

Question: In Theorem 1, we generally only get stationary points of (3) rather than local minimum. So it is not clear whether such a point is also a stationary point of (2). And by adding $\gamma \|X - Y\|^2$, will more spurious local minima be introduced (i.e. local but not globally optimal), or more saddle points?

Answer: Thank you very much for your valuable comments. We supplement the following Corollary 1 that affirms: when the condition of Theorem 1 is satisfied, a stationary point (\bar{X}, \bar{Y}) of (3) provides \bar{X} as a stationary point of (2), a local minimizer (\bar{X}, \bar{Y}) of (3) provides \bar{X} as a local minimizer of (2). So, no more spurious local minima or saddle points are introduced.

Corollary 1. *When the condition of Theorem 1 is satisfied, we have: (i) \bar{X} is a stationary point of (2) if (\bar{X}, \bar{Y}) is a stationary point of (3); (ii) \bar{X} is a local minimizer of (2) if (\bar{X}, \bar{Y}) is a local minimizer of (3).*

Proof. (i) From $f_s(X, Y) = \frac{1}{2}[\tilde{f}(XY^\top) + \tilde{f}(YX^\top)]$, we have the Fréchet derivative of f_s in term of X as

$$\nabla_X f_s(X, Y) = \frac{1}{2}[\nabla_Z \tilde{f}(Z)\big|_{Z=XY^\top} + \nabla_Z \tilde{f}(Z)^\top\big|_{Z=YX^\top}]Y. \quad (26)$$

The Fréchet derivative of \tilde{f} in problem (2) is

$$\nabla_X \tilde{f}(XX^\top) = [\nabla_Z \tilde{f}(Z) + \nabla_Z \tilde{f}(Z)^\top]\big|_{Z=XX^\top} X. \quad (27)$$

From that (\bar{X}, \bar{Y}) is a stationary point of $F(X, Y; \gamma)$ and $\bar{X} = \bar{Y}$, we have

$$\begin{aligned} \nabla_X F(\bar{X}, \bar{Y}; \gamma)\big|_{\bar{Y}=\bar{X}} &= 0 \\ \Rightarrow [\nabla_X f_s(\bar{X}, \bar{Y}) + \gamma(\bar{X} - \bar{Y})]\big|_{\bar{Y}=\bar{X}} &= 0 \\ \Rightarrow \nabla_X f_s(\bar{X}, \bar{X}) &= 0 \\ \Rightarrow \frac{1}{2}[\nabla_Z \tilde{f}(\bar{Z})\big|_{\bar{Z}=\bar{X}\bar{X}^\top} + \nabla_Z \tilde{f}(\bar{Z})^\top\big|_{\bar{Z}=\bar{X}\bar{X}^\top}]\bar{X} &= 0 \\ \Rightarrow \frac{1}{2}[\nabla_Z \tilde{f}(\bar{Z}) + \nabla_Z \tilde{f}(\bar{Z})^\top]\big|_{\bar{Z}=\bar{X}\bar{X}^\top} \bar{X} &= 0 \\ \Rightarrow \frac{1}{2}\nabla_X \tilde{f}(\bar{X}\bar{X}^\top) &= 0 \\ \Rightarrow \nabla_X \tilde{f}(\bar{X}\bar{X}^\top) &= 0, \end{aligned}$$

where the fourth equality uses (26) and the sixth equality uses (27). The last equality implies that \bar{X} is a stationary point of problem (2).

(ii) If (\bar{X}, \bar{Y}) is a local minimizer of $F(X, Y; \gamma)$, we have

$$F(\bar{X}, \bar{Y}; \gamma) \leq F(\bar{X} + \Theta, \bar{Y} + \Theta; \gamma), \forall \Theta \in \{\Theta \in \mathbb{R}^{n \times r}, \|\Theta\| \rightarrow 0\}. \quad (28)$$

When the condition of Theorem 1 is satisfied, we have $\bar{X} = \bar{Y}$, so that

$$F(\bar{X}, \bar{Y}; \gamma) = f_s(\bar{X}, \bar{Y}) = \tilde{f}(\bar{X}\bar{X}^\top), F(\bar{X} + \Theta, \bar{Y} + \Theta; \gamma) = \tilde{f}((\bar{X} + \Theta)(\bar{X} + \Theta)^\top). \quad (29)$$

Combining (28) and (29), we have

$$G(\bar{X}) \stackrel{\text{def}}{=} \tilde{f}(\bar{X}\bar{X}^\top) \leq \tilde{f}((\bar{X} + \Theta)(\bar{X} + \Theta)^\top) = G(\bar{X} + \Theta).$$

The last inequality means that \bar{X} is a local minimizer of $G(X) = \tilde{f}(XX^\top)$. □

Question: How does the method compare with existing methods that use $Z = XX^\top$ in terms of finding globally minima and local minima?

Answer: A local minimizer \bar{X} of (2) will solve (1) if \bar{X} is rank deficient [Li and Tang, 2016]. Hence, if we set $r(\leq n)$ big enough, a local minima of problem (3) is a local minima of problem (2) (by Corollary 1), and thus it is globally minima of problem (1). Therefore we can conclude that, if the initial rank r is big enough and the condition of Theorem 1 is satisfied, any local minima of (3) is globally minima of (3), (2) and (1).

Question: “Bring higher-order nonlinearity, resulting in optimization difficulties” does sound quantitatively convincing.

Answer: Thank you very much for your valuable comments. We have deleted this improper sentence.

[Li and Tang, 2016] Qiuwei Li and Gongguo Tang, The nonconvex geometry of low-rank matrix optimizations with general objective functions. CoRR, abs/1611.03060, 2016.