The proof of the convergence of the ADMM algorithm for EDOHA

The scaled augmented Lagrangian can be rewritten as

\[
L_\rho(X, \tilde{X}, \tilde{V}, U, \tilde{U}_V) = \Phi(X) + h_1(\tilde{V}) + \Psi(\tilde{X}) + U^T(X - \tilde{X}) + \frac{\rho}{2}\|X - \tilde{X}\|_F^2 + \tilde{U}_V^T(\tilde{V} - V) + \frac{\rho}{2}\|\tilde{V} - V\|_F^2
\]

(1)

where \(U = \rho W\) and \(\tilde{U} = \rho \tilde{W}_V\). Let \((X^*, \tilde{X}^*, \tilde{V}^*)\) be a primal optimal and \((U^*, U^*_V)\) be a dual optimal point. According to convex optimization theory, KKT conditions guarantee that the optimal duality gap is zero \([1]\), that is,

\[
L_0(X^*, \tilde{X}^*, \tilde{V}^*, U, \tilde{U}_V) \leq L_0(X_t^*, \tilde{X}^*_t, \tilde{V}^*_t, U_t, \tilde{U}_V^*_t) \leq L_0(X_t+1, \tilde{X}_t+1, \tilde{V}_t+1, U^*_t, \tilde{U}_V^*_t)
\]

holds for all \(X, \tilde{X}, \tilde{V}, U, \tilde{U}_V\). Since \((X^*, \tilde{X}^*, \tilde{V}^*, U^*_V)\) is a saddle point for \(L_0\), we have

\[
L_0(X^*, \tilde{X}^*, \tilde{V}^*, U^*_V) \leq L_0(X_{t+1}, \tilde{X}_{t+1}, \tilde{V}_{t+1}, U^*_t, \tilde{U}_V^*_t).
\]

Using \(X^* = \tilde{X}^*\) and \(V^* = \tilde{V}^*\),

\[
p^* - p_{t+1} \leq U^T(X_{t+1} - \tilde{X}_{t+1}) + \tilde{U}_V^T(\tilde{V}_{t+1} - V_{t+1}).
\]

(2)

By definition, \(X_{t+1}\) minimizes \(L_\rho(X, \tilde{X}_t, \tilde{V}_t, U_t, \tilde{U}_V_t)\). The optimality condition is

\[
0 \in \partial L_\rho(X_{t+1}, \tilde{X}_t, \tilde{V}_t, U_t, \tilde{U}_V_t).
\]

Since it is separable with respect to \((\Theta_{t+1}, Z_{t+1}, V_{t+1})\), we have

\[
0 \in \partial f(\Theta_{t+1}) + U_{\Theta_t} + \rho(\Theta_{t+1} - \tilde{\Theta}_t).
\]

Since \(U_{\Theta_{t+1}} = U_{\Theta_t} + \rho(\Theta_{t+1} - \tilde{\Theta}_{t+1})\), we can obtain

\[
0 \in \partial f(\Theta_{t+1}) + U_{\Theta_{t+1}} + \rho(\tilde{\Theta}_{t+1} - \tilde{\Theta}_t).
\]

This implies that \(\Theta_{t+1}\) minimizes

\[
f(\Theta) + (U_{\Theta_{t+1}} + \rho(\tilde{\Theta}_{t+1} - \tilde{\Theta}_t))^T \Theta.
\]
It follows that
\[ f(\Theta_{t+1}) + (U_{\Theta_{t+1}} + \rho(\tilde{\Theta}_{t+1} - \tilde{\Theta}_t))^T \Theta_{t+1} \leq f(\Theta^*) + (U_{\Theta_{t+1}} + \rho(\tilde{\Theta}_{t+1} - \tilde{\Theta}_t))^T \Theta^* \]

Similarly, we have
\[ g(Z_{t+1}) + (U_{Z_{t+1}} + \rho(\tilde{Z}_{t+1} - \tilde{Z}_t))^T Z_{t+1} \leq g(Z^*) + (U_{Z_{t+1}} + \rho(\tilde{Z}_{t+1} - \tilde{Z}_t))^T Z^* , \]

\[ h(V_{t+1}) + (U_{V_{t+1}} - \tilde{U}_{V_{t+1}} + \rho(\tilde{V}_{t+1} - \tilde{V}_t + \hat{V}_{t+1} - \hat{V}_t))^T V_{t+1} \leq h(V^*) + (U_{V_{t+1}} - \tilde{U}_{V_{t+1}} + \rho(\tilde{V}_{t+1} - \tilde{V}_t + \hat{V}_{t+1} - \hat{V}_t))^T V^* \]

\[ h_2(\hat{V}_{t+1} + \check{U}_{V_{t+1}} \hat{V}_{t+1} \leq h_2(\hat{V}^*) + \check{U}_{V_{t+1}} \hat{V}^* \]

\[ \Psi(\check{X}_{t+1}) - U^T_{t+1} \check{X}_{t+1} \leq \Psi(\check{X}^*) - U^T_{t+1} \check{X}^* \]

Add these inequalities above, using \( X^* = \check{X}^* \) and \( V^* = \hat{V}^* \), we obtain
\[ p_{t+1} - p^* \leq U^T_{t+1}(\check{X}_{t+1} - X_{t+1}) + \tilde{U}^T_{V_{t+1}}(V_{t+1} - \tilde{V}_{t+1}) \]
\[ \rho(\check{X}_{t+1} - \check{X}_t)^T (X^* - X_{t+1}) + \rho(\hat{V}_{t+1} - \hat{V}_t)^T (V^* - V_{t+1}) \]

Adding (2) and (3),
\[ (U_{t+1} - U^*)^T (\check{X}_{t+1} - X_{t+1}) + \rho(\check{X}_{t+1} - \check{X}_t)^T (X^* - X_{t+1}) \]
\[ + (\tilde{U}_{V_{t+1}} - \check{U}^T_{V_{t+1}})(V_{t+1} - \tilde{V}_{t+1}) + \rho(\hat{V}_{t+1} - \hat{V}_t)^T (V^* - V_{t+1}) \geq 0 \]

We begin by the first two terms. Substituting \( X^* - X_{t+1} = X^* - \check{X}_{t+1} + \check{X}_{t+1} - X_{t+1} \) gives
\[ (U_{t+1} - U^*)^T (\check{X}_{t+1} - X_{t+1}) + \rho(\check{X}_{t+1} - \check{X}_t)^T (\check{X}_{t+1} - X_{t+1}) + \rho(\check{X}_{t+1} - \check{X}_t)^T (\check{X}^* - \check{X}_{t+1}) \]

For the first term of (5),
\[ (U_t + \rho(X_{t+1} - \check{X}_{t+1}) - U^*)^T (\check{X}_{t+1} - X_{t+1}) \]
\[ = (U_t - U^*)^T (\check{X}_{t+1} - X_{t+1}) - \frac{\rho}{2} \left\| \check{X}_{t+1} - X_{t+1} \right\|^2 - \frac{\rho}{2} \left\| \check{X}_{t+1} - X_{t+1} \right\|^2 \]
\[ = (U_t - U^*)^T (\check{X}_{t+1} - X_{t+1}) - \frac{\rho}{2} \left\| (U_{t+1} - U_t) \right\|^2 - \frac{\rho}{2} \left\| \check{X}_{t+1} - X_{t+1} \right\|^2 \]
\[ = -\frac{1}{\rho}(U_t - U^*)^T (U_{t+1} - U^* - (U_t - U^*)) \]
\[ - \frac{1}{2\rho} \left\| U_{t+1} - U^* - (U_t - U^*) \right\|^2 - \frac{\rho}{2} \left\| \check{X}_{t+1} - X_{t+1} \right\|^2 \]
\[ = \frac{1}{2\rho} \left\| U_t - U^* \right\|^2 - \frac{1}{2\rho} \left\| U_{t+1} - U^* \right\|^2 - \frac{\rho}{2} \left\| \check{X}_{t+1} - X_{t+1} \right\|^2. \]
For the last two terms of (5), taking $\frac{\rho}{2} \|\hat{X}_{t+1} - X_{t+1}\|^2$ from above,

$$-\frac{\rho}{2} \|\hat{X}_{t+1} - X_{t+1}\|^2 + \rho (\hat{X}_{t+1} - \hat{X}_t)^T (\hat{X}_{t+1} - X_{t+1})
+ \rho (\hat{X}_{t+1} - \hat{X}_t)^T (\hat{X}^* - \hat{X}_t + \hat{X}_t - \hat{X}_{t+1})$$

$$= -\frac{\rho}{2} \|\hat{X}_{t+1} - X_{t+1} - (\hat{X}_{t+1} - \hat{X}_t)\|^2 + \rho (\hat{X}_{t+1} - \hat{X}_t)(\hat{X}^* - \hat{X}_t) - \frac{\rho}{2} \|\hat{X}_{t+1} - \hat{X}_t\|^2$$

$$= -\frac{\rho}{2} \|\hat{X}_{t+1} - X_{t+1} - (\hat{X}_{t+1} - \hat{X}_t)\|^2 - \frac{\rho}{2} \|\hat{X}_{t+1} - \hat{X}^*\|^2 + \frac{\rho}{2} \|\hat{X}_{t} - \hat{X}^*\|^2$$

Hence the first two terms of (4) can be rewritten as

$$\frac{1}{\rho} \|U_t - U^*\|^2 - \frac{1}{2 \rho} \|U_{t+1} - U^*\|^2 - \frac{\rho}{2} \|\hat{X}_{t+1} - X_{t+1} - (\hat{X}_{t+1} - \hat{X}_t)\|^2 + \frac{\rho}{2} \|\hat{X}_{t} - \hat{X}^*\|^2 - \frac{\rho}{2} \|\hat{X}_{t+1} - \hat{X}^*\|^2$$

Similarly, the last two terms of (4) can be rewritten as

$$\frac{1}{\rho} \|\tilde{U}_{V_t} - \tilde{U}^*_V\|^2 - \frac{1}{\rho} \|\tilde{U}_{V_{t+1}} - \tilde{U}^*_V\|^2 - \frac{\rho}{2} \|\tilde{V}_{t+1} - V_{t+1} - (\tilde{V}_{t+1} - \tilde{V}_t)\|^2 + \frac{\rho}{2} \|\tilde{V}_t - \tilde{V}^*\|^2 - \frac{\rho}{2} \|\tilde{V}_{t+1} - \tilde{V}^*\|^2$$

This implies that (4) can be written as

$$\frac{1}{\rho} \|U_t - U^*\|^2 + \frac{1}{\rho} \|\tilde{U}_{V_t} - \tilde{U}^*_V\|^2 + \frac{\rho}{2} \|\tilde{X}_t - \tilde{X}^*\|^2 + \frac{\rho}{2} \|\tilde{V}_t - \tilde{V}^*\|^2$$

$$- \frac{1}{\rho} \|U_{t+1} - U^*\|^2 - \frac{1}{\rho} \|\tilde{U}_{V_{t+1}} - \tilde{U}^*_V\|^2 - \frac{\rho}{2} \|\tilde{X}_{t+1} - \tilde{X}^*\|^2 - \frac{\rho}{2} \|\tilde{V}_{t+1} - \tilde{V}^*\|^2$$

$$- \frac{\rho}{2} \|\hat{X}_{t+1} - X_{t+1} - (\hat{X}_{t+1} - \hat{X}_t)\|^2 - \frac{\rho}{2} \|\tilde{V}_{t+1} - V_{t+1} - (\tilde{V}_{t+1} - \tilde{V}_t)\|^2 \geq 0.$$ (6)

Let

$$D_t = \frac{1}{2 \rho} \|U_t - U^*\|^2 + \frac{1}{\rho} \|\tilde{U}_{V_t} - \tilde{U}^*_V\|^2 + \frac{\rho}{2} \|\tilde{X}_t - \tilde{X}^*\|^2 + \frac{\rho}{2} \|\tilde{V}_t - \tilde{V}^*\|^2$$

it follows that

$$D_t - D_{t+1} \geq \frac{\rho}{2} \|\tilde{X}_{t+1} - X_{t+1} - (\hat{X}_{t+1} - \hat{X}_t)\|^2 - \frac{\rho}{2} \|\tilde{V}_{t+1} - V_{t+1} - (\tilde{V}_{t+1} - \tilde{V}_t)\|^2$$ (7)

Recall that $\hat{X}_{t+1}$ minimizes $\Psi(X_t) - U_{t+1}^T \hat{X}$ and $\hat{X}_t$ minimizes $\Psi(X_t) - U_t^T \hat{X}$, so we can add

$$\Psi(\hat{X}_{t+1}) - U_{t+1}^T \hat{X}_{t+1} \leq \Psi(\hat{X}_t) - U_{t+1}^T \hat{X}_t$$

and

$$\Psi(\hat{X}_t) - U_t^T \hat{X}_t \leq \Psi(\hat{X}_{t+1}) - U_{t+1}^T \hat{X}_{t+1}$$
to get
\[(U_{t+1} - U_t)^T(\tilde{X}_{t+1} - \tilde{X}_t) \geq 0.\]

Hence by \(\rho > 0\),
\[(\tilde{X}_{t+1} - X_{t+1})^T(\tilde{X}_{t+1} - \tilde{X}_t) \leq 0.\]

Likewise,
\[(\tilde{V}_{t+1} - V_{t+1})^T(\tilde{V}_{t+1} - \tilde{V}_t) \leq 0.\]

Combined with (7),
\[D_t - D_{t+1} \geq \frac{\rho}{2} \left\| \tilde{X}_{t+1} - X_{t+1} \right\|^2 + \frac{\rho}{2} \left\| \tilde{X}_{t+1} - \tilde{X}_t \right\|^2 + \frac{\rho}{2} \left\| \tilde{V}_{t+1} - V_{t+1} \right\|^2 + \frac{\rho}{2} \left\| \tilde{V}_{t+1} - \tilde{V}_t \right\|^2.\]

Iterating these inequalities gives that
\[
\sum_{t=0}^{\infty} \frac{\rho}{2} \left\| \tilde{X}_{t+1} - X_{t+1} \right\|^2 + \frac{\rho}{2} \left\| \tilde{X}_{t+1} - \tilde{X}_t \right\|^2 + \frac{\rho}{2} \left\| \tilde{V}_{t+1} - V_{t+1} \right\|^2 + \frac{\rho}{2} \left\| \tilde{V}_{t+1} - \tilde{V}_t \right\|^2 \leq D_0
\]

which implies that \(\tilde{X}_{t+1} - X_{t+1} \to 0\) and \(\tilde{V}_{t+1} - V_{t+1} \to 0\). The right hand in (2) goes to zero as \(t \to \infty\). Hence we have \(\lim_{t \to \infty} p_t = p^*\).

Reference