Supplementary Material for "Dynamic Principal Component Analysis in High Dimensions"

S.1 Algorithm

Practitioners may use the retraction-based proximal gradient method (ManPG) (Chen et al., 2020) to solve our manifold optimization problem (5). Denote $\mathcal{M} = \mathcal{V}_{p,d}$ and $F(V) = -\mathrm{Tr}\{V(t)^T \hat{\Sigma}(t) V(t)\} + \lambda \|V(t)\|_1$ where $f(V) = -\mathrm{Tr}\{V(t)^T \hat{\Sigma}(t) V(t)\}$ is smooth and its gradient is Lipschitz continuous with the Lipschitz constant L and $h(V) = \lambda\|V(t)\|_1$. ManPG first computes a descent direction D_k (k-th step) by solving the following problem:

$$\begin{aligned} \min_D \quad & < \nabla f(V_k), D > + \frac{1}{2t}\|D\|_F^2 + h(V_k + D) \\ \mathrm{s.t.} \quad & D^T V_k + V_k^T D = 0, \end{aligned} \tag{S.1}$$

where V_k is obtained in the k-th iteration, $t > 0$ is a step size and D is a descent direction of F in the tangent space $T_{V_k}\mathcal{M}$. Based on the Lagrangian function and KKT system, we get that

$$E(\Lambda) = \mathcal{A}_k(D(\Lambda)) = 0, \tag{S.2}$$

where $\mathcal{A}_k(D) = D^T V_k + V_k^T D$, $D(\Lambda) = \mathrm{prox}_{\lambda t h}(B(\Lambda)) - V_k$ with $B(\Lambda) = V_k - t(\nabla f(V_k) - \mathcal{A}^*(\Lambda))$, $\mathcal{A}^*(\Lambda))$ denotes the adjoint operator of $\mathcal{A}_k$, where Λ is a $d \times d$ symmetric matrix. The semi-smooth Newton method (SSN) (Xiao et al., 2018) could be used to solve (S.2).

Retraction operation is an important concept in manifold optimization, see Absil et al. (2009) for more details. There are many common retractions for the Stiefel manifold, including exponential mapping, the polar decomposition and the Cayley transformation. For

inequality and the Cauchy-Schwarz inequality we have

$$
\begin{aligned}
Ee^{a\sum_{l=1}^{m_i}\tilde{w}_{il}x_{ijl}x_{ikl}} &\le 1+\sum_{r=1}^{\infty}\frac{a^r}{r!}E\left(\sum_{l=1}^{m_i}\tilde{w}_{il}x_{ijl}x_{ikl}\right)^r \\
&\le 1+\sum_{r=1}^{\infty}\frac{a^r m_i^{r-1}}{r!}\sum_{l=1}^{m_i}E|\tilde{w}_{il}x_{ijl}x_{ikl}|^r \\
&\le 1+\sum_{r=1}^{\infty}\frac{a^r m_i^{r-1}}{r!}\sum_{l=1}^{m_i}\{E(|\tilde{w}_{il}|^{1/2}x_{ijl})^{2r}+E(|\tilde{w}_{il}|^{1/2}x_{ikl})^{2r}\} \\
&\le \frac{1}{m_i}\sum_{l=1}^{m_i}\left(1+\sum_{r=1}^{\infty}\frac{(am_i)^r}{r!}E(|\tilde{w}_{il}|x_{ijl}^2)^r+\sum_{r=1}^{\infty}\frac{(am_i)^r}{r!}E(|\tilde{w}_{il}|x_{ikl}^2)^r\right) \\
&\le \frac{1}{m_i}\sum_{l=1}^{m_i}\left(Ee^{am_i|\tilde{w}_{il}|x_{ijl}^2}+Ee^{am_i|\tilde{w}_{il}|x_{ikl}^2}\right).
\end{aligned}
$$

Denote the event $E_n=\{E_{\mathrm{x}}e^{a\sum_{l=1}^{m_i}\tilde{w}_{il}x_{ijl}x_{ikl}}<\infty\}$, where E_{x} means that the expectation is taken on x conditional on t_{il}. Then, it holds for all i by picking some appropriate $\tilde{w}_{il}$ and a such that $\max_i am_i|\tilde{w}_{il}|$ is sufficiently small, since $x_j^2(t)$ is sub-exponential uniformly in t by Assumption 2.

Define $B:=\sum_{i=1}^{n}b_{ijk}=O(n^3m^3h^3+n^3m^4h^4)$. For sufficiently large n, we have

$$
\begin{aligned}
P\left(\sum_{i=1}^{n}W_{ijk}\ge\delta_n\,\Big|\,E_n\right) &\le \exp\{-a\delta_n\}E\left\{\exp\left(a\sum_{i=1}^{n}W_{ijk}\right)\Big|\,E_n\right\} \\
&= \exp\{-a\delta_n\}\prod_{i=1}^{n}E\left(\exp(aW_{ijk})\,|\,E_n\right) \\
&\le \exp\{-a\delta_n+Ba^2\}. \qquad \text{(S.6)}
\end{aligned}
$$

Note that (S.6) is minimized when $a=\delta_n/(2B)$ and that the minimizer is $\exp\{-\delta_n^2/(4B)\}$. Thus, there exists some positive constant C such that

$$
P\left(\sum_{i=1}^{n}W_{ijk}\ge\delta_n\,\Big|\,E_n\right)\le\exp\left(-C\delta_n^2/(n^3m^3h^3+n^3m^4h^4)\right).
$$

Similarly, we obtain

$$
P\left(\sum_{i=1}^{n}W_{ijk}\le-\delta_n\,\Big|\,E_n\right)\le\exp\left(-C\delta_n^2/(n^3m^3h^3+n^3m^4h^4)\right).
$$

The following obtained by a simple union bound holds for each $t \in \mathcal{T}$,

$$
\begin{aligned}
& P\left(\max_{j,k} \left| \sum_{i=1}^{n} \sum_{l=1}^{m_i} \{ \bar{w}_{il} x_{ijl} x_{ikl} - E(\bar{w}_{il} x_{ijl} x_{ikl}) \} \right| \geq {}_n \,\middle|\, E_n \right) \\
\leq\; & 2p^2 \exp\left(-C {}_n^2 / (n^3 m^3 h^3 + n^3 m^4 h^4) \right).
\end{aligned}
$$

Let ${}_n = O\{(\log p)^{1/2}(n^3m^3h^3 + n^3m^4h^4)^{1/2}\}$. Note that $\bar{w}_{il} = O_p\{(n^2m^2h^3)^{1/2}\}$ from Lemma 4, then with probability tending to 1, the event E_n holds from Assumption 4. Consequently,

$$
\max_{j,k} \left| \sum_{i=1}^{n} \sum_{l=1}^{m_i} \{ \bar{w}_{il} x_{ijl} x_{ikl} - E(\bar{w}_{il} x_{ijl} x_{ikl}) \} \right| = O_p\{(\log p)^{1/2}(n^3m^3h^3 + n^3m^4h^4)^{1/2}\}.
$$

□

References

Absil, P.-A., Mahony, R., and Sepulchre, R. (2009), *Optimization algorithms on matrix manifolds*, Princeton University Press.

Chen, S., Ma, S., Man-Cho So, A., and Zhang, T. (2020), "Proximal gradient method for nonsmooth optimization over the Stiefel manifold," *SIAM Journal on Optimization*, 30, 210–239.

Edelman, A., Arias, T. A., and Smith, S. T. (1998), "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, 20, 303–353.

Vu, V. Q. and Lei, J. (2013), "Minimax sparse principal subspace estimation in high dimensions," *The Annals of Statistics*, 41, 2905–2947.

Wainwright, M. J. (2019), *High-dimensional statistics: A non-asymptotic viewpoint*, Cambridge University Press.

Xiao, X., Li, Y., Wen, Z., and Zhang, L. (2018), \A regularized semi-smooth Newton method with projection steps for composite convex programs," *Journal of Scienti c Computing*, 1{26.