

A. Convergence for BPDN-ADMM (Offline)

The proof of convergence for Algorithm 1 parallels the one given in (Yang & Zhang, 2011) for compressed sensing and the one in (Ma et al., 2012) for latent variable Gaussian graphical model selection. Hence, we only give an outline of the major steps here.

First, we define the convex functions

$$\begin{aligned} F(\boldsymbol{\alpha}) &:= \mathcal{I}(\|\boldsymbol{\alpha}\| \leq \epsilon) \\ G(\boldsymbol{\beta}) &:= \|\boldsymbol{\beta}\|_1, \end{aligned}$$

so that problem (11) can be expressed as

$$\begin{aligned} \min_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \quad & F(\boldsymbol{\alpha}) + G(\boldsymbol{\beta}) \\ \text{s.t.} \quad & \boldsymbol{\alpha} = \mathbf{C}\boldsymbol{\beta} + \mathbf{d}. \end{aligned}$$

The KKT conditions which an optimal solution $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \mathbf{v}^*)$ to this problem satisfies are

$$0 \in \partial F(\boldsymbol{\alpha}^*) - \mathbf{v} \quad (25)$$

$$0 \in \partial G(\boldsymbol{\beta}^*) - \mathbf{C}^\top \mathbf{v}^* \quad (26)$$

$$\boldsymbol{\alpha}^* = \mathbf{C}\boldsymbol{\beta}^* + \mathbf{d}, \quad (27)$$

where \mathbf{v}^* is the optimal dual variable associated with the equality constraint $\boldsymbol{\alpha} = \mathbf{C}\boldsymbol{\beta}$. We then write a single iteration of Algorithm 1 in the following form:

$$\boldsymbol{\alpha}_{j+1} = \arg \min_{\boldsymbol{\alpha}} \left\{ F(\boldsymbol{\alpha}) - \mathbf{v}^\top (\mathbf{C}\boldsymbol{\beta}_j + \mathbf{d} - \boldsymbol{\alpha}) + \frac{1}{2\mu} \|\mathbf{C}\boldsymbol{\beta}_j + \mathbf{d} - \boldsymbol{\alpha}\|^2 \right\} \quad (28)$$

$$\begin{aligned} \boldsymbol{\beta}_{j+1} = \arg \min_{\boldsymbol{\beta}} \left\{ G(\boldsymbol{\beta}) + \right. \\ \left. \frac{1}{2\mu} \|\boldsymbol{\beta} - (\boldsymbol{\beta}_j - \tau \mathbf{C}^\top (\mathbf{C}\boldsymbol{\beta}_j + \mathbf{d} - \boldsymbol{\alpha}_{j+1} - \mu \mathbf{v}))\|^2 \right\} \quad (29) \end{aligned}$$

$$\mathbf{v}_{j+1} = \mathbf{v}_j - \frac{1}{\mu} (\mathbf{C}\boldsymbol{\beta}_{j+1} + \mathbf{d} - \boldsymbol{\alpha}_{j+1}) \quad (30)$$

By the first-order optimality conditions of the sub-problems (28) and (29) and the fact that $\partial F(\cdot)$ and $\partial G(\cdot)$ are monotone operators (since F and G are convex functions), we have that

$$\langle \boldsymbol{\alpha}_{j+1} - \boldsymbol{\alpha}^*, \mathbf{v}_{j+1} - \mathbf{v}^* + \frac{1}{\mu} \mathbf{C}(\boldsymbol{\beta}_{j+1} - \boldsymbol{\beta}_j) \rangle \geq 0, \quad (31)$$

$$\begin{aligned} \langle \boldsymbol{\beta}_{j+1} - \boldsymbol{\beta}^*, \frac{1}{\mu\tau} (\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}) - \frac{1}{\mu} \mathbf{C}^\top \mathbf{C}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}) \\ + \mathbf{C}^\top (\mathbf{v}_{j+1} - \mathbf{v}^*) \rangle \geq 0. \quad (32) \end{aligned}$$

Combining the above two inequalities and using (27) and (30), we have

$$\begin{aligned} \frac{1}{\mu\tau} \langle \boldsymbol{\beta}_{j+1} - \boldsymbol{\beta}^*, \boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1} \rangle + \mu \langle \mathbf{v}_{j+1} - \mathbf{v}^*, \mathbf{v}_j - \mathbf{v}_{j+1} \rangle \\ \geq \langle \mathbf{v}_j - \mathbf{v}_{j+1}, \mathbf{C}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}) \rangle. \quad (33) \end{aligned}$$

Now, define $\mathbf{u}_j := \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{v} \end{pmatrix}$, $\mathbf{u}^* := \begin{pmatrix} \boldsymbol{\beta}^* \\ \mathbf{v}^* \end{pmatrix}$, and $\mathbf{H} := \begin{pmatrix} \frac{1}{\mu\tau} \mathbf{I} & 0 \\ 0 & \mu \mathbf{I} \end{pmatrix}$. Define the norm $\|\mathbf{u}\|_{\mathbf{H}} := \mathbf{u}^\top \mathbf{H} \mathbf{u}$ and the inner product $\langle \mathbf{u}, \mathbf{w} \rangle_{\mathbf{H}} := \mathbf{u}^\top \mathbf{H} \mathbf{w}$. Then, it can be shown that

$$\begin{aligned} \|\mathbf{u}_{j+1} - \mathbf{u}^*\|_{\mathbf{H}} - \|\mathbf{u}_{j+1} - \mathbf{u}^*\|_{\mathbf{H}} \geq \|\mathbf{u}_{j+1} - \mathbf{u}_j\|_{\mathbf{H}} + \\ 2 \langle \mathbf{v}_j - \mathbf{v}_{j+1}, \mathbf{C}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}) \rangle. \end{aligned}$$

Hence, by noting that $2 \langle \mathbf{v}_j - \mathbf{v}_{j+1}, \mathbf{C}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}) \rangle \geq -\rho \|\mathbf{v}_j - \mathbf{v}_{j+1}\|^2 - \frac{\lambda_{\max}(\mathbf{C}^\top \mathbf{C})}{\rho} \|\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j+1}\|^2$ for any $\rho > 0$, we can prove the following result.

Lemma A.1. *For any step size τ satisfying $0 < \tau < \frac{1}{\lambda_{\max}(\mathbf{C}^\top \mathbf{C})}$, there exists a positive η such that the sequence $\{(\boldsymbol{\alpha}_j, \boldsymbol{\beta}_j, \mathbf{v}_j)\}$ generated by Algorithm 1 satisfies*

$$\|\mathbf{u}_j - \mathbf{u}^*\|_{\mathbf{H}}^2 - \|\mathbf{u}_{j+1} - \mathbf{u}^*\|_{\mathbf{H}}^2 \geq \eta \|\mathbf{u}_j - \mathbf{u}_{j+1}\|_{\mathbf{H}}^2. \quad (34)$$

Now, Lemma A.1 implies that

- $\|\mathbf{u}_j - \mathbf{u}_{j+1}\|_{\mathbf{H}} \rightarrow 0$;
- The sequence $\{\mathbf{u}_j\}$ lies in a compact region;
- The sequence $\{\|\mathbf{u}_j - \mathbf{u}^*\|_{\mathbf{H}}^2\}$ is monotonically non-increasing and thus converges.

It is then straightforward to show that any limit point of the sequence $\{(\boldsymbol{\alpha}_j, \boldsymbol{\beta}_j, \mathbf{v}_j)\}$ satisfies the KKT conditions (25), (26), and (27), and therefore, is an optimal solution to problem (11):