ X_2)$.

Proof *According to Lemma 1, the right-hand side is*

$$\mathcal{P}_{sk}(\overbrace{\mathcal{P}_{sk}(X_1)} \circ X_2) = \mathcal{P}_{sk}(\overbrace{X_1 \circ SK^{(X_1)}} \circ X_2) \tag{47}$$

$$= \mathcal{P}_{sk}(X_1 \circ X_2), \tag{48}$$

which proves this Lemma.

Lemma 3 *Sinkhorn-Hadamard power*
Let $X_1, X_2 \in \mathbb{R}_{+}^{n\times n}$, then $\mathcal{P}_{sk}(X^{\circ(ab)}) = \mathcal{P}_{sk}(\mathcal{P}_{sk}(X^{\circ a})^{\circ b})$, where a and b are two constants not equal to zero.

Proof *According to Lemma 1, the right-hand side is*

$$\mathcal{P}_{sk}(\mathcal{P}_{sk}(X^{\circ a})^{\circ b}) = \mathcal{P}_{sk}((X^{\circ a} \circ SK^{(X^{\circ a})})^{\circ b}) \tag{49}$$

$$= \mathcal{P}_{sk}(X^{\circ(ab)} \circ (SK^{(X^{\circ a})})^{\circ b}) \tag{50}$$

$$= \mathcal{P}_{sk}(X^{\circ(ab)}) \tag{51}$$

which completes the proof.

E. Relation with the proximal point method

In this subsection, we shall demonstrate the equivalence and difference between the adaptive softassign and the proximal point method proposed by [40]. The linear convergence rate of the adaptive softassign methods can be inferred from the convergence of the proximal point method. While the difference brings computational efficiency.

Proposition 4 *The softassign transition (25) can solve*

$$S_X^{\beta_2} = \arg\max_{s \in \Sigma_{n\times n}} \langle X, S\rangle - (\beta_2 - \beta_1) D_h\left(S, S_X^{\beta_1}\right), \tag{52}$$

$$D_h(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} x_i \log \frac{x_i}{y_i} - \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i. \tag{53}$$

Proof *The solution of (52) in the proximal point method is*

$$\mathcal{P}_{sk}(S_X^{\beta_1} \circ \exp((\beta_2 - \beta_1)X)). \tag{54}$$

According to the Hadamard-Equipped Sinkhorn Theorem, we have

$$S_X^{\beta_1} = \exp(\beta_1 X) \circ SK^{(\exp(\beta_1 X))}. \tag{55}$$

Then

$$\begin{aligned}
&\mathcal{P}_{sk}(S_X^{\beta_1} \circ \exp((\beta_2 - \beta_1)X)) \\
&= \mathcal{P}_{sk}(\exp(\beta_1 X) \circ \exp((\beta_2 - \beta_1)X)) \\
&= \mathcal{P}_{sk}(\exp(\beta_2 X)) \\
&= S_X^{\beta_2},
\end{aligned} \tag{56}$$

which is equivalent to the softassign transition (25).

According to Proposition 4, we can rewrite the iterative formula of the adaptive softassign as a proximal point method in [40]

$$S_X^{(k)} = \arg\max_{s \in \Sigma_{n\times n}} \langle X, S\rangle - (\Delta\beta) D_h\left(S, S_X^{(k-1)}\right), \tag{57}$$

where $D_h(\cdot)$, the *Bregman divergence*, is a regularization term to define the proximal operator. This indicates adaptive softassign is a variant of the proximal point method for problem (12) and enjoys a linear convergence rate [40].

Let us discuss the difference between adaptive softassign and the proximal point method. Adaptive softassign aims at obtaining a sub-optimal solution and β_ϵ with a given error bound, where β_ϵ can be used as a good initial β_0 in the next adaptive softassign in the whole graph matching process. While the proximal point method aims to find the exact solution, its efficiency is secondary and the change of β is implicit. As to the computation aspect, the proximal point method solves (52) according to

$$S_X^{\beta_2} = \mathcal{P}_{sk}(S_X^{\beta_1} \circ \exp((\beta_2 - \beta_1)X)). \tag{58}$$

Softassign transition only adapts a power operation and does not need the X, which indicates the change of β more clearly. One can track and analyze the explicit change of β easily.

(a) ubc

(b) wall

(c) trees

(d) leuven

Figure 7. Graphs from real images matching. The yellow lines represent the correspondence between key points of the pictures.

F. Baselines and visualization of experiments

Visualization of the matching results is shown in Figure 7.

Baselines:

- DSPFP [23] is a fast doubly stochastic projected fixed-point method with an alternating projection.
- GA [8] can be considered a softassign-based projected fixed-point method with an outer annealing process.
- AIPFP [19, 23] is an integer projected fixed point method with a fast greedy integer projection.
- SCG [32] is a constrained gradient method with a dynamic softassign invariant to the nodes' number.
- GWL [42] measures the distance between two graphs by Gromov-Wasserstein discrepancy and matches graphs by optimal transport.
- S-GWL [41] is a scalable variant of GWL. It divides matching graphs into small graphs to match.
- MAGNA++ [36] is a global network alignment method for protein-protein interaction network matching, which focuses on node and edge conservation.
- GRASP [11] aligns nodes based on functions derived from Laplacian matrix eigenvectors.