Supplementary Material for Linearized Alternating Direction Method with Adaptive Penalty for Low-Rank Representation

Zhouchen Lin
Visual Computing Group
Microsoft Research Asia

Risheng Liu   **Zhixun Su**
School of Mathematical Sciences
Dalian University of Technology

1   Proof of Proposition 2

We present the proof of Proposition 2 in a more general setting. Namely, λ is updated as

$$\lambda_{k+1} = \lambda_k + \gamma\beta_k[\mathcal{A}(\mathbf{x}_{k+1}) + \mathcal{B}(\mathbf{y}_{k+1}) - \mathbf{c}]. \tag{21}$$

Even with this extra parameter γ, the proof of Theorem 3 is almost unchanged. We have a more general Proposition 2 as follows:

Proposition 2 *If $\{\beta_k\}$ is non-decreasing and upper bounded, $\eta_A > \|\mathcal{A}\|^2$, $\gamma \in (0,2)$, $\eta_B(2-\gamma) > \|\mathcal{B}\|^2$, and $(\mathbf{x}^*, \mathbf{y}^*, \lambda^*)$ is any KKT point of problem (1), then:*

1. *$\{\eta_A\|\mathbf{x}_k - \mathbf{x}^*\|^2 - \|\mathcal{A}(\mathbf{x}_k - \mathbf{x}^*)\|^2 + \eta_B\|\mathbf{y}_k - \mathbf{y}^*\|^2 + \gamma^{-1}\beta_k^{-2}\|\lambda_k - \lambda^*\|^2\}$ is non-increasing.*
2. *$\|\mathbf{x}_{k+1} - \mathbf{x}_k\| \to 0$, $\|\mathbf{y}_{k+1} - \mathbf{y}_k\| \to 0$, $\|\lambda_{k+1} - \lambda_k\| \to 0$.*

The proof of Proposition 2 is based on the following lemma.

Lemma 1

$$
\begin{aligned}
&\eta_A\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 - \|\mathcal{A}(\mathbf{x}_{k+1} - \mathbf{x}^*)\|^2 + \eta_B\|\mathbf{y}_{k+1} - \mathbf{y}^*\|^2 + \gamma^{-1}\beta_k^{-2}\|\lambda_{k+1} - \lambda^*\|^2 \\
&= \eta_A\|\mathbf{x}_k - \mathbf{x}^*\|^2 - \|\mathcal{A}(\mathbf{x}_k - \mathbf{x}^*)\|^2 + \eta_B\|\mathbf{y}_k - \mathbf{y}^*\|^2 + \gamma^{-1}\beta_k^{-2}\|\lambda_k - \lambda^*\|^2 \\
&\quad - \left\{ (2-\gamma)(\gamma\beta_k)^{-2}\|\lambda_{k+1} - \lambda_k\|^2 + \eta_B\|\mathbf{y}_{k+1} - \mathbf{y}_k\|^2 \right. \\
&\qquad \left. -2(\gamma\beta_k)^{-1}\langle\lambda_{k+1} - \lambda_k, \mathcal{B}(\mathbf{y}_{k+1} - \mathbf{y}_k)\rangle \right\} \\
&\quad -(\eta_A\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 - \|\mathcal{A}(\mathbf{x}_{k+1} - \mathbf{x}_k)\|^2) \\
&\quad -2\beta_k^{-1}\left\langle \mathbf{x}_{k+1} - \mathbf{x}^*, [-\beta_k\eta_A(\mathbf{x}_{k+1} - \mathbf{x}_k) - \mathcal{A}^*(\tilde{\lambda}_{k+1})] + \mathcal{A}^*(\lambda^*)\right\rangle \\
&\quad -2\beta_k^{-1}\left\langle \mathbf{y}_{k+1} - \mathbf{y}^*, [-\beta_k\eta_B(\mathbf{y}_{k+1} - \mathbf{y}_k) - \mathcal{B}^*(\hat{\lambda}_{k+1})] + \mathcal{B}^*(\lambda^*)\right\rangle.
\end{aligned}
\tag{22}
$$

This identity can be routinely checked, by using the definitions of $\tilde{\lambda}_{k+1}$ and $\hat{\lambda}_{k+1}$ and the following facts:

1. $2\langle\mathbf{a}_{k+1} - \mathbf{a}^*, \mathbf{a}_{k+1} - \mathbf{a}_k\rangle = \|\mathbf{a}_{k+1} - \mathbf{a}^*\|^2 - \|\mathbf{a}_k - \mathbf{a}^*\|^2 + \|\mathbf{a}_{k+1} - \mathbf{a}_k\|^2$.
2. $\mathcal{A}(\mathbf{x}^*) + \mathcal{B}(\mathbf{y}^*) = \mathbf{c}$.
3. $\langle\lambda, \mathcal{A}(\mathbf{x})\rangle = \langle\mathcal{A}^*(\lambda), \mathbf{x}\rangle$, $\langle\lambda, \mathcal{B}(\mathbf{y})\rangle = \langle\mathcal{B}^*(\lambda), \mathbf{y}\rangle$.

As it is lengthy and tedious, we omit the complete details.

Proof (of Proposition 2) By Lemma 1 and the given conditions, it is easy to check that

$$\eta_A\|\mathbf{w}\|^2 - \|\mathcal{A}(\mathbf{w})\|^2 \geq 0, \quad \text{for } \mathbf{w} = \mathbf{x}_{k+1} - \mathbf{x}^*, \mathbf{x}_k - \mathbf{x}^*, \mathbf{x}_{k+1} - \mathbf{x_k},$$

1

$$(2-\gamma)(\gamma\beta_k)^{-2}\|\lambda_{k+1}-\lambda_k\|^2+\eta_B\|\mathbf{y}_{k+1}-\mathbf{y}_k\|^2-2(\gamma\beta_k)^{-1}\langle\lambda_{k+1}-\lambda_k,\mathcal{B}(\mathbf{y}_{k+1}-\mathbf{y}_k)\rangle\geq 0.$$

The last two terms in (22) are also nonnegative due to Proposition 1 and the monotonicity of sub-gradient mapping. So Proposition 2 (1) is obvious due to the non-decrement of $\{\beta_k\}$.

Then as $\{\eta_A\|\mathbf{x}_k-\mathbf{x}^*\|^2-\|\mathcal{A}(\mathbf{x}_k-\mathbf{x}^*)\|^2+\eta_B\|\mathbf{x}_k-\mathbf{x}^*\|^2+\gamma^{-1}\beta_k^{-2}\|\lambda_k-\lambda^*\|^2\}$ is non-increasing and non-negative, it has a limit. Then we can see that

$$\eta_A\|\mathbf{x}_{k+1}-\mathbf{x_k}\|^2-\|\mathcal{A}(\mathbf{x}_{k+1}-\mathbf{x_k})\|^2\to 0,$$

$$(2-\gamma)(\gamma\beta_k)^{-2}\|\lambda_{k+1}-\lambda_k\|^2+\eta_B\|\mathbf{y}_{k+1}-\mathbf{y}_k\|^2-2(\gamma\beta_k)^{-1}\langle\lambda_{k+1}-\lambda_k,\mathcal{B}(\mathbf{y}_{k+1}-\mathbf{y}_k)\rangle\to 0,$$

due to their non-negativity. So $\|\mathbf{x}_{k+1}-\mathbf{x}_k\|\to 0$ follows from the first limit.

Note that

$$
\begin{aligned}
&(2-\gamma)(\gamma\beta_k)^{-2}\|\lambda_{k+1}-\lambda_k\|^2+\eta_B\|\mathbf{y}_{k+1}-\mathbf{y}_k\|^2-2(\gamma\beta_k)^{-1}\langle\lambda_{k+1}-\lambda_k,\mathcal{B}(\mathbf{y}_{k+1}-\mathbf{y}_k)\rangle\\
&\geq (2-\gamma)(\gamma\beta_k)^{-2}\|\lambda_{k+1}-\lambda_k\|^2+\eta_B\|\mathbf{y}_{k+1}-\mathbf{y}_k\|^2-2(\gamma\beta_k)^{-1}\|\lambda_{k+1}-\lambda_k\|\|\mathcal{B}(\mathbf{y}_{k+1}-\mathbf{y}_k)\|\\
&= \left((2-\gamma)^{1/2}(\gamma\beta_k)^{-1}\|\lambda_{k+1}-\lambda_k\|-(2-\gamma)^{-1/2}\|\mathcal{B}(\mathbf{y}_{k+1}-\mathbf{y}_k)\|\right)^2\\
&\quad+\eta_B\|\mathbf{y}_{k+1}-\mathbf{y}_k\|^2-(2-\gamma)^{-1}\|\mathcal{B}(\mathbf{y}_{k+1}-\mathbf{y}_k)\|^2\\
&\geq \eta_B\|\mathbf{y}_{k+1}-\mathbf{y}_k\|^2-(2-\gamma)^{-1}\|\mathcal{B}(\mathbf{y}_{k+1}-\mathbf{y}_k)\|^2.
\end{aligned}
$$

So we have that $\|\mathbf{y}_{k+1}-\mathbf{y}_k\|\to 0$. On the other hand,

$$
\begin{aligned}
&(2-\gamma)(\gamma\beta_k)^{-2}\|\lambda_{k+1}-\lambda_k\|^2+\eta_B\|\mathbf{y}_{k+1}-\mathbf{y}_k\|^2-2(\gamma\beta_k)^{-1}\langle\lambda_{k+1}-\lambda_k,\mathcal{B}(\mathbf{y}_{k+1}-\mathbf{y}_k)\rangle\\
&= \left((2-\gamma)^{1/2}(\gamma\beta_k)^{-1}\|\lambda_{k+1}-\lambda_k\|-\sqrt{\eta_B}\|\mathbf{y}_{k+1}-\mathbf{y}_k\|\right)^2\\
&\quad+2(\gamma\beta_k)^{-1}\left(\sqrt{\eta_B(2-\gamma)}\|\lambda_{k+1}-\lambda_k\|\|\mathbf{y}_{k+1}-\mathbf{y}_k\|-\langle\lambda_{k+1}-\lambda_k,\mathcal{B}(\mathbf{y}_{k+1}-\mathbf{y}_k)\rangle\right)\\
&\geq \left((2-\gamma)^{1/2}(\gamma\beta_k)^{-1}\|\lambda_{k+1}-\lambda_k\|-\sqrt{\eta_B}\|\mathbf{y}_{k+1}-\mathbf{y}_k\|\right)^2.
\end{aligned}
$$

So $(2-\gamma)^{1/2}(\gamma\beta_k)^{-1}\|\lambda_{k+1}-\lambda_k\|-\sqrt{\eta_B}\|\mathbf{y}_{k+1}-\mathbf{y}_k\|\to 0$. This together with $\|\mathbf{y}_{k+1}-\mathbf{y}_k\|\to 0$ results in $\|\lambda_{k+1}-\lambda_k\|\to 0$.

2 Solving LRR via APG

The LRR problem can also be relaxed to the following unconstrained optimization problem:

$$\min \beta\|\mathbf{Z}\|_*+\beta\mu\|\mathbf{E}\|_{2,1}+\frac{1}{2}\|\mathbf{X}-\mathbf{XZ}-\mathbf{E}\|^2, \tag{23}$$

where $\beta > 0$ is a relaxation parameter. Then we can apply APG to solve this problem. The two subproblems to update $\mathbf{E}$ and $\mathbf{Z}$ are:

$$\mathbf{E}_{k+1}=\arg\min_{\mathbf{E}}\mu\beta\|\mathbf{E}\|_{2,1}+\frac{\tau}{2}\|\mathbf{E}-(\bar{\mathbf{E}}_k-\frac{1}{2\tau}\nabla_E\|\mathbf{X}-\mathbf{XZ}-\mathbf{E}\|^2|_{\bar{\mathbf{E}}_k,\bar{\mathbf{Z}}_k})\|^2, \tag{24a}$$

$$\mathbf{Z}_{k+1}=\arg\min_{\mathbf{Z}}\beta\|\mathbf{Z}\|_*+\frac{\tau}{2}\|\mathbf{Z}-(\bar{\mathbf{Z}}_k-\frac{1}{2\tau}\nabla_Z\|\mathbf{X}-\mathbf{XZ}-\mathbf{E}\|^2|_{\bar{\mathbf{E}}_k,\bar{\mathbf{Z}}_k})\|^2, \tag{24b}$$

where $\tau\geq\sigma_{\max}^2(\mathbf{X})$ is a Lipschitz constant.

The APG approach, with the continuation technique, for the LRR problem is described in Algorithm 3.

3 Convergence Behaviors of Tested Algorithms

In Figure 1, we plot the relative changes of $\mathbf{E}_k$ and $\mathbf{Z}_k$ and the feasibility errors at all iterations for four test algorithms, respectively. We can see the errors of LADMAP in the two KKT conditions drop much quicker than other methods.

2

Algorithm 3 APG for LRR

Input: Observation matrix $\mathbf{X}$ and parameter $\mu > 0$.
Initialize: Set $\mathbf{E}_0 = \mathbf{E}_{-1} = \mathbf{0}$ and $\mathbf{Z}_0 = \mathbf{Z}_{-1} = \mathbf{0}$.
Set $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $\beta_0 \gg \beta_{\min} > 0$, $t_0 = t_{-1} = 1$, $\theta < 1$, $\tau \geq \sigma_{\max}^2(\mathbf{X})$, and $k \leftarrow 0$.
while not converged **do**

    Step 1: Update $\bar{\mathbf{E}}_k = \mathbf{E}_k + \frac{t_{k-1}-1}{t_k}(\mathbf{E}_k - \mathbf{E}_{k-1})$, $\bar{\mathbf{Z}}_k = \mathbf{Z}_k + \frac{t_{k-1}-1}{t_k}(\mathbf{Z}_k - \mathbf{Z}_{k-1})$.

    Step 2: Update $\mathbf{G}_k^E = \bar{\mathbf{E}}_k + \frac{1}{\tau}(\mathbf{X} - \mathbf{X}\bar{\mathbf{Z}}_k - \bar{\mathbf{E}}_k)$.

    Step 3: Update $\mathbf{E}_{k+1} = \mathcal{S}_{\frac{\mu\beta_k}{\tau}}(\mathbf{G}_k^E)$, where $\mathcal{S}$ is the shrinkage operator.

    Step 4: Update $\mathbf{G}_k^Z = \bar{\mathbf{Z}}_k + \frac{1}{\tau}\mathbf{X}^T(\mathbf{X} - \mathbf{X}\bar{\mathbf{Z}}_k - \bar{\mathbf{E}}_k)$.

    Step 5: Update $\mathbf{Z}_{k+1} = \mathbf{U}\mathcal{S}_{\frac{\beta_k}{\tau}}(\Sigma)\mathbf{V}^T$, where $\mathbf{U}\Sigma\mathbf{V}^T$ is the SVD of $\mathbf{G}_k^Z$.

    Step 6: Update $t_{k+1} = \frac{1+\sqrt{4t_k^2+1}}{2}$, $\beta_{k+1} = \max(\beta_{\min}, \theta\beta_k)$.

    Step 7: Check the convergence conditions:
    $\frac{\|\mathbf{X}\mathbf{Z}_{k+1}+\mathbf{E}_{k+1}-\mathbf{X}\|}{\|\mathbf{X}\|} \leq \varepsilon_1$ and $\max\left(\frac{\|\mathbf{Z}_{k+1}-\mathbf{Z}_k\|}{\|\mathbf{X}\|}, \frac{\|\mathbf{E}_{k+1}-\mathbf{E}_k\|}{\|\mathbf{X}\|}\right) \leq \varepsilon_2$.
    If they are satisfied, break.

    Step 8: $k \leftarrow k + 1$.
end while

(a) Relative Changes of $\mathbf{E}_k$ (b) Relative Changes of $\mathbf{Z}_k$ (c) Feasibility Errors
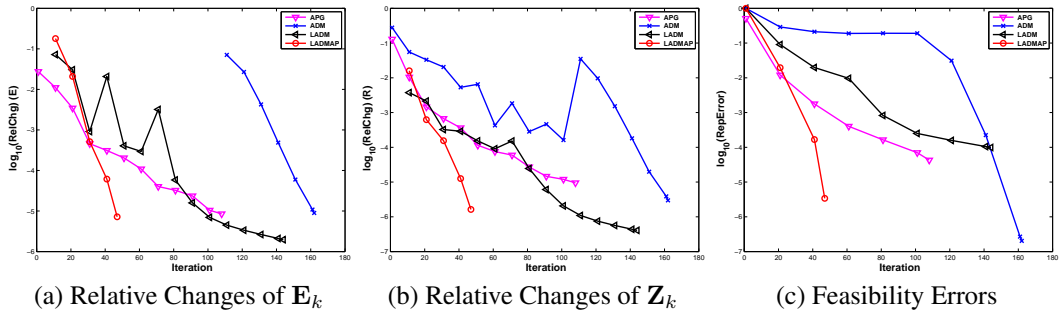
Figure 1: Convergence behaviors of APG, ADM, LADM, LADMAP on the toy data $\mathbf{X}$ generated with parameters (5, 20, 100, 5). The changes and errors are in $\log_{10}$ scale. In (a) and (b), as the relative changes of $\mathbf{E}_k$ and $\mathbf{Z}_k$ in the first several iterations are zeros, which corresponds to $-\infty$ in the plots, we only report the nonzero relative changes of $\mathbf{E}_k$ and $\mathbf{Z}_k$.

3