Convergence of ADMM

$$\min_{\boldsymbol{x},\boldsymbol{y}} f(\boldsymbol{x}) + g(\boldsymbol{y}) \text{ s.t. } \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} = \boldsymbol{0}$$

f, g convex, $\boldsymbol{A}, \boldsymbol{B}$ full rank.

$$\min_{\boldsymbol{x},\boldsymbol{y}} \max_{\boldsymbol{\lambda}} \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda})$$

$$\mathcal{L}_\rho = f(\boldsymbol{x}) + g(\boldsymbol{y}) + \langle \boldsymbol{\lambda}, \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} \rangle + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c}\|_2^2.$$

for $k = 1, 2, \ldots$ **do**

$$\boldsymbol{x}_{k+1} = \operatorname*{argmin}_{\boldsymbol{x}} \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}_k, \boldsymbol{\lambda}_k)$$

$$\boldsymbol{y}_{k+1} = \operatorname*{argmin}_{\boldsymbol{y}} \mathcal{L}_\rho(\boldsymbol{x}_{k+1}, \boldsymbol{y}, \boldsymbol{\lambda}_k)$$

$$\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \rho(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})$$

Andersen Ang
Dept. Combinatorics & Optimization, University of Waterloo, Canada
ms$\boldsymbol{x}$ang@uwaterloo.ca, where $\boldsymbol{x} = \lfloor \pi \rfloor$ Homepage: angms.science

First draft: November 3, 2022 Last update: November 14, 2022

Convergence of ADMM vs Convergence of gradient descent

Convergence of ADMM

$$\min_{\boldsymbol{x},\boldsymbol{y}} f(\boldsymbol{x}) + g(\boldsymbol{y}) \text{ s.t. } \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} = \boldsymbol{0}, \; f, g \text{ convex}, \boldsymbol{A}, \boldsymbol{B} \text{ full rank}.$$

$$\min_{\boldsymbol{x},\boldsymbol{y}} \max_{\boldsymbol{\lambda}} \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + g(\boldsymbol{y}) + \langle \boldsymbol{\lambda}, \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} \rangle + \frac{\rho}{2} \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c}\|_2^2.$$

- Convergence of primal function value $f(\boldsymbol{x}) + g(\boldsymbol{y}) \longrightarrow f^* + g^*$
- Convergence of primal variables $\boldsymbol{x} \longrightarrow \boldsymbol{x}^*, \; \boldsymbol{y} \longrightarrow \boldsymbol{y}^*$
- Convergence of feasibility gap $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} \longrightarrow \boldsymbol{0}$
- Convergence of dual function value $d(\boldsymbol{\lambda}) \longrightarrow d^*$
- Convergence of dual variable $\boldsymbol{\lambda} \longrightarrow \boldsymbol{\lambda}^*$
- Convergence of 1st-order optimality on $\mathcal{L}_\rho(\boldsymbol{x})$ $\partial f(\boldsymbol{x}) + \boldsymbol{A}^\top \boldsymbol{\lambda} + \rho \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c}) \overset{\leftarrow}{\ni} \boldsymbol{0}$
- Convergence of 1st-order optimality on $\mathcal{L}_\rho(\boldsymbol{y})$ $\partial g(\boldsymbol{x}) + \boldsymbol{B}^\top \boldsymbol{\lambda} + \rho \boldsymbol{B}^\top (\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c}) \overset{\leftarrow}{\ni} \boldsymbol{0}$
- Convergence to a KKT point combination of above
- Convergence rate of above

Gradient descent

$$\min_{\boldsymbol{x}} F(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x}).$$

- Convergence of primal function value $F(\boldsymbol{x}) \longrightarrow F^*$
- Convergence of primal variables $\boldsymbol{x} \longrightarrow \boldsymbol{x}^*$
- Convergence of 1st-order optimality $\partial F(\boldsymbol{x}) \overset{\leftarrow}{\ni} \boldsymbol{0}$
- (If f is twice differentiable) Convergence of 2nd-order optimality
- Convergence rate of above

Remarks

$$\min_{\boldsymbol{x},\boldsymbol{y}} f(\boldsymbol{x}) + g(\boldsymbol{y}) \text{ s.t. } \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} = \boldsymbol{0}, \ f, g \text{ convex, } \boldsymbol{A}, \boldsymbol{B} \text{ full rank.}$$

$$\min_{\boldsymbol{x},\boldsymbol{y}} \max_{\boldsymbol{\lambda}} \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + g(\boldsymbol{y}) + \langle \boldsymbol{\lambda}, \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} \rangle + \frac{\rho}{2} \| \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} \|_2^2.$$

for $k = 1, 2, \ldots$ **do**

$$\boldsymbol{x}_{k+1} = \operatorname*{argmin}_{\boldsymbol{x}} \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}_k, \boldsymbol{\lambda}_k)$$

$$\boldsymbol{y}_{k+1} = \operatorname*{argmin}_{\boldsymbol{y}} \mathcal{L}_\rho(\boldsymbol{x}_{k+1}, \boldsymbol{y}, \boldsymbol{\lambda}_k)$$

$$\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \rho(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})$$

- Does ADMM iterate $(\boldsymbol{x}_k, \boldsymbol{y}_k, \boldsymbol{\lambda}_k)$ solves $\mathcal{P}$? No.
- Does ADMM iterate $(\boldsymbol{x}_k, \boldsymbol{y}_k, \boldsymbol{\lambda}_k)$ is feasible for $\mathcal{P}$? No.
- When does ADMM iterate $(\boldsymbol{x}_k, \boldsymbol{y}_k, \boldsymbol{\lambda}_k)$ solves $\mathcal{P}$? $k \to \infty$
- When does ADMM iterate $(\boldsymbol{x}_k, \boldsymbol{y}_k, \boldsymbol{\lambda}_k)$ is feasible for $\mathcal{P}$? $k \to \infty$
- As a result, ADMM is not for applications that requires feasibility at all iterates.

What does ADMM iterate $(\boldsymbol{x}_k, \boldsymbol{y}_k, \boldsymbol{\lambda}_k)$ solves?

$$\min_{\boldsymbol{x},\boldsymbol{y}} f(\boldsymbol{x}) + g(\boldsymbol{y}) \text{ s.t. } \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} = \boldsymbol{0}, \quad f, g \text{ convex, } \boldsymbol{A}, \boldsymbol{B} \text{ full rank.}$$

$$\min_{\boldsymbol{x},\boldsymbol{y}} \max_{\boldsymbol{\lambda}} \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + g(\boldsymbol{y}) + \langle \boldsymbol{\lambda}, \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} \rangle + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c}\|_2^2.$$

for $k = 1, 2, \ldots$ **do**

$$\boldsymbol{x}_{k+1} = \operatorname*{argmin}_{\boldsymbol{x}} \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}_k, \boldsymbol{\lambda}_k)$$

$$\boldsymbol{y}_{k+1} = \operatorname*{argmin}_{\boldsymbol{y}} \mathcal{L}_\rho(\boldsymbol{x}_{k+1}, \boldsymbol{y}, \boldsymbol{\lambda}_k)$$

$$\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \rho(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})$$

- Recall: for $\min \phi(\boldsymbol{\zeta})$, gradient descent

  $$\boldsymbol{\zeta}_{k+1} = \boldsymbol{\zeta}_k - \alpha_k \nabla\phi(\boldsymbol{\zeta}_k) = \operatorname*{argmin}_{\boldsymbol{\zeta}} \phi(\boldsymbol{\zeta}_k) + \langle \nabla\phi(\boldsymbol{\zeta}_k), \boldsymbol{\zeta} - \boldsymbol{\zeta}_k \rangle + \frac{1}{2\alpha_k}\|\boldsymbol{\zeta} - \boldsymbol{\zeta}_k\|_2^2.$$

- Now same for dual ascent: $\boldsymbol{\lambda}_k + \rho\nabla_{\boldsymbol{\lambda}} d(\boldsymbol{\lambda}) = \operatorname*{argmax}_{\boldsymbol{\lambda}} d(\boldsymbol{\lambda}_k) - \langle \nabla d(\boldsymbol{\lambda}_k), \boldsymbol{\lambda} - \boldsymbol{\lambda}_k \rangle + \frac{1}{2\rho}\|\boldsymbol{\lambda} - \boldsymbol{\lambda}_k\|_2^2.$

  So dual ascent = find a new $\boldsymbol{\lambda}$ based on a local quadratic model of d at iteration k.

- Optimality condition of ADMM iterate $(\boldsymbol{x}_k, \boldsymbol{y}_k, \boldsymbol{\lambda}_k)$
  - $\boldsymbol{0} \in \partial f(\boldsymbol{x}_{k+1}) + \boldsymbol{A}^\top\boldsymbol{\lambda}_k + \rho\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c})$ — subgradient 1st-order optimality on $\boldsymbol{x}_{k+1}$
  - $\boldsymbol{0} \in \partial g(\boldsymbol{y}_{k+1}) + \boldsymbol{B}^\top\boldsymbol{\lambda}_k + \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})$ — subgradient 1st-order optimality on $\boldsymbol{y}_{k+1}$
  - $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \rho(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})$ — gradient ascent is already 1st-order optimal

Optimality condition of ADMM iterate $(\boldsymbol{x}_k, \boldsymbol{y}_k, \boldsymbol{\lambda}_k)$ and $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\lambda}^*)$

The convergence of ADMM builds upon these 6 expressions:

1. $\mathbf{0} \in \partial f(\boldsymbol{x}_{k+1}) + \boldsymbol{A}^\top \boldsymbol{\lambda}_k + \rho \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c})$ — 1st-order optimality on $\boldsymbol{x}_{k+1}$
2. $\mathbf{0} \in \partial g(\boldsymbol{y}_{k+1}) + \boldsymbol{B}^\top \boldsymbol{\lambda}_k + \rho \boldsymbol{B}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})$ — 1st-order optimality on $\boldsymbol{y}_{k+1}$
3. $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \rho(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})$ — gradient ascent

4. $\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}^* + \boldsymbol{c} = \mathbf{0}$ — primal feasibility
5. $\mathbf{0} \in \partial f(\boldsymbol{x}^*) + \boldsymbol{A}^\top \boldsymbol{\lambda}^*$ — 1st-order optimality on $\boldsymbol{x}^*$
6. $\mathbf{0} \in \partial g(\boldsymbol{y}^*) + \boldsymbol{B}^\top \boldsymbol{\lambda}^*$ — 1st-order optimality on $\boldsymbol{y}^*$

What about the KKT condition?

- (4)-(6) are basically KKT condition of equality-constrained problem

$$\boldsymbol{x}^* \text{ solves } \begin{cases} \min_{\boldsymbol{x}} & f(\boldsymbol{x}) \\ \text{s.t.} & c(\boldsymbol{x}) = \mathbf{0} \end{cases} \overset{\text{KKT}}{\Longleftrightarrow} \underbrace{c(\boldsymbol{x}^*) = \mathbf{0}}_{\text{primal feasibility}}, \quad \underbrace{\mathbf{0} \in \partial f(\boldsymbol{x}^*) + \langle \overbrace{J(\boldsymbol{x}^*)}^{\text{Jacobian of } c}, \boldsymbol{\lambda}^* \rangle}_{\text{dual feasibility}}.$$

- In fact, we can derive KKT condition from subgradient optimality.

Convergence of ADMM (minimum condition)

$$\min_{\boldsymbol{x},\boldsymbol{y}} f(\boldsymbol{x}) + g(\boldsymbol{y}) \text{ s.t. } \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} = \boldsymbol{0}, \quad f, g \text{ convex, } \boldsymbol{A}, \boldsymbol{B} \text{ full rank.}$$

$$\min_{\boldsymbol{x},\boldsymbol{y}} \max_{\boldsymbol{\lambda}} \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + g(\boldsymbol{y}) + \langle \boldsymbol{\lambda}, \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} \rangle + \frac{\rho}{2} \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c}\|_2^2.$$

for $k = 1, 2, \ldots$ **do**

$$\boldsymbol{x}_{k+1} = \operatorname*{argmin}_{\boldsymbol{x}} \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}_k, \boldsymbol{\lambda}_k)$$
$$\boldsymbol{y}_{k+1} = \operatorname*{argmin}_{\boldsymbol{y}} \mathcal{L}_\rho(\boldsymbol{x}_{k+1}, \boldsymbol{y}, \boldsymbol{\lambda}_k)$$
$$\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \rho(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})$$

Theorem The ADMM iterate $(\boldsymbol{x}_{k+1}, \boldsymbol{y}_{k+1})$ satisfies

- Primal objective function converges

$$f(\boldsymbol{x}_{k+1}) + g(\boldsymbol{y}_{k+1}) \to f(\boldsymbol{x}^*) + g(\boldsymbol{y}^*).$$

- Primal feasibility gap goes to zero

$$\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c} \to \boldsymbol{0}.$$

  - $\implies \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k \to \boldsymbol{0}$
  - $\implies$ dual feasibility gap goes to zero

Convergence of ADMM

$$\min_{\boldsymbol{x},\boldsymbol{y}} f(\boldsymbol{x}) + g(\boldsymbol{y}) \text{ s.t. } \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} = \boldsymbol{0}, \quad f, g \text{ convex, } \boldsymbol{A}, \boldsymbol{B} \text{ full rank.}$$

$$\min_{\boldsymbol{x},\boldsymbol{y}} \max_{\boldsymbol{\lambda}} \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + g(\boldsymbol{y}) + \langle \boldsymbol{\lambda}, \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} \rangle + \frac{\rho}{2} \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c}\|_2^2.$$

Convergence of primal function value	$f(\boldsymbol{x}) + g(\boldsymbol{y}) \longrightarrow f^* + g^*$
Convergence of primal variables	$\boldsymbol{x} \longrightarrow \boldsymbol{x}^*, \ \boldsymbol{y} \longrightarrow \boldsymbol{y}^*$
Convergence of feasibility gap	$\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c} \longrightarrow \boldsymbol{0}$
Convergence of dual function value	$d(\boldsymbol{\lambda}) \longrightarrow d^*$
Convergence of dual variable	$\boldsymbol{\lambda} \longrightarrow \boldsymbol{\lambda}^*$
Convergence of 1st-order optimality on $\mathcal{L}_\rho(\boldsymbol{x})$	$\partial f(\boldsymbol{x}) + \boldsymbol{A}^\top \boldsymbol{\lambda} + \rho \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c}) \stackrel{\leftarrow}{\ni} \boldsymbol{0}$
Convergence of 1st-order optimality on $\mathcal{L}_\rho(\boldsymbol{y})$	$\partial g(\boldsymbol{x}) + \boldsymbol{B}^\top \boldsymbol{\lambda} + \rho \boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{c}) \stackrel{\leftarrow}{\ni} \boldsymbol{0}$
Convergence to a KKT point	combination of above
Convergence rate of above	

Preparation

- Shorthand notation
  - $f^* := f(\boldsymbol{x}^*)$
  - $g^* := g(\boldsymbol{y}^*)$
  - $f_k = f(\boldsymbol{x}_k)$
  - $g_k = g(\boldsymbol{y}_k)$
  - $\nabla f_k = \nabla f(\boldsymbol{x}_k)$
  - $\nabla g_k = \nabla g(\boldsymbol{y}_k)$

- Tools we need
  - Subgradient inequality of convexity of f and g

$$\boldsymbol{q} \in \partial f(\bar{\boldsymbol{x}}) \iff f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}}) + \langle \boldsymbol{q}, \boldsymbol{x} - \bar{\boldsymbol{x}}\rangle \iff \langle \boldsymbol{q}, \bar{\boldsymbol{x}} - \boldsymbol{x}\rangle \geq f(\bar{\boldsymbol{x}}) - f(\boldsymbol{x}).$$

  - Subgradient is monotone

$$\langle \partial f_k - \partial f_{k-1}, \boldsymbol{x}_k - \boldsymbol{x}_{k-1}\rangle \geq 0.$$

  - A tricky inequality

$$\langle \boldsymbol{a}, \boldsymbol{b}\rangle = \frac{\|\boldsymbol{a}\|_2^2 + \|\boldsymbol{b}\|_2^2 - \|\boldsymbol{a}-\boldsymbol{b}\|_2^2}{2}.$$

  - KKT point is a saddle point for (non-augmented) Lagrangian

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\lambda}) \leq \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\lambda}^*) \leq \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}^*).$$

Important subgradients

- Subgradient of f and g at the $k+1$ iteration
  - $-\boldsymbol{A}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_k+\boldsymbol{c}) \in \partial f(\boldsymbol{x}_{k+1})$
  - $\underbrace{-\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c})}_{=-\boldsymbol{B}^\top\boldsymbol{\lambda}_{k+1}} \in \partial g(\boldsymbol{y}_{k+1})$
- Subgradient of f and g at convergence
  - $-\boldsymbol{A}^\top\boldsymbol{\lambda}^* \in \partial f^*$
  - $-\boldsymbol{B}^\top\boldsymbol{\lambda}^* \in \partial g^*$
- Convexity of f

$$f^* \geq f(\boldsymbol{x}_{k+1}) + \langle \partial f_{k+1}, \boldsymbol{x}^*-\boldsymbol{x}_{k+1}\rangle \implies f^* \geq f(\boldsymbol{x}_{k+1}) + \langle -\boldsymbol{A}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_k+\boldsymbol{c}), \boldsymbol{x}^*-\boldsymbol{x}_{k+1}\rangle$$

- Convexity of g

$$g^* \geq g(\boldsymbol{y}_{k+1}) + \langle \partial g_{k+1}, \boldsymbol{y}^*-\boldsymbol{y}_{k+1}\rangle \implies g^* \geq g(\boldsymbol{y}_{k+1}) + \langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c}), \boldsymbol{y}^*-\boldsymbol{y}_{k+1}\rangle$$

- Sum

$$f^*+g^* \geq \begin{array}{l} f_{k+1}+g_{k+1} \\ +\langle -\boldsymbol{A}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_k+\boldsymbol{c}), \boldsymbol{x}^*-\boldsymbol{x}_{k+1}\rangle \\ +\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c}), \boldsymbol{y}^*-\boldsymbol{y}_{k+1}\rangle \end{array}$$

- Rearrange

$$f_{k+1}+g_{k+1}-f^*-g^* \leq \begin{array}{l} \langle -\boldsymbol{A}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_k+\boldsymbol{c}), \boldsymbol{x}_{k+1}-\boldsymbol{x}^*\rangle \\ +\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c}), \boldsymbol{y}_{k+1}-\boldsymbol{y}^*\rangle \end{array}$$

$\min_{\boldsymbol{x},\boldsymbol{y}} f(\boldsymbol{x})+g(\boldsymbol{y})$ s.t. $\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}+\boldsymbol{c}=\boldsymbol{0}$, f,g cvx, $\boldsymbol{A},\boldsymbol{B}$ full rank, $\min_{\boldsymbol{x},\boldsymbol{y}}\max_{\boldsymbol{\lambda}} \mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda}) = f(\boldsymbol{x})+g(\boldsymbol{y})+\langle \boldsymbol{\lambda}, \boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}+\boldsymbol{c}\rangle+\frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}+\boldsymbol{c}\|_2^2$.

1. $\boldsymbol{0} \in \partial f(\boldsymbol{x}_{k+1}) + \boldsymbol{A}^\top \boldsymbol{\lambda}_k + \rho \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c})$
2. $\boldsymbol{0} \in \partial g(\boldsymbol{y}_{k+1}) + \boldsymbol{B}^\top \boldsymbol{\lambda}_k + \rho \boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})$
3. $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \rho(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})$

- $-\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}) \in \partial f(\boldsymbol{x}_{k+1})$
- $\underbrace{-\boldsymbol{B}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})}_{=-\boldsymbol{B}^\top\boldsymbol{\lambda}_{k+1}} \in \partial g(\boldsymbol{y}_{k+1})$

4. $\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}^* + \boldsymbol{c} = \boldsymbol{0}$
5. $\boldsymbol{0} \in \partial f(\boldsymbol{x}^*) + \boldsymbol{A}^\top \boldsymbol{\lambda}^*$
6. $\boldsymbol{0} \in \partial g(\boldsymbol{y}^*) + \boldsymbol{B}^\top \boldsymbol{\lambda}^*$

- $-\boldsymbol{A}^\top \boldsymbol{\lambda}^* \in \partial f^*$
- $-\boldsymbol{B}^\top \boldsymbol{\lambda}^* \in \partial g^*$

Lemma 1a $\Big\langle -\boldsymbol{B}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}), \boldsymbol{y}_{k+1} - \boldsymbol{y}\Big\rangle = \Big\langle -\boldsymbol{\lambda}_{k+1}, \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}\Big\rangle$

Proof

$$\begin{aligned} & \Big\langle -\boldsymbol{B}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}), \boldsymbol{y}_{k+1} - \boldsymbol{y}\Big\rangle && \text{copy the statement} \\ = & \Big\langle -\boldsymbol{\lambda}_k - \rho(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}), \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}\Big\rangle && \text{move } \boldsymbol{B} \text{ to the right} \\ = & \Big\langle -\boldsymbol{\lambda}_{k+1}, \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}\Big\rangle && \text{definition of } \boldsymbol{\lambda}_{k+1} \quad \blacksquare \end{aligned}$$

Comment: why start with $\boldsymbol{y}$? Because we can use definition of $\boldsymbol{\lambda}_{k+1}$ after the $\boldsymbol{y}$-step.

$\min_{\boldsymbol{x},\boldsymbol{y}} f(\boldsymbol{x})+g(\boldsymbol{y})$ s.t. $\boldsymbol{Ax}+\boldsymbol{By}+\boldsymbol{c}=\boldsymbol{0}$, f, g cvx, $\boldsymbol{A}$, $\boldsymbol{B}$ full rank,

1. $\boldsymbol{0} \in \partial f(\boldsymbol{x}_{k+1}) + \boldsymbol{A}^\top \boldsymbol{\lambda}_k + \rho \boldsymbol{A}^\top (\boldsymbol{Ax}_{k+1} + \boldsymbol{By}_k + \boldsymbol{c})$
2. $\boldsymbol{0} \in \partial g(\boldsymbol{y}_{k+1}) + \boldsymbol{B}^\top \boldsymbol{\lambda}_k + \rho \boldsymbol{B}^\top (\boldsymbol{Ax}_{k+1} + \boldsymbol{By}_{k+1} + \boldsymbol{c})$
3. $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \rho(\boldsymbol{Ax}_{k+1} + \boldsymbol{By}_{k+1} + \boldsymbol{c})$

- $-\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{A}^\top (\boldsymbol{Ax}_{k+1} + \boldsymbol{By}_k + \boldsymbol{c}) \in \partial f(\boldsymbol{x}_{k+1})$
- $\underbrace{-\boldsymbol{B}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{B}^\top (\boldsymbol{Ax}_{k+1} + \boldsymbol{By}_{k+1} + \boldsymbol{c})}_{=-\boldsymbol{B}^\top \boldsymbol{\lambda}_{k+1}} \in \partial g(\boldsymbol{y}_{k+1})$

$\min_{\boldsymbol{x},\boldsymbol{y}} \max_{\boldsymbol{\lambda}} \mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda}) = f(\boldsymbol{x})+g(\boldsymbol{y})+\langle \boldsymbol{\lambda}, \boldsymbol{Ax}+\boldsymbol{By}+\boldsymbol{c}\rangle + \frac{\rho}{2}\|\boldsymbol{Ax}+\boldsymbol{By}+\boldsymbol{c}\|_2^2.$

4. $\boldsymbol{Ax}^* + \boldsymbol{By}^* + \boldsymbol{c} = \boldsymbol{0}$
5. $\boldsymbol{0} \in \partial f(\boldsymbol{x}^*) + \boldsymbol{A}^\top \boldsymbol{\lambda}^*$
6. $\boldsymbol{0} \in \partial g(\boldsymbol{y}^*) + \boldsymbol{B}^\top \boldsymbol{\lambda}^*$

- $-\boldsymbol{A}^\top \boldsymbol{\lambda}^* \in \partial f^*$
- $-\boldsymbol{B}^\top \boldsymbol{\lambda}^* \in \partial g^*$

Lemma 1b

$$\Big\langle -\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{A}^\top (\boldsymbol{Ax}_{k+1} + \boldsymbol{By}_k + \boldsymbol{c}), \boldsymbol{x}_{k+1} - \boldsymbol{x} \Big\rangle = \Big\langle -\boldsymbol{\lambda}_{k+1}, \boldsymbol{Ax}_{k+1} - \boldsymbol{Ax} \Big\rangle + \rho \Big\langle \boldsymbol{By}_{k+1} - \boldsymbol{By}_k, \boldsymbol{Ax}_{k+1} - \boldsymbol{Ax} \Big\rangle$$

Proof

$$\begin{aligned}
& \Big\langle -\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{A}^\top (\boldsymbol{Ax}_{k+1} + \boldsymbol{By}_k + \boldsymbol{c}), \boldsymbol{x}_{k+1} - \boldsymbol{x} \Big\rangle && \text{copy the statement} \\
= & \Big\langle -\boldsymbol{\lambda}_k - \rho(\boldsymbol{Ax}_{k+1} + \boldsymbol{By}_k + \boldsymbol{c}), \boldsymbol{Ax}_{k+1} - \boldsymbol{Ax} \Big\rangle && \text{move } \boldsymbol{A} \text{ to right} \\
= & \Big\langle -\boldsymbol{\lambda}_k - \rho(\boldsymbol{Ax}_{k+1} + \boldsymbol{B}(\boldsymbol{y}_k + \boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}) + \boldsymbol{c}), \boldsymbol{Ax}_{k+1} - \boldsymbol{Ax} \Big\rangle \\
= & \Big\langle -\boldsymbol{\lambda}_k - \rho(\boldsymbol{Ax}_{k+1} + \boldsymbol{B}(\boldsymbol{y}_{k+1}) + \boldsymbol{c}), \boldsymbol{Ax}_{k+1} - \boldsymbol{Ax} \Big\rangle + \Big\langle -\rho \boldsymbol{B}(\boldsymbol{y}_k - \boldsymbol{y}_{k+1}), \boldsymbol{Ax}_{k+1} - \boldsymbol{Ax} \Big\rangle \\
= & \Big\langle -\boldsymbol{\lambda}_{k+1}, \boldsymbol{Ax}_{k+1} - \boldsymbol{Ax} \Big\rangle + \rho \Big\langle \boldsymbol{By}_{k+1} - \boldsymbol{By}_k, \boldsymbol{Ax}_{k+1} - \boldsymbol{Ax} \Big\rangle && \text{definition of } \boldsymbol{\lambda}_{k+1} \ \blacksquare
\end{aligned}$$

Comment: for the $\boldsymbol{x}$-step, we have to create $\boldsymbol{y}_{k+1}$ to use $\boldsymbol{\lambda}_{k+1}$.

$\min_{\boldsymbol{x},\boldsymbol{y}} f(\boldsymbol{x})+g(\boldsymbol{y})$ s.t. $\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}+\boldsymbol{c}=\boldsymbol{0}$, f, g cvx, $\boldsymbol{A}, \boldsymbol{B}$ full rank, $\min_{\boldsymbol{x},\boldsymbol{y}} \max_{\boldsymbol{\lambda}} \mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda}) = f(\boldsymbol{x})+g(\boldsymbol{y})+\langle \boldsymbol{\lambda}, \boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}+\boldsymbol{c}\rangle + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}+\boldsymbol{c}\|_2^2$.

1. $\boldsymbol{0} \in \partial f(\boldsymbol{x}_{k+1}) + \boldsymbol{A}^\top \boldsymbol{\lambda}_k + \rho \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c})$
2. $\boldsymbol{0} \in \partial g(\boldsymbol{y}_{k+1}) + \boldsymbol{B}^\top \boldsymbol{\lambda}_k + \rho \boldsymbol{B}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})$
3. $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \rho(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})$
4. $\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}^* + \boldsymbol{c} = \boldsymbol{0}$
5. $\boldsymbol{0} \in \partial f(\boldsymbol{x}^*) + \boldsymbol{A}^\top \boldsymbol{\lambda}^*$
6. $\boldsymbol{0} \in \partial g(\boldsymbol{y}^*) + \boldsymbol{B}^\top \boldsymbol{\lambda}^*$

- $-\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}) \in \partial f(\boldsymbol{x}_{k+1})$
- $\underbrace{-\boldsymbol{B}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{B}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})}_{=-\boldsymbol{B}^\top \boldsymbol{\lambda}_{k+1}} \in \partial g(\boldsymbol{y}_{k+1})$

- $-\boldsymbol{A}^\top \boldsymbol{\lambda}^* \in \partial f^*$
- $-\boldsymbol{B}^\top \boldsymbol{\lambda}^* \in \partial g^*$

Lemma 1c

$$\begin{aligned}&\left\langle -\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}), \boldsymbol{x}_{k+1} - \boldsymbol{x} \right\rangle \\ &+ \left\langle -\boldsymbol{B}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{B}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}), \boldsymbol{y}_{k+1} - \boldsymbol{y} \right\rangle = \left\langle -\boldsymbol{\lambda}_{k+1}, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y} \right\rangle \\ &\qquad + \rho \left\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x} \right\rangle\end{aligned}$$

Proof

$$\left\langle -\boldsymbol{B}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{B}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}), \boldsymbol{y}_{k+1} - \boldsymbol{y} \right\rangle = \left\langle -\boldsymbol{\lambda}_{k+1}, \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y} \right\rangle \quad \text{lemma 1a}$$

$$\left\langle -\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}), \boldsymbol{x}_{k+1} - \boldsymbol{x} \right\rangle = \left\langle -\boldsymbol{\lambda}_{k+1}, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x} \right\rangle + \rho\left\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x} \right\rangle \quad \text{lemma 1b}$$

Lemma 1c = Lemma 1a + Lemma 1b. ■

$\min_{\boldsymbol{x},\boldsymbol{y}} f(\boldsymbol{x})+g(\boldsymbol{y})$ s.t. $\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}+\boldsymbol{c}=\boldsymbol{0}$, f,g cvx, $\boldsymbol{A},\boldsymbol{B}$ full rank, $\min_{\boldsymbol{x},\boldsymbol{y}}\max_{\boldsymbol{\lambda}} \mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda}) = f(\boldsymbol{x})+g(\boldsymbol{y})+\langle\boldsymbol{\lambda},\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}+\boldsymbol{c}\rangle+\frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}+\boldsymbol{c}\|_2^2$.

1. $\boldsymbol{0} \in \partial f(\boldsymbol{x}_{k+1}) + \boldsymbol{A}^\top\boldsymbol{\lambda}_k + \rho\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_k+\boldsymbol{c})$
2. $\boldsymbol{0} \in \partial g(\boldsymbol{y}_{k+1}) + \boldsymbol{B}^\top\boldsymbol{\lambda}_k + \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c})$
3. $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \rho(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c})$

- $-\boldsymbol{A}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_k+\boldsymbol{c}) \in \partial f(\boldsymbol{x}_{k+1})$
- $\underbrace{-\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c})}_{=-\boldsymbol{B}^\top\boldsymbol{\lambda}_{k+1}} \in \partial g(\boldsymbol{y}_{k+1})$

4. $\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}^* + \boldsymbol{c} = \boldsymbol{0}$
5. $\boldsymbol{0} \in \partial f(\boldsymbol{x}^*) + \boldsymbol{A}^\top\boldsymbol{\lambda}^*$
6. $\boldsymbol{0} \in \partial g(\boldsymbol{y}^*) + \boldsymbol{B}^\top\boldsymbol{\lambda}^*$

- $-\boldsymbol{A}^\top\boldsymbol{\lambda}^* \in \partial f^*$
- $-\boldsymbol{B}^\top\boldsymbol{\lambda}^* \in \partial g^*$

Recall why lemma 1c is important

$$\begin{aligned}&\Big\langle -\boldsymbol{A}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_k+\boldsymbol{c}),\, \boldsymbol{x}_{k+1}-\boldsymbol{x}\Big\rangle \\ &+\Big\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c}),\, \boldsymbol{y}_{k+1}-\boldsymbol{y}\Big\rangle\end{aligned} = \begin{aligned}&\Big\langle -\boldsymbol{\lambda}_{k+1},\, \boldsymbol{A}\boldsymbol{x}_{k+1}-\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}_{k+1}-\boldsymbol{B}\boldsymbol{y}\Big\rangle \\ &+\rho\Big\langle \boldsymbol{B}\boldsymbol{y}_{k+1}-\boldsymbol{B}\boldsymbol{y}_k,\, \boldsymbol{A}\boldsymbol{x}_{k+1}-\boldsymbol{A}\boldsymbol{x}\Big\rangle\end{aligned}$$

can be "understood" as

$$\Big\langle \partial f(\boldsymbol{x}_{k+1}),\, \boldsymbol{x}_{k+1}-\boldsymbol{x}\Big\rangle + \Big\langle \partial g(\boldsymbol{y}_{k+1}),\, \boldsymbol{y}_{k+1}-\boldsymbol{y}\Big\rangle = \begin{aligned}&\Big\langle -\boldsymbol{\lambda}_{k+1},\, \boldsymbol{A}\boldsymbol{x}_{k+1}-\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}_{k+1}-\boldsymbol{B}\boldsymbol{y}\Big\rangle \\ &+\rho\Big\langle \boldsymbol{B}\boldsymbol{y}_{k+1}-\boldsymbol{B}\boldsymbol{y}_k,\, \boldsymbol{A}\boldsymbol{x}_{k+1}-\boldsymbol{A}\boldsymbol{x}\Big\rangle\end{aligned} \qquad (*)$$

which is similar to the key inequality for proving gradient descent converges.

Remarks

- Lemma 1c $\not\Longrightarrow$ $(*)$ because lemma 1c is only one particular subgradient of $(*)$.
- If f,g are differentiable then lemma 1c $\Longrightarrow$ $(*)$ because now subdifferential is singleton. But for ADMM with minimum assumption we do not assume f,g are differentiable.

3. $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \rho(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c})$

4. $\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}^* + \boldsymbol{c} = \boldsymbol{0}$

$$\begin{aligned}&\left\langle -\boldsymbol{A}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}), \boldsymbol{x}_{k+1} - \boxed{\boldsymbol{x}}\right\rangle \\ +&\left\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}), \boldsymbol{y}_{k+1} - \boxed{\boldsymbol{y}}\right\rangle\end{aligned} = \begin{aligned}&\left\langle -\boldsymbol{\lambda}_{k+1}, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boxed{\boldsymbol{x}} + \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boxed{\boldsymbol{y}}\right\rangle \\ &+\rho\left\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boxed{\boldsymbol{x}}\right\rangle\end{aligned} \qquad \text{(Lemma 1c)}$$

Lemma 2 (Preparation lemma for telescoping)

$$\begin{aligned}&\left\langle -\boldsymbol{A}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}), \boldsymbol{x}_{k+1} - \boldsymbol{x}^*\right\rangle \\ +&\left\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}), \boldsymbol{y}_{k+1} - \boldsymbol{y}^*\right\rangle \\ &+\left\langle \boldsymbol{\lambda}^*, \boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}\right\rangle\end{aligned} \le \begin{aligned}&\tfrac{1}{2\rho}\|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}^*\|_2^2 - \tfrac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*\|_2^2 \\ &+\tfrac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_k - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 - \tfrac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 \\ &-\tfrac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\|_2^2 - \tfrac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k\|_2^2\end{aligned} \qquad \text{(Lemma 2)}$$

Proof. Put $(\boxed{\boldsymbol{x}}, \boxed{\boldsymbol{y}}) = (\boldsymbol{x}^*, \boldsymbol{y}^*)$ in lemma 1c, then add $\left\langle \boldsymbol{\lambda}^*, \boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}\right\rangle$:

$$\begin{aligned}
\text{LHS} &= -\left\langle \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*, \boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}\right\rangle + \rho\left\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x}^*\right\rangle \\
&= -\left\langle \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*, \frac{\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k}{\rho}\right\rangle + \rho\left\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k, \boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}^* + \boldsymbol{c}\right\rangle \\
&= -\frac{1}{\rho}\left\langle \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\right\rangle + \rho\left\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k, \boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c} + \boldsymbol{B}\boldsymbol{y}^* - \boldsymbol{B}\boldsymbol{y}_{k+1}\right\rangle \\
&= -\frac{1}{\rho}\left\langle \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\right\rangle + \left\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k, \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k + \rho(\boldsymbol{B}\boldsymbol{y}^* - \boldsymbol{B}\boldsymbol{y}_{k+1})\right\rangle \\
&= -\frac{1}{\rho}\left\langle \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\right\rangle + \left\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k, \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\right\rangle - \rho\left\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k, \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}^*\right\rangle \\
&= -\frac{1}{\rho}\left\langle \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*, \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\right\rangle + \left\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k, \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\right\rangle - \rho\left\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k, \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}^*\right\rangle \\
&= \frac{-1}{2\rho}\left(\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*\|_2^2 + \|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\|_2^2 - \|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}^*\|_2^2\right) \\
&\quad + -\frac{\rho}{2}\left(\|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k\|_2^2 + \|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 - \|\boldsymbol{B}\boldsymbol{y}_k - \boldsymbol{B}\boldsymbol{y}^*\|_2^2\right) + \left\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k, \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\right\rangle
\end{aligned}$$

(blue) and (green) are exactly the RHS of lemma2, we finish the proof by showing (red) ≤ 0.

Showing $\left\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k \,,\, \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k \right\rangle \leq 0$

$$\left\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}) \,,\, \boldsymbol{y}_{k+1} - \boldsymbol{y} \right\rangle = \left\langle -\boldsymbol{\lambda}_{k+1} \,,\, \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y} \right\rangle \qquad \text{(Lemma 1a)}$$

---

(1)	$\left\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}) \,,\, \boldsymbol{y}_{k+1} - \boldsymbol{y}_k \right\rangle = \left\langle -\boldsymbol{\lambda}_{k+1} \,,\, \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k \right\rangle$	$\boldsymbol{y} = \boldsymbol{y}^k$ in lemma 1a
(2)	$\left\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_{k-1} - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}) \,,\, \boldsymbol{y}_k - \boldsymbol{y} \right\rangle = \left\langle -\boldsymbol{\lambda}_k \,,\, \boldsymbol{B}\boldsymbol{y}_k - \boldsymbol{B}\boldsymbol{y} \right\rangle$	decrease k by 1 in lemma 1a
(3)	$\left\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_{k-1} - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}) \,,\, \boldsymbol{y}_k - \boldsymbol{y}_{k+1} \right\rangle = \left\langle -\boldsymbol{\lambda}_k \,,\, \boldsymbol{B}\boldsymbol{y}_k - \boldsymbol{B}\boldsymbol{y}_{k+1} \right\rangle$	$\boldsymbol{y} = \boldsymbol{y}^{k+1}$ in (2)

(1)+(3) gives

$$\begin{aligned}&\left\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}) \,,\, \boldsymbol{y}_{k+1} - \boldsymbol{y}_k \right\rangle \\ +&\left\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_{k-1} - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}) \,,\, \boldsymbol{y}_k - \boldsymbol{y}_{k+1} \right\rangle\end{aligned} = \left\langle -\boldsymbol{\lambda}_{k+1} \,,\, \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k \right\rangle + \left\langle -\boldsymbol{\lambda}_k \,,\, \boldsymbol{B}\boldsymbol{y}_k - \boldsymbol{B}\boldsymbol{y}_{k+1} \right\rangle.$$

Regroup (swap $\boldsymbol{y}_k - \boldsymbol{y}_{k+1}$ to $\boldsymbol{y}_{k+1} - \boldsymbol{y}_k$)

$$\left\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}) - \left(-\boldsymbol{B}^\top\boldsymbol{\lambda}_{k-1} - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c})\right) \,,\, \boldsymbol{y}_{k+1} - \boldsymbol{y}_k \right\rangle = \left\langle -\boldsymbol{\lambda}_{k+1} + \boldsymbol{\lambda}_k \,,\, \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k \right\rangle$$

We now have

$$\left\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}) - \left(-\boldsymbol{B}^\top\boldsymbol{\lambda}_{k-1} - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c})\right) , \boldsymbol{y}_{k+1} - \boldsymbol{y}_k \right\rangle + \left\langle \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k , \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k \right\rangle = 0.$$

Both [blue] and [green] are subgradients. Subgradient is monotone so the first inner product is ≥ 0. Hence [red] $\leq \mathbf{0}$. ■

$$\begin{array}{c}\left\langle -\boldsymbol{A}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}), \boldsymbol{x}_{k+1} - \boldsymbol{x}^*\right\rangle \\ + \left\langle -\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}), \boldsymbol{y}_{k+1} - \boldsymbol{y}^*\right\rangle \\ + \left\langle \boldsymbol{\lambda}^*, \boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}\right\rangle \end{array} \le \begin{array}{c} \frac{1}{2\rho}\|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}^*\|_2^2 - \frac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*\|_2^2 \\ + \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_k - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 - \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 \\ - \frac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\|_2^2 - \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k\|_2^2 \end{array} \quad \text{(Lemma 2)}$$

Lemma 3

- The ADMM iterate $(\boldsymbol{x}_{k+1}, \boldsymbol{y}_{k+1}, \boldsymbol{\lambda}_{k+1})$ satisfy

$$\begin{array}{c} f_{k+1} + g_{k+1} - f^* - g^* \\ + \langle \boldsymbol{\lambda}^*, \boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}\rangle \end{array} \le \begin{array}{c} \frac{1}{2\rho}\|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}^*\|_2^2 - \frac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*\|_2^2 \\ + \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_k - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 - \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 \\ - \frac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\|_2^2 - \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k\|_2^2 \end{array} \quad \text{(Lemma 3)}$$

- **Proof**. By subgradient convexity inequality of f and g:

$$f^* \ge f_{k+1} + \langle \partial f_{k+1}, \boldsymbol{x}^* - \boldsymbol{x}_{k+1}\rangle \iff f_{k+1} - f^* \le \langle \partial f_{k+1}, \boldsymbol{x}_{k+1} - \boldsymbol{x}^*\rangle$$

$$g^* \ge g_{k+1} + \langle \partial g_{k+1}, \boldsymbol{y}^* - \boldsymbol{y}_{k+1}\rangle \iff g_{k+1} - g^* \le \langle \partial g_{k+1}, \boldsymbol{y}_{k+1} - \boldsymbol{y}^*\rangle$$

- $-\boldsymbol{A}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}) \in \partial f_{k+1}$ and $-\boldsymbol{B}^\top\boldsymbol{\lambda}_k - \rho\boldsymbol{B}^\top(\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}) \in \partial g_{k+1}$ means lemma 2 $\implies$ lemma 3. ■

What does lemma 3 represent

$$\begin{array}{c} f_{k+1}+g_{k+1}-f^*-g^* \\ +\langle \boldsymbol{\lambda}^*, \boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c}\rangle \end{array} \leq \begin{array}{c} \frac{1}{2\rho}\|\boldsymbol{\lambda}_k-\boldsymbol{\lambda}^*\|_2^2-\frac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}^*\|_2^2 \\ +\frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_k-\boldsymbol{B}\boldsymbol{y}^*\|_2^2-\frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1}-\boldsymbol{B}\boldsymbol{y}^*\|_2^2. \\ -\frac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k\|_2^2-\frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1}-\boldsymbol{B}\boldsymbol{y}_k\|_2^2 \end{array} \qquad \text{(Lemma 3)}$$

- Focus on LHS

$$\begin{aligned} & f_{k+1}+g_{k+1}-f^*-g^*+\langle \boldsymbol{\lambda}^*, \boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c}\rangle \\ = \; & \underbrace{f_{k+1}+g_{k+1}+\langle \boldsymbol{\lambda}^*, \boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c}\rangle}_{=\mathcal{L}(\boldsymbol{x}_{k+1},\boldsymbol{y}_{k+1},\boldsymbol{\lambda}^*)} \underbrace{-f^*-g^*-\langle \boldsymbol{\lambda}^*, \overbrace{\boldsymbol{A}\boldsymbol{x}^*+\boldsymbol{B}\boldsymbol{y}^*+\boldsymbol{c}}^{=\boldsymbol{0}}\rangle}_{=\mathcal{L}(\boldsymbol{x}^*,\boldsymbol{y}^*,\boldsymbol{\lambda}^*)} \\ = \; & \mathcal{L}(\boldsymbol{x}_{k+1},\boldsymbol{y}_{k+1},\boldsymbol{\lambda}^*)-\mathcal{L}(\boldsymbol{x}^*,\boldsymbol{y}^*,\boldsymbol{\lambda}^*) \end{aligned}$$

- Thus lemma 3 bounds the optimality gap, in terms of the un-augmented Lagrangian, between $(\boldsymbol{x}_{k+1},\boldsymbol{y}_{k+1},\boldsymbol{\lambda}^*)$ and $(\boldsymbol{x}^*,\boldsymbol{y}^*,\boldsymbol{\lambda}^*)$.
- The gap only focus on the primal variables $\boldsymbol{x},\boldsymbol{y}$ and ignore the dual variable $\boldsymbol{\lambda}$.
- As KKT points is a saddle point, we have $\mathcal{L}(\boldsymbol{x}_{k+1},\boldsymbol{y}_{k+1},\boldsymbol{\lambda}^*)-\mathcal{L}(\boldsymbol{x}^*,\boldsymbol{y}^*,\boldsymbol{\lambda}^*)\geq 0$.

Convergence of ADMM: primal feasibility ... 1/2

- **Theorem** The ADMM sequence $(\boldsymbol{x}_{k+1}, \boldsymbol{y}_{k+1})$ satisfies $\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c} \to \boldsymbol{0}$.
- **Proof**. Applying saddle point property $\mathcal{L}(\boldsymbol{x}_{k+1}, \boldsymbol{y}_{k+1}, \boldsymbol{\lambda}^*) - \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\lambda}^*) \geq 0$ in lemma 3, we have

$$0 \leq \begin{array}{c} f_{k+1} + g_{k+1} - f^* - g^* \\ + \langle \boldsymbol{\lambda}^*, \boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c} \rangle \end{array} \leq \begin{array}{c} \frac{1}{2\rho}\|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}^*\|_2^2 - \frac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*\|_2^2 \\ + \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_k - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 - \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 \\ - \frac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\|_2^2 - \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k\|_2^2 \end{array}. \qquad \text{(Lemma 3)}$$

Thus we have

$$0 \leq \frac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\|_2^2 + \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k\|_2^2 \leq \begin{array}{c} \frac{1}{2\rho}\|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}^*\|_2^2 + \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_k - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 \\ - \frac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*\|_2^2 - \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 \end{array}. \qquad (*)$$

- From $(*)$ we have

$$0 \leq \begin{array}{c} \frac{1}{2\rho}\|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}^*\|_2^2 + \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_k - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 \\ - \frac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*\|_2^2 - \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 \end{array}$$

$$\iff \quad 0 \leq \frac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}^*\|_2^2 + \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}^*\|_2^2 \leq \frac{1}{2\rho}\|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}^*\|_2^2 + \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_k - \boldsymbol{B}\boldsymbol{y}^*\|_2^2,$$

meaning that the sequence $\frac{1}{2\rho}\|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}^*\|_2^2 + \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_k - \boldsymbol{B}\boldsymbol{y}^*\|_2^2$ is bounded below and non-increasing.

  - So $\|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}^*\|_2^2$ and $\boxed{\|\boldsymbol{B}\boldsymbol{y}_k - \boldsymbol{B}\boldsymbol{y}^*\|_2^2}$ are bounded for all k.
  - $\|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}^*\|_2^2$ is bounded implies $\|\boldsymbol{\lambda}_k\|_2^2$ is bounded.

Convergence of ADMM: primal feasibility ... 2/2

- Sum $(*)$ from $k=0$ to $k=\infty$ with $\boldsymbol{\lambda}_\infty = \boldsymbol{\lambda}^*$ and $\boldsymbol{y}_\infty = \boldsymbol{y}^*$,

$$\sum_{k=0}^{\infty}\left(\frac{1}{2\rho}\|\boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k\|_2^2 + \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_{k+1}-\boldsymbol{B}\boldsymbol{y}_k\|_2^2\right) \leq \frac{1}{2\rho}\|\boldsymbol{\lambda}_0-\boldsymbol{\lambda}^*\|_2^2 + \frac{\rho}{2}\|\boldsymbol{B}\boldsymbol{y}_0-\boldsymbol{B}\boldsymbol{y}^*\|_2^2.$$

  So we have $\boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k \to \boldsymbol{0}$ and $\boldsymbol{B}\boldsymbol{y}_{k+1}-\boldsymbol{B}\boldsymbol{y}_k \to \boldsymbol{0}$.

- Now consider the ADMM iteration $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \rho(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c})$.
  1. We have $\boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k = \rho(\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c})$.
     By $\boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k \to \boldsymbol{0}$ so $\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c} \to \boldsymbol{0}$.
  2. We have $\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c} \overset{\boldsymbol{A}\boldsymbol{x}^*+\boldsymbol{B}\boldsymbol{y}^*+\boldsymbol{c}=\boldsymbol{0}}{=} (\boldsymbol{A}\boldsymbol{x}_{k+1}-\boldsymbol{A}\boldsymbol{x}^*)+(\boldsymbol{B}\boldsymbol{y}_{k+1}-\boldsymbol{B}\boldsymbol{y}^*)$.
     By $\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c} \to \boldsymbol{0}$ and $\|\boldsymbol{B}\boldsymbol{y}_k-\boldsymbol{B}\boldsymbol{y}^*\|_2^2$ is bounded we have $\boldsymbol{A}\boldsymbol{x}_{k+1}-\boldsymbol{A}\boldsymbol{x}^*$ is bounded. ■

- Remarks: we have proved
  - convergence in primal feasibility $\boldsymbol{A}\boldsymbol{x}_{k+1}+\boldsymbol{B}\boldsymbol{y}_{k+1}+\boldsymbol{c} \to \boldsymbol{0}$
  - convergence in dual variable fixed-point $\boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k \to \boldsymbol{0}$
  - convergence in primal variable fixed-point $\boldsymbol{B}\boldsymbol{y}_{k+1}-\boldsymbol{B}\boldsymbol{y}_k \to \boldsymbol{0}$
  - $\|\boldsymbol{B}\boldsymbol{y}_k-\boldsymbol{B}\boldsymbol{y}^*\|_2^2$ and $\boldsymbol{A}\boldsymbol{x}_{k+1}-\boldsymbol{A}\boldsymbol{x}^*$ is bounded for all k

Convergence of ADMM: primal function value convergence ... 1/2

- **Theorem** ADMM iterate $(\boldsymbol{x}_{k+1}, \boldsymbol{y}_{k+1})$ satisfies $f(\boldsymbol{x}_{k+1}) + g(\boldsymbol{y}_{k+1}) \to f(\boldsymbol{x}^*) + g(\boldsymbol{y}^*)$.
- By convexity of f, g

$$
\begin{aligned}
f^* &\geq f_{k+1} + \langle \partial f_{k+1}, \boldsymbol{x}^* - \boldsymbol{x}_{k+1} \rangle && f \text{ is convex} \\
g^* &\geq g_{k+1} + \langle \partial g_{k+1}, \boldsymbol{y}^* - \boldsymbol{y}_{k+1} \rangle && g \text{ is convex} \\
f^* + g^* &\geq f_{k+1} + \langle \partial f_{k+1}, \boldsymbol{x}^* - \boldsymbol{x}_{k+1} \rangle + g_{k+1} + \langle \partial g_{k+1}, \boldsymbol{y}^* - \boldsymbol{y}_{k+1} \rangle && \text{sum} \\
f_{k+1} + g_{k+1} - f^* - g^* &\leq \langle \partial f_{k+1}, \boldsymbol{x}_{k+1} - \boldsymbol{x}^* \rangle + \langle \partial g_{k+1}, \boldsymbol{y}_{k+1} - \boldsymbol{y}^* \rangle && \text{rearrange}
\end{aligned}
$$

- As $-\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}) \in \partial f_{k+1}$ and $-\boldsymbol{B}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{B}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}) \in \partial g_{k+1}$,

$$
f^* + g^* - f_{k+1} - g_{k+1} \leq \begin{array}{c} \langle -\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}), \boldsymbol{x}_{k+1} - \boldsymbol{x}^* \rangle \\ + \langle -\boldsymbol{B}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{B}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}), \boldsymbol{y}_{k+1} - \boldsymbol{y}^* \rangle \end{array}
$$

- Apply lemma 1c

$$
\begin{array}{c} \Big\langle -\boldsymbol{A}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_k + \boldsymbol{c}) \, , \, \boldsymbol{x}_{k+1} - \boldsymbol{x} \Big\rangle \\ + \Big\langle -\boldsymbol{B}^\top \boldsymbol{\lambda}_k - \rho \boldsymbol{B}^\top (\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c}), \boldsymbol{y}_{k+1} - \boldsymbol{y} \Big\rangle \end{array} = \begin{array}{c} \Big\langle -\boldsymbol{\lambda}_{k+1} \, , \, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y} \Big\rangle \\ + \rho \Big\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k \, , \, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x} \Big\rangle \end{array} \quad \text{(Lemma 1c)}
$$

we have

$$
f^* + g^* - f_{k+1} - g_{k+1} \leq \begin{array}{c} \Big\langle -\boldsymbol{\lambda}_{k+1} \, , \, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}^* \Big\rangle \\ + \rho \Big\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k \, , \, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x}^* \Big\rangle \end{array} = \begin{array}{c} \Big\langle -\boldsymbol{\lambda}_{k+1} \, , \, \boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c} \Big\rangle \\ + \rho \Big\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k \, , \, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x}^* \Big\rangle \end{array}
$$

- We now have

$$
f^* + g^* - f_{k+1} - g_{k+1} \leq \Big\langle -\boldsymbol{\lambda}_{k+1} \, , \, \boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c} \Big\rangle + \rho \Big\langle \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k \, , \, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x}^* \Big\rangle \to 0,
$$

by $\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c} \to \boldsymbol{0}$, $\boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}_k \to \boldsymbol{0}$ and $\boxed{\boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x}^*}$ is bounded.

Convergence of ADMM: primal function value convergence ... 2/2

- By convexity of f, g

$$\begin{aligned} f_{k+1} &\geq f^* + \langle \partial f^*, \boldsymbol{x}_{k+1} - \boldsymbol{x}^* \rangle && f \text{ is convex} \\ g_{k+1} &\geq g^* + \langle \partial g^*, \boldsymbol{y}_{k+1} - \boldsymbol{y}^* \rangle && g \text{ is convex} \\ f_{k+1} + g_{k+1} &\geq f^* + g^* + \langle \partial f^*, \boldsymbol{x}_{k+1} - \boldsymbol{x}^* \rangle + \langle \partial g^*, \boldsymbol{y}_{k+1} - \boldsymbol{y}^* \rangle && \text{sum} \\ f_{k+1} + g_{k+1} - f^* - g^* &\geq \langle \partial f^*, \boldsymbol{x}_{k+1} - \boldsymbol{x}^* \rangle + \langle \partial g^*, \boldsymbol{y}_{k+1} - \boldsymbol{y}^* \rangle && \text{rearrange} \end{aligned}$$

- Subgradient of f and g at convergence: $-\boldsymbol{A}^\top \boldsymbol{\lambda}^* \in \partial f^*$, $-\boldsymbol{B}^\top \boldsymbol{\lambda}^* \in \partial g^*$

$$\begin{aligned} f_{k+1} + g_{k+1} - f^* - g^* &\geq \langle -\boldsymbol{A}^\top \boldsymbol{\lambda}^*, \boldsymbol{x}_{k+1} - \boldsymbol{x}^* \rangle + \langle -\boldsymbol{B}^\top \boldsymbol{\lambda}^*, \boldsymbol{y}_{k+1} - \boldsymbol{y}^* \rangle \\ &= \langle -\boldsymbol{\lambda}^*, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x}^* \rangle + \langle -\boldsymbol{\lambda}^*, \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}^* \rangle \\ &= \langle -\boldsymbol{\lambda}^*, \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}_{k+1} - \boldsymbol{B}\boldsymbol{y}^* \rangle \\ &\overset{\boldsymbol{A}\boldsymbol{x}^*+\boldsymbol{B}\boldsymbol{y}^*+\boldsymbol{c}=\boldsymbol{0}}{=} \langle -\boldsymbol{\lambda}^*, \boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c} \rangle \end{aligned}$$

So we have

$$f_{k+1} + g_{k+1} - f^* - g^* \geq -\langle \boldsymbol{\lambda}^*, \boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c} \rangle \to 0$$

by $\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{y}_{k+1} + \boldsymbol{c} \to \boldsymbol{0}$ and $\|\boldsymbol{\lambda}_k\|_2^2$ is bounded for all k.

- We have

$$0 \leq f_{k+1} + g_{k+1} - f^* - g^* \leq 0.$$

By squeeze theorem, $f_{k+1} + g_{k+1} - f^* - g^* \to 0$.

End of document