Supplemental Material for "Factorized Graph Matching"

Abstract

In this supplementary material, we provide more details of the four steps of our factorized graph matching (FGM) algorithm: (1) the line search for λ, (2) the modified Frank-Wolfe algorithm, (3) the concave-convex procedure and (4) the local vs. global strategy.

1. The line search for λ

The Frank-Wolfe's algorithm (FW) [1] involves two steps: (1) the computation of the optimal direction $\mathbf{Y} \in \mathbb{R}^{n_1 \times n_2}$, and (2) the line search for the optimal step size $\lambda \in [0, 1]$. In the submitted paper, we mentioned that the former issue can be efficiently solved using the Hungarian algorithm. In this supplementary material we solve the second one.

Given the optimal direction $\mathbf{Y} \in \mathbb{R}^{n_1 \times n_2}$, the line search step maximizes $J_\alpha(\mathbf{X} + \lambda\mathbf{Y})$ over $\lambda \in [0, 1]$. The optimal λ can be computed in a closed form by finding the optimum point of the following parabola:

$$\begin{aligned} J_\alpha(\mathbf{X} + \lambda\mathbf{Y}) &= a\lambda^2 + b\lambda + \text{const}, \quad (1) \\ \text{where} \quad a &= (1-\alpha)a_{vex} + \alpha a_{cav}, \\ b &= (1-\alpha)b_{vex} + \alpha b_{cav}. \end{aligned}$$

To compute the convex components, a_{vex} and b_{vex}, we firstly expand the convex relaxation as follows:

$$\begin{aligned} J_{vex}(\mathbf{X}) &= J_{gm}(\mathbf{X}) - \frac{1}{2}C(\mathbf{X}) \\ &= \operatorname{tr}\Big(\mathbf{L}^T(\mathbf{H}_1^T\mathbf{X}\mathbf{H}_2 \circ \mathbf{H}_1^T\mathbf{X}\mathbf{H}_2)\Big) \\ &\quad - \frac{1}{2}\sum_i \operatorname{tr}(\mathbf{A}_i^1\mathbf{A}_i^1\mathbf{X}\mathbf{X}^T) + \operatorname{tr}(\mathbf{A}_i^2\mathbf{A}_i^2\mathbf{X}^T\mathbf{X}) \\ &= \operatorname{tr}\Big(\mathbf{L}^T(\mathbf{H}_1^T\mathbf{X}\mathbf{H}_2 \circ \mathbf{H}_1^T\mathbf{X}\mathbf{H}_2)\Big) \\ &\quad - \frac{1}{2}\sum_i \operatorname{tr}\Big(\operatorname{diag}(\mathbf{u}_i)\mathbf{H}_1^T\mathbf{H}_1 \operatorname{diag}(\mathbf{u}_i)\mathbf{H}_1^T\mathbf{X}\mathbf{X}^T\mathbf{H}_1\Big) \\ &\quad + \operatorname{tr}\Big(\operatorname{diag}(\mathbf{v}_i)\mathbf{H}_2^T\mathbf{H}_2 \operatorname{diag}(\mathbf{v}_i)\mathbf{H}_2^T\mathbf{X}^T\mathbf{X}\mathbf{H}_2\Big) \\ &= \operatorname{tr}\Big(\mathbf{L}^T(\mathbf{H}_1^T\mathbf{X}\mathbf{H}_2 \circ \mathbf{H}_1^T\mathbf{X}\mathbf{H}_2)\Big) \\ &\quad - \frac{1}{2}\operatorname{tr}\Big(\mathbf{U}\mathbf{U}^T(\mathbf{H}_1^T\mathbf{H}_1 \circ \mathbf{H}_1^T\mathbf{X}\mathbf{X}^T\mathbf{H}_1)\Big) \\ &\quad - \frac{1}{2}\operatorname{tr}\Big(\mathbf{V}\mathbf{V}^T(\mathbf{H}_2^T\mathbf{H}_2 \circ \mathbf{H}_2^T\mathbf{X}^T\mathbf{X}\mathbf{H}_2)\Big). \end{aligned}$$

Therefore, we can obtain:

$$\begin{aligned} J_{vex}(\mathbf{X} + \lambda\mathbf{Y}) &= a_{vex}\lambda^2 + b_{vex}\lambda + \text{const}, \\ \text{where } a_{vex} &= \operatorname{tr}\Big(\mathbf{L}^T(\mathbf{H}_1^T\mathbf{Y}\mathbf{H}_2 \circ \mathbf{H}_1^T\mathbf{Y}\mathbf{H}_2)\Big) \\ &\quad - \frac{1}{2}\operatorname{tr}\Big(\mathbf{U}\mathbf{U}^T(\mathbf{H}_1^T\mathbf{H}_1 \circ \mathbf{H}_1^T\mathbf{Y}\mathbf{Y}^T\mathbf{H}_1\Big) \\ &\quad - \frac{1}{2}\operatorname{tr}\Big(\mathbf{V}\mathbf{V}^T(\mathbf{H}_2^T\mathbf{H}_2 \circ \mathbf{H}_2^T\mathbf{Y}^T\mathbf{Y}\mathbf{H}_2\Big), \\ b_{vex} &= 2\operatorname{tr}\Big(\mathbf{L}^T(\mathbf{H}_1^T\mathbf{X}\mathbf{H}_2 \circ \mathbf{H}_1^T\mathbf{Y}\mathbf{H}_2)\Big) \\ &\quad - \operatorname{tr}\Big(\mathbf{U}\mathbf{U}^T(\mathbf{H}_1^T\mathbf{H}_1 \circ \mathbf{H}_1^T\mathbf{X}\mathbf{Y}^T\mathbf{H}_1\Big) \\ &\quad - \operatorname{tr}\Big(\mathbf{V}\mathbf{V}^T(\mathbf{H}_2^T\mathbf{H}_2 \circ \mathbf{H}_2^T\mathbf{X}^T\mathbf{Y}\mathbf{H}_2\Big). \end{aligned}$$

Similarly, we can compute the concave components, a_{cav} and b_{cav} as follows:

$$\begin{aligned} J_{cav}(\mathbf{X} + \lambda\mathbf{Y}) &= a_{cav}\lambda^2 + b_{cav}\lambda + \text{const}, \\ \text{where } a_{cav} &= \operatorname{tr}\Big(\mathbf{K}_q^T(\mathbf{G}_1^T\mathbf{Y}\mathbf{G}_2 \circ \mathbf{G}_1^T\mathbf{Y}\mathbf{G}_2)\Big), \\ b_{cav} &= 2\operatorname{tr}\Big(\mathbf{K}_q^T(\mathbf{G}_1^T\mathbf{X}\mathbf{G}_2 \circ \mathbf{G}_1^T\mathbf{Y}\mathbf{G}_2)\Big) \\ &\quad - \operatorname{tr}\Big((\mathbf{G}_1\mathbf{K}_q\mathbf{G}_2^T)^T\mathbf{Y}\Big) + \operatorname{tr}\Big(\mathbf{K}_p^T\mathbf{Y}\Big). \end{aligned}$$

The optimum point of the parabola (Eq. 1) must be one of the three points, 0, 1 or $-\frac{b}{2a}$. Therefore, a straightforward way to obtain the optimal λ^* is to compute the objective values $J_\alpha(\mathbf{X} + \lambda\mathbf{Y})$ for $\lambda \in \{0, 1, -\frac{b}{2a}\}$ and pick the one yielding the largest value. However, it is more efficient to choose λ^* based on the geometry of the parabola. In our implementation, we adopted the following strategy:

$$\lambda^* = \begin{cases} 0, & \text{if } -\frac{b}{2a} \le 0 \text{ and } a \le 0, \\ 1, & \text{if } \Big(-\frac{b}{2a} \le 0 \text{ and } a > 0\Big) \\ & \text{or } \Big(-\frac{b}{2a} > 0 \text{ and } a > 0\Big) \\ & \text{or } \Big(-\frac{b}{2a} > 1 \text{ and } a < 0\Big), \\ -\frac{b}{2a}, & \text{otherwise.} \end{cases}$$

2. The modified Frank-Wolfe algorithm

In the submitted paper, we adopted a modified Frank-Wolfe algorithm (MFW) presented in [2] to improve the convergence rate of the original FW algorithm. The basic idea of MFW is that the direction computed by FW might not be optimal. To improve that, we consider an alternative direction which is a convex combination of several previously computed directions. Note that this new direction is still feasible due to the convex combination. More specifically, suppose that at step t we compute the optimum $\mathbf{Y}_t$, and meanwhile, we keep a record of the m previously computed directions, $\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \cdots, \mathbf{Y}_{t-m}$. Then the optimum $\mathbf{Y}^*$ would be either $\mathbf{Y}_t$ or $\hat{\mathbf{Y}} \doteq \frac{1}{m}\sum_{i=1}^{m} \mathbf{Y}_{t-i}$ based on the following criteria,

$$\mathbf{Y}^* = \begin{cases} \mathbf{Y}_t, & \text{if } \frac{\operatorname{tr}(\nabla J_\alpha(\mathbf{X})^T \mathbf{Y}_t)}{\|\mathbf{Y}_t\|_F} \geq \frac{\operatorname{tr}(\nabla J_\alpha(\mathbf{X})^T \hat{\mathbf{Y}})}{\|\hat{\mathbf{Y}}\|_F}, \\ \hat{\mathbf{Y}}, & \text{otherwise.} \end{cases}$$

In the experiments, we set $m = 10$ which consistently led to a faster convergence rate.

3. The concave-convex procedure

In the submitted paper, we adopted the concave-convex procedure (CCCP) presented in [4] to improve the optimization performance. The main intuition of using CCCP is to take the advantage of the concave-convex structure of the objective, $J_\alpha(\mathbf{X}) = (1-\alpha)J_{vex}(\mathbf{X}) + \alpha J_{cav}(\mathbf{X})$. In our case, CCCP approximates the non-convex $J_\alpha(\mathbf{X})$ by optimizing a series of convex sub-problems:

$$\begin{aligned} \max_{\mathbf{X}} \quad & J_\alpha(\mathbf{X}) = (1-\alpha)J_{vex}(\mathbf{X}) + \alpha \operatorname{tr}(\nabla J_{cav}(\mathbf{X}_0)^T \mathbf{X}) \quad (2) \\ \text{s. t.} \quad & \mathbf{X}\mathbf{1}_{n_2} \leq \mathbf{1}_{n_1}, \mathbf{X}^T \mathbf{1}_{n_1} \leq \mathbf{1}_{n_2}, \mathbf{X} \geq \mathbf{0}_{n_1 \times n_2} \end{aligned}$$

where Eq. 2 is obtained by the 1^{st}-order Taylor expansion around the previous solution $\mathbf{X}_0$. We iteratively optimize each sub-problem of Eq. 2 via the MFW.

4. The local vs. global strategy

In the submitted paper, we adopted a heuristic strategy (*i.e.*, step 9 of Algorithm 1) to improve the optimization performance of the path-following strategy. The basic idea is to compute $\mathbf{X}^*$ by optimizing $J_{gm}(\mathbf{X})$ instead of $J_\alpha(\mathbf{X})$ in the case when MFW or CCCP does not improve $J_{gm}(\mathbf{X})$. This step is similar to [3], in which Leordeanu *et al*. adopted FW to optimize $J_{gm}(\mathbf{X})$ over $\mathbf{X}$. We take the same step as [3] of using FW except for the computation of the gradient. This is because we never computed the costly $\mathbf{K}$. In a nutshell, we compute the gradient as follows:

$$\nabla J_{gm}(\mathbf{X}) = 2\mathbf{H}_1(\mathbf{H}_1^T \mathbf{X} \mathbf{H}_2 \circ \mathbf{L})\mathbf{H}_2^T$$

which is equivalent to the one proposed in [3]:

$$\nabla \mathbf{x}^T \mathbf{K} \mathbf{x} = 2\mathbf{K}\mathbf{x}$$

References

[1] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, 1956. 1

[2] M. Fukushima. A modified Frank-Wolfe algorithm for solving the traffic assignment problem. *Transp. Res. Part B: Methodological*, 18(2):169–177, 1984. 1

[3] M. Leordeanu, M. Hebert, and R. Sukthankar. An integer projected fixed point method for graph matching and MAP inference. In *NIPS*, 2009. 2

[4] A. L. Yuille and A. Rangarajan. The concave-convex procedure. *Neural Comput.*, 15(4):915–936, 2003. 2