Online Robust PCA via Stochastic Optimization Supplementary Material

Jiashi Feng
ECE Department
National University of Singapore
jiashi@nus.edu.sg

Huan Xu
ME Department
National University of Singapore
mpexuh@nus.edu.sg

Shuicheng Yan
ECE Department
National University of Singapore
eleyans@nus.edu.sg

1 Algorithm solving Problem (7)

For the data projection $\mathbf{r}$ and noise estimation $\mathbf{e}$, we can get the closed-form solutions for them respectively, as shown in Algorithm 1.1. In particular, the closed-form solution to a projection to ℓ_1-ball in updating $\mathbf{e}$ involves a soft thresholding operator $\mathcal{S}_\lambda[\cdot]$ Hale et al. (2008), which is defined as:

$$\mathcal{S}_\lambda[x] \triangleq \begin{cases} x-\lambda, & \text{if } x > \lambda, \\ x+\lambda, & \text{if } x < -\lambda, \\ 0, & \text{otherwise.} \end{cases}$$

And it is conducted element-wisely on the involved vectors. The optimization iteration is terminated when the following convergence criterion is met:

$$\max(\|\mathbf{r}_{k+1}-\mathbf{r}_k\|/\|\mathbf{z}\|, \|\mathbf{e}_{k+1}-\mathbf{e}_k\|/\|\mathbf{z}\|) < \varepsilon.$$

Here ε is set as 1×10^{-6} throughout the simulations.

The details of the algorithm are summarized as follows,

Algorithm 1.1 Data Projection

Input: $L = [\mathbf{l}_1, \ldots, \mathbf{l}_r] \in \mathbb{R}^{p\times r}$ (input basis), $\mathbf{z} \in \mathbb{R}^p$, parameters λ_1 and λ_2.
$\mathbf{e} \leftarrow \mathbf{0}$.
while not converged **do**
Update the coefficient $\mathbf{r}$:

$$\mathbf{r} \leftarrow (L^T L + \lambda_1 I)^{-1} L^T (\mathbf{z}-\mathbf{e}).$$

Update the sparse error $\mathbf{e}$:

$$\mathbf{e} \leftarrow \mathcal{S}_{\lambda_2}[\mathbf{z}-L\mathbf{r}].$$

end while
Return L.

2 Proof of Lemma 1

Lemma 1 [optimality condition of Problem (4)] $\mathbf{r}^* \in \mathrm{R}^r$ and $\mathbf{e}^* \in \mathrm{R}^p$ is a solution of Problem (4) if and only if

$$\begin{aligned} &C_\Lambda(\mathbf{z}_\Lambda - \mathbf{e}^*_\Lambda) = \lambda_2 \mathrm{sign}(\mathbf{e}^*_\Lambda), \\ &|C_{\Lambda^c}(\mathbf{z}_{\Lambda^c} - \mathbf{e}^*_{\Lambda^c})| \le \lambda_2, \text{ otherwise}, \\ &\mathbf{r}^* = (L^T L + \lambda_1 I)^{-1} L^T (\mathbf{z} - \mathbf{e}^*), \end{aligned}$$

where $C = I - L(L^T L + \lambda_1 I)^{-1} L^T$ and C_Λ denotes the column vectors of matrix C indexed by the set $\Lambda = \{j | \mathbf{e}^*[j] \neq 0\}$ and Λ^c denotes the complementary set of Λ. Moreover, the optimal solution is unique.

Proof. Denote the subgradient of $\|\mathbf{e}\|_1$ as $\partial\|\mathbf{e}\|_1$ and it is known that

$$\partial\|\mathbf{e}\|_1 = \{\mathbf{u} | \mathbf{u}_i = \mathrm{sign}(\mathbf{e}_i) \text{ if } \mathbf{e}_i \neq 0, \text{ and } |\mathbf{e}_i| \le 1 \text{ otherwise}\}.$$

The point $(\mathbf{r}^*, \mathbf{e}^*)$ is a global minimum of (5) if and only if the vector zero is in its subgradient at $(\mathbf{r}^*, \mathbf{e}^*)$:

$$\exists \mathbf{u} \in \partial\|\mathbf{e}^*\|_1 \text{ such that } \mathbf{e}^* + L\mathbf{r}^* - \mathbf{z} + \lambda_2 \mathbf{u} = 0, \tag{2.1}$$

$$-L^T \mathbf{z} + L^T L \mathbf{r}^* + L^T \mathbf{e}^* + \lambda_1 \mathbf{r}^* = 0. \tag{2.2}$$

From (2.2), we have,

$$\mathbf{r}^* = (L^T L + \lambda_1 I)^{-1} L^T (\mathbf{z} - \mathbf{e}^*).$$

This proves the third inequality in the lemma. Substituting back into (2.1) yields

$$I - L(L^T L + \lambda_1 I)^{-1} L^T \ (\mathbf{z} - \mathbf{e}^*) = \lambda_2 \mathbf{u}, \text{ where } \mathbf{u} \in \partial\|\mathbf{e}^*\|_1.$$

Define the matrix

$$C \triangleq I - L(L^T L + \lambda_1 I)^{-1} L^T.$$

According to Woodbury matrix identity, we have

$$C = \left(I + \frac{1}{\lambda_1} L^T L\right)^{-1}.$$

Thus C is invertible. We then have

$$C(\mathbf{z} - \mathbf{e}^*) = \lambda_2 \mathbf{u}, \text{ where } \mathbf{u} \in \partial\|\mathbf{e}^*\|_1.$$

Let $\Lambda = \{j | \mathbf{e}^*[j] \neq 0\}$ be the index set of nonzero elements of the optimal solution $\mathbf{e}^*$. Then we can show that

$$C_\Lambda(\mathbf{z}_\Lambda - \mathbf{e}^*_\Lambda) = \lambda_2 \mathrm{sign}(\mathbf{e}^*_\Lambda)$$

Here C_Λ denotes submatrix of C consisting of the column vectors of matrix C indexed by Λ. Then we can solve out that

$$\begin{aligned} &\mathbf{e}^*_\Lambda = (C_\Lambda^T C_\Lambda)^{-1} (C_\Lambda \mathbf{z}_\Lambda - \lambda \mathrm{sign}(\mathbf{e}^*_\Lambda)), \\ &\mathbf{e}^*_{\Lambda^c} = 0. \end{aligned}$$

Since C is invertible, C is column full rank. Thus C_Λ is column full rank and $C_\Lambda^T C_\Lambda$ is invertible, the solution $\mathbf{e}^*$ is unique and thus $\mathbf{r}^*$ is also unique. □

3 Proof of Lemma 2

Lemma 2 Assume the observations $\mathbf{z}$ are always bounded. Define

$$\{\mathbf{r}^*, \mathbf{e}^*\} = \arg\min_{\mathbf{r},\mathbf{e}} \frac{1}{2}\|\mathbf{z} - L\mathbf{r} - \mathbf{e}\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{r}\|_2^2 + \lambda_2 \|\mathbf{e}\|_1.$$

Then, 1) the function $\ell(\mathbf{z}, L)$ defined in (4) is continuously differentiable and

$$\nabla_L \ell(\mathbf{z}, L) = (L\mathbf{r}^* + \mathbf{e}^* - \mathbf{z})\mathbf{r}^{*T};$$

2) $\nabla f(L) = \mathrm{E}_{\mathbf{z}}[\nabla_L \ell(\mathbf{z}, L)]$; and 3) $\nabla_L f(L)$ is Lipschitz.

Proof. To reveal the regularity of the expected loss function f and its derivative ∇f, we need first to prove the regularity of the loss function ℓ as stated in the first claim.

Proof of the first claim

Define a function $\tilde{f}$ as

$$\tilde{f}(\mathbf{r}, \mathbf{e}, \mathbf{z}, L) \triangleq \frac{1}{2}\|\mathbf{z} - L\mathbf{r} - \mathbf{e}\|_2^2 + \frac{\lambda_1}{2}\|L\|_F^2 + \frac{\lambda_1}{2}\|\mathbf{r}\|_2^2 + \lambda_2\|\mathbf{e}\|_1.$$

Thus the loss function ℓ can be expressed as

$$\ell(\mathbf{z}, L) = \min_{\mathbf{r},\mathbf{e}} \tilde{f}(\mathbf{r}, \mathbf{e}, \mathbf{z}, L).$$

The function $\tilde{f}(\mathbf{r}, \mathbf{e}, \mathbf{z}, L)$ is continuous, and for all $\mathbf{r} \in \mathrm{R}^r, \mathbf{e} \in \mathrm{R}^p$, the function $\tilde{f}(\mathbf{r}, \mathbf{e}, \cdot, \cdot)$ is differentiable, and the derivative $\nabla_L \tilde{f}(\mathbf{r}, \mathbf{e}, \cdot, \cdot) = (L\mathbf{r} + \mathbf{e} - \mathbf{z})\mathbf{r}^T$ is continuous. Furthermore, according to Lemma 1, $\tilde{f}(\cdot, \cdot, \mathbf{z}, L)$ has unique minimizer $(\mathbf{r}^*, \mathbf{e}^*)$, thus Lemma 3 directly applies and we obtain that $\ell(\mathbf{z}, L)$ is differentiable in L and

$$\nabla_L \ell(\mathbf{z}, L) = \nabla_L \tilde{f}(\mathbf{r}^*, \mathbf{e}^*, \mathbf{z}, L) = (L\mathbf{r}^* + \mathbf{e}^* - \mathbf{z})\mathbf{r}^{*T} + \lambda_1 L.$$

Thus, we complete the proof of the first claim.

Proof of the second claim

According to the first claim, the function $\ell_{\mathbf{z},L}$ is continuously differentiable, thus

$$\nabla_L f(L) = \nabla_L \mathrm{E}_{\mathbf{z}}[\ell(\mathbf{z}, L)] = \mathrm{E}_{\mathbf{z}}[\nabla_L \ell(\mathbf{z}, L)].$$

Equipped with the above two results, we are ready to prove that the derivative $\nabla_L f(L)$ is Lipschitz.

Proof of the third claim

To prove that $\nabla f(L)$ is Lipschitz, we will show that for all bounded observations $\mathbf{z}$, $\mathbf{r}^*(\mathbf{z}, \cdot)$ and $\mathbf{e}^*(\mathbf{z}, \cdot)$ are Lipschitz with constants independent of $\mathbf{z}$. First, the loss function $\ell(\mathbf{z}, L)$ defined in (4) is continuous in $\mathbf{r}, \mathbf{e}, L, \mathbf{z}$ and has a unique minimum (according to Lemma 1) for fixed $\mathbf{z}$ and L, thus the optimal solutions $\mathbf{r}^*$ and $\mathbf{e}^*$ are continuous in L and $\mathbf{z}$.

Consider a matrix L and a sample $\mathbf{z}$, and denote $\mathbf{r}^*$ and $\mathbf{e}^*$ as the corresponding optimal solutions. Denote by Λ the set of the indices such that $|(\mathbf{z}_\Lambda - \mathbf{e}^*_\Lambda)| = \lambda_2$ (see Lemma 1). Here the matrix C is defined as $C = I - L(L^T L + \lambda_1 I)^{-1} L^T$. Since C_Λ is nonsingular, $(\mathbf{z}_\Lambda - \mathbf{e}^*_\Lambda)$ is continuous in L and $\mathbf{z}$. Thus we consider a small perturbation of $(\mathbf{z}, L)$ in one of their open neighborhood V, such that for all $(\mathbf{z}^0, L^0)$ in V, we have if $j \notin \Lambda$, $|(\mathbf{z} - L\mathbf{r}^*)[j] - (\mathbf{e}^{*0})[j]| < \lambda_2$ and $\mathbf{e}^{*0}[j] = 0$, where $\mathbf{e}^{*0} = \mathbf{e}^*(\mathbf{z}^0, L^0)$. Namely the support set of $\mathbf{e}^*$ is not changed.

Based on the about continuity, we consider the following function

$$\tilde{\ell}(\mathbf{z}_\Lambda, L_\Lambda, \mathbf{r}, \mathbf{e}_\Lambda) \triangleq \frac{1}{2}\|\mathbf{z}_\Lambda - L_\Lambda\mathbf{r} - \mathbf{e}_\Lambda\|_2^2 + \frac{\lambda_1}{2}\|L_\Lambda\|_F^2 + \frac{\lambda_1}{2}\|\mathbf{r}\|_2^2 + \lambda_2\|\mathbf{e}_\Lambda\|_1.$$

Since the Hessian matrix of the function $\tilde{\ell}(\mathbf{z}_\Lambda, L_\Lambda, \cdot, \cdot)$ w.r.t. $\mathbf{r}$, $I \otimes (L_\Lambda^T L_\Lambda + \lambda_1 I)$, and the Hessian matrix w.r.t. $\mathbf{e}_\Lambda$, $I \otimes \lambda_2 I$, are positive definite, we have the function $\tilde{\ell}(\mathbf{z}_\Lambda, L_\Lambda, \cdot, \cdot)$ is strictly convex and

$$\begin{aligned} & \tilde{\ell}(\mathbf{z}_\Lambda, L_\Lambda, \mathbf{r}^{*0}, \mathbf{e}^{*0}_\Lambda) - \tilde{\ell}(\mathbf{z}_\Lambda, L_\Lambda, \mathbf{r}^*, \mathbf{e}^*_\Lambda) \\ \geq\ & \lambda_1\|\mathbf{r}^{*0} - \mathbf{r}^*\|_2^2 + \lambda_2\|\mathbf{e}^{*0}_\Lambda - \mathbf{e}^*_\Lambda\|_2^2 \\ \geq\ & \min(\lambda_1, \lambda_2)(\|\mathbf{r}^{*0} - \mathbf{r}^*\|_2^2 + \|\mathbf{e}^{*0}_\Lambda - \mathbf{e}^*_\Lambda\|_2^2). \end{aligned} \tag{3.1}$$

Moreover, it is easy to show that $\tilde{\ell}(\mathbf{z}_\Lambda, L_\Lambda, \cdot, \cdot)$ is Lipschitz with constant c_2 independent of $\mathbf{z}$.

To see this, consider the following difference of the function $\tilde{\ell}(\mathbf{z}_\Lambda, L_\Lambda, \cdot, \cdot)$,

$$\begin{aligned}
& \tilde{\ell}(\mathbf{z}_\Lambda, L_\Lambda, \mathbf{r}', \mathbf{e}'_\Lambda) - \tilde{\ell}(\mathbf{z}_\Lambda, L_\Lambda, \mathbf{r}'', \mathbf{e}''_\Lambda) \\
\leq\ & \frac{1}{2}\|L_\Lambda(\mathbf{r}' + \mathbf{r}'')\| + \|(\mathbf{e}'_\Lambda + \mathbf{e}''_\Lambda)\| + 2\|\mathbf{z}_\Lambda\|_\infty \|L_\Lambda(\mathbf{r}' - \mathbf{r}'')\| + \|(\mathbf{e}'_\Lambda - \mathbf{e}''_\Lambda)\| \\
& + \frac{\lambda_1}{2}\|\mathbf{r}' + \mathbf{r}''\|_\infty \|\mathbf{r}' - \mathbf{r}''\| + \lambda_2 \|\mathbf{e}'_\Lambda - \mathbf{e}''_\Lambda\| \\
\leq\ & (c_1\|L_\Lambda\| + c_2)\|\mathbf{r}' - \mathbf{r}''\| + (c_1 + \lambda_2)\|\mathbf{e}'_\Lambda - \mathbf{e}''_\Lambda\| \\
\leq\ & c_3(\|\mathbf{r}' - \mathbf{r}''\| + \|\mathbf{e}'_\Lambda - \mathbf{e}''_\Lambda\|).
\end{aligned}$$

Here $c_1 = \frac{1}{2}\|L_\Lambda(\mathbf{r}' + \mathbf{r}'')\| + \|(\mathbf{e}'_\Lambda + \mathbf{e}''_\Lambda)\| + 2\|\mathbf{z}_\Lambda\|_\infty$, $c_2 = \frac{\lambda_1}{2}\|\mathbf{r}' + \mathbf{r}''\|_\infty$ and $c_3 = \max(c_1\|L_\Lambda\| + c_2, c_1 + \lambda_2)$. Since the observations are from a bounded set, the constants c_1 and c_2 are upper bounded and the upper bound is independent of the sample $\mathbf{z}$.

Combining with (3.1), we can show that

$$\|\mathbf{r}^{t+1} - \mathbf{r}^t\|_2 + \|\mathbf{e}^{t+1} - \mathbf{e}^t\|_2 = \|\mathbf{r}^{t+1} - \mathbf{r}^t\|_2 + \|\mathbf{e}''_\Lambda - \mathbf{e}'_\Lambda\|_2 \leq \frac{c_3}{\min(\lambda_1, \lambda_2)}.$$

Therefore, $\mathbf{r}^t$ and $\mathbf{e}^t$ are locally Lipschitz. Since they are in a compact set, $\mathbf{r}^t$ and $\mathbf{e}^t$ are uniformly Lipschitz with constant independent of the sample $\mathbf{z}$.

Thus, combining with the second claim, we have that the function $\nabla_L f$ is Lipschitz continuous with the same constant as the function ℓ. This gives the third claim. □

4 Proof of Theorem 2

Theorem 2 [Convergence of the surrogate function g_t] Let g_t denote the surrogate function defined in (6). Then, $g_t(L_t)$ converges almost surely when the solution L_t is given by Algorithm 1.

Proof. We prove the convergence of the sequence $g_t(L_t)$ by showing that the stochastic positive process

$$u_t \triangleq g_t(L_t) \geq 0,$$

is a quasi-martingale Fisk (1965). According to Lemma 4, if the sum of the positive difference of u_t is bounded, u_t is a quasi-martingale. And the sum converges almost surely. Thus, we compute the difference of u_t and obtain

$$\begin{aligned}
& u_{t+1} - u_t \\
=\ & g_{t+1}(L_{t+1}) - g_t(L_t) \\
=\ & g_{t+1}(L_{t+1}) - g_{t+1}(L_t) + g_{t+1}(L_t) - g_t(L_t) \\
=\ & g_{t+1}(L_{t+1}) - g_{t+1}(L_t) + \frac{\ell(\mathbf{z}_{t+1}, L_t) - f_t(L_t)}{t+1} + \frac{f_t(L_t) - g_t(L_t)}{t+1}.
\end{aligned} \tag{4.1}$$

Here the third equality is from the fact that $g_{t+1}(L_t) = \frac{1}{t+1}\ell(\mathbf{z}_{t+1}, L_t) + \frac{t}{t+1}g_t(L_t)$. Since L_{t+1} minimizes g_{t+1}, $g_{t+1}(L_{t+1}) - g_{t+1}(L_t) \leq 0$. Since the surrogate g_t upperbounds the empirical cost f_t, $g_t \geq f_t$, we also have $f_t(L_t) - g_t(L_t) \leq 0$. Thus we have

$$u_{t+1} - u_t \leq \frac{\ell(\mathbf{z}_{t+1}, L_t) - f_t(L_t)}{t+1}.$$

Since the above inequality is valid for the variation of each pair of u_{t+1} and u_t, in particular it will be valid for the positive variation,

$$[u_{t+1} - u_t]^+ \leq \frac{\ell(\mathbf{z}_{t+1}, L_t) - f_t(L_t)}{t+1},$$

where $[\cdot]^+$ denotes the positive variation.

According to Lemma 4, we consider the expectation of the variation of u_t conditioned on the past information $\mathcal{F}_t = \{\mathbf{z}_1, \ldots, \mathbf{z}_t, L_1, \ldots, L_t, \mathbf{r}_1, \ldots, \mathbf{r}_t, \mathbf{e}_1, \ldots, \mathbf{e}_t\}$ and apply the above inequality,

$$\mathbb{E}[[u_{t+1} - u_t | \mathcal{F}_t]^+] \leq \frac{\mathbb{E}[\ell(\mathbf{z}_{t+1}, L_t)|\mathcal{F}_t] - f_t(L_t)}{t+1} = \frac{f(L_t) - f_t(L_t)}{t+1} \leq \frac{\|f - f_t\|_\infty}{t+1}. \tag{4.2}$$

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：https://d.book118.com/877040125056006035