Linear Rate Convergence of the ADMM for Multi-Block Convex Conic Programming

Defeng Sun

National University of Singapore

August 11, 2016

Joint work with: Deren Han (Nanjing) and Liwei Zhang (Dalian)

$\mathcal{X}$: real finite dimensional Euclidean space endowed inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. Consider a maximal monotone operator $\mathcal{T} : \mathcal{X} \to \mathcal{X}$.

Solve the following inclusion problem: $0 \in \mathcal{T}(z)$

Given $c > 0$, the proximal mapping associated with $c\mathcal{T}$

$$P := (\mathcal{I} + c\mathcal{T})^{-1}$$

The proximal point algorithm (PPA):

$$z^{k+1} \approx P_k(z^k), \quad P_k = (\mathcal{I} + c_k\mathcal{T})^{-1}$$

Criterion for approximate calculation of $P_k(z^k)$:

$$(\mathrm{A}) : \quad \|z^{k+1} - P_k(z^k)\| \leq \delta_k \|z^{k+1} - z^k\|, \qquad \sum_{k=0}^{\infty} \delta_k < \infty$$

Let $\overline{Z} := \{z \in \mathcal{X} \mid 0 \in \mathcal{T}(z)\} \neq \emptyset$.

Error bound condition for $\mathcal{T}$: $\exists\, a > 0, \tau > 0$

$$\text{dist}(z, \overline{Z}) \leq a\|w\|, \quad \forall z \in \mathcal{T}^{-1}(w),\ \|w\| \leq \tau$$

Theorem 1 (Luque 1984, based on Rockafellar 1976)

Let z^k be generated by PPA using criterion (A) with c_k nondecreasing ($c_k \uparrow c_\infty \leq +\infty$). Suppose that the above error bound condition holds for $\mathcal{T}$. Then,

- *$\text{dist}(z^k, \overline{Z}) \to 0$ linearly with a rate bounded from above by*

$$\frac{a}{\sqrt{a^2 + c_\infty^2}} < 1 \qquad \text{(fast linear convergence)}$$

- *If $c_\infty = +\infty$, the convergence is superlinear.*

Consider

$$\begin{aligned} \min \quad & f(u) + g(v) \\ \text{s.t.} \quad & \mathcal{F}^* u + \mathcal{G}^* v = c. \end{aligned}$$

Closed proper convex functions: $f : \mathcal{U} \to (-\infty, +\infty]$, $g : \mathcal{V} \to (-\infty, +\infty]$

Linear maps $\mathcal{F} : \mathcal{X} \to \mathcal{U}$, $\mathcal{G} : \mathcal{X} \to \mathcal{V}$

Given $\sigma > 0$, the augmented Lagrangian function:

$$\mathcal{L}_\sigma(u, v; x) = f(u) + g(v) + \langle x, \, \mathcal{F}^* u + \mathcal{G}^* v - c \rangle + \frac{\sigma}{2} \|\mathcal{F}^* u + \mathcal{G}^* v - c\|^2.$$

Its dual:

$$\max \left\{ -\langle c, \, x \rangle - f^*(-\mathcal{F}x) - g^*(-\mathcal{G}x) \right\}$$

Let $A(x) := c - \mathcal{F}^* \partial f^*(-\mathcal{F}x), \quad B(x) := -\mathcal{G}^* \partial g^*(-\mathcal{G}x)$.

Algorithm ADMM(τ): An ADMM with step-length τ

For $k = 0, 1, 2, ...,$ iterate

1. Compute
$$u^{k+1} = \arg\min_u \mathcal{L}_\sigma(u, v^k; x^k)$$
2. Compute
$$v^{k+1} = \arg\min_v \mathcal{L}_\sigma(u^{k+1}, v; x^k)$$
3. Compute $x^{k+1} = x^k + \tau\sigma(\mathcal{F}^* u^{k+1} + \mathcal{G}^* v^{k+1} - c)$.

Consider a general inclusion model

$$0 \in (A+B)(z), \tag{1}$$

where $A, B : \mathcal{Z} \rightrightarrows \mathcal{Z}$ are maximal monotone operators. Given $\sigma > 0$, let

$$J_{\sigma A}(z) := (I+\sigma A)^{-1}(z), \quad J_{\sigma B}(z) := (I+\sigma B)^{-1}(z)$$

be the resolvents of σA and σB, respectively. Let $\{t^k\}_{k=0}^{\infty}$ be the sequence generated by the following Douglas-Rachford splitting (DR-splitting) method

$$t^{k+1} = (J_{\sigma A} \circ (2J_{\sigma B} - I) + (I - J_{\sigma B}))t^k. \tag{2}$$

Assume that the KKT solution set is nonempty.

• Gabay (1983) showed that ADMM(1) with $\mathcal{F}\mathcal{F}^* \succ 0$ and $\mathcal{G}\mathcal{G}^* \succ 0$ can be viewed as a DR-splitting method applied to $0 \in (A + B)(x)$

• Eckstein (PhD Thesis, 1989) showed that the DR-splitting is in fact a PPA corresponding to the following maximal monotone operator:

$$\mathcal{S}_{A,B} := \{(y + \sigma b, z - y) \,|\, b \in B(z),\ a \in A(y),\ y + \sigma a = z - \sigma b\}$$

Caution: ADMM(1) is PPA. But in general ADMM(τ) is not PPA.

Theorem 2 (J. Sun, PhD Thesis (1986))

Let f be a closed proper convex function. Then f is piecewise linear-quadratic iff the graph of ∂f is piecewise polyhedral. Moreover, f is piecewise linear-quadratic iff f^* (Fenchel conjugate function) is piecewise linear-quadratic.

Theorem 3 (Robinson 1981)

If the multi-valued mapping $F : \mathcal{X} \rightrightarrows \mathcal{Y}$ is piecewise polyhedral, then F is calm at x^0, i.e., $\exists\, \kappa_0 > 0$ and a neighborhood V of x^0 such that

$$F(x) \subseteq F(x^0) + \kappa_0 \|x - x^0\| \mathbf{B}_{\mathcal{Y}}, \quad \forall\, x \in V.$$

Thus, the error bound condition holds for $\mathcal{S}_{A,B}$ corresponding to the convex piecewise quadratic programming (QP).

Linear convergence rate of ADMM(1) for convex piecewise QP:

(1). ADMM(1) with $\mathcal{F}\mathcal{F}^* \succ 0$ and $\mathcal{G}\mathcal{G}^* \succ 0$ is a PPA corresponding to $\mathcal{S}_{A,B}$.

(2). The error bound condition holds for the corresponding maximal monotone operator $\mathcal{S}_{A,B}$ [Robinson + J. Sun].

- (1) + (2) $\Longrightarrow$ Linear convergence rate.

All the above are essentially known by 1989. There are several recent papers with more direct proofs ...

Convex composite optimization

The convex composite optimization problem

$$\min_{y\in\mathcal{Y},\, z\in\mathcal{Z}} \{\underline{\vartheta(y) + g(y)} + \underline{\varphi(z) + h(z)} : \mathcal{A}^*y + \mathcal{B}^*z = c\},$$

where $\mathcal{Y}$ and $\mathcal{Z}$ are two finite-dimensional real Euclidean spaces each equipped with an inner product $\langle\cdot,\cdot\rangle$ and its induced norm $\|\cdot\|$.

- $\vartheta : \mathcal{Y} \to (-\infty, +\infty]$, $\varphi : \mathcal{Z} \to (-\infty, +\infty]$ proper closed convex functions
- $g : \mathcal{Y} \to (-\infty, +\infty)$ and $h : \mathcal{Z} \to (-\infty, +\infty)$ are two C^1 convex functions (e.g., convex quadratic functions)

The augmented Lagrangian function

- Write $\vartheta_g(\cdot) \equiv \vartheta(\cdot) + g(\cdot)$ and $\varphi_h(\cdot) \equiv \varphi(\cdot) + h(\cdot)$.
- The augmented Lagrangian function is defined by

$$\begin{aligned}\mathcal{L}_\sigma(y, z; x) := \quad & \vartheta_g(y) + \varphi_h(z) + \langle x, \mathcal{A}^* y + \mathcal{B}^* z - c\rangle \\ & + \frac{\sigma}{2}\|\mathcal{A}^* y + \mathcal{B}^* z - c\|^2, \\ & \forall\, (y, z, x) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}.\end{aligned}$$

Step 0. Input $(y^0, z^0, x^0) \in \mathrm{dom}\,\vartheta \times \mathrm{dom}\,\varphi \times \mathcal{X}$. Let $\tau \in (0, \infty)$ be a positive parameter (e.g., $\tau \in (0, (1+\sqrt{5})/2)$), and $\mathcal{S} : \mathcal{Y} \to \mathcal{Y}$ and $\mathcal{T} : \mathcal{Z} \to \mathcal{Z}$ be two self-adjoint positive semi-definite, not necessarily positive definite, linear operators. Set $k := 0$.

Step 1. Set

$$\begin{aligned} y^{k+1} &\in \arg\min \, \mathcal{L}_\sigma(y, z^k; x^k) + \tfrac{1}{2}\|y - y^k\|_{\mathcal{S}}^2 \,, \\ z^{k+1} &\in \arg\min \, \mathcal{L}_\sigma(y^{k+1}, z; x^k) + \tfrac{1}{2}\|z - z^k\|_{\mathcal{T}}^2 \,, \\ x^{k+1} &= x^k + \tau\sigma(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^{k+1} - c). \end{aligned} \tag{3}$$

Step 2. If a termination criterion is not met, set $k := k + 1$ and go to Step 1.

Assumption 1

The KKT system has a non-empty solution set.

Denote $u := (y, z, x)$ for $y \in \mathcal{Y}$, $z \in \mathcal{Z}$ and $x \in \mathcal{X}$. Let

$$\mathcal{U} := \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}.$$

Define the KKT mapping $R : \mathcal{U} \to \mathcal{U}$ as

$$R(u) := \begin{pmatrix} y - \mathrm{Pr}_{\vartheta}[y - (\nabla g(y) + \mathcal{A}x)] \\ z - \mathrm{Pr}_{\varphi}[z - (\nabla h(z) + \mathcal{B}x)] \\ c - \mathcal{A}^* y - \mathcal{B}^* z \end{pmatrix}, \quad \forall\, u \in \mathcal{U},$$

where $\mathrm{Pr}_{\theta}(\cdot)$ denotes the Moreau-Yosida proximal mapping. The mapping $R(\cdot)$ is at least continuous on $\mathcal{U}$ and

$$\forall\, u \in \mathcal{U}, \quad R(u) = 0 \Longleftrightarrow u \in \overline{\Omega}.$$

Global convergence of sPADMM

The global convergence of the sPADMM is established in Appendix B of Fazel-Pong-S.-Tseng (2013). For the iteration complexity on sPADMM(1), see Shefi and Teboulle (2014).

There exist two self-adjoint and positive semi-definite linear operators (could be zero operators) Σ_g and Σ_h such that for all $y', y \in \operatorname{dom}\vartheta$, and for all $z', z \in \operatorname{dom}\varphi$,

$$\langle \nabla g(y') - \nabla g(y), y' - y \rangle \geq \|y' - y\|_{\Sigma_g}^2,$$
$$\langle \nabla h(z') - \nabla h(z), z' - z \rangle \geq \|z' - z\|_{\Sigma_h}^2.$$

Linear convergence rate of sPADMM

For any $\tau \in (0, \infty)$, define

$$s_\tau := \frac{5 - \tau - 3\min\{\tau, \tau^{-1}\}}{4} \quad \& \quad t_\tau := \frac{1 - \tau + \min\{\tau, \tau^{-1}\}}{8}.$$

$$1/4 \leq s_\tau \leq 5/4 \quad \& \quad 0 < t_\tau \leq 1/8, \quad \forall\, \tau \in (0, (1+\sqrt{5})/2).$$

Let $\mathcal{E} : \mathcal{X} \to \mathcal{U} := \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$ be a linear operator such that its adjoint $\mathcal{E}^*$ satisfies

$$\mathcal{E}^*(y, z, x) = \mathcal{A}^* y + \mathcal{B}^* z$$

for any $(y, z, x) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$.

Denote

$$\mathcal{M} := \mathrm{Diag}\left(\mathcal{S} + \Sigma_g, \mathcal{T} + \Sigma_h + \sigma\mathcal{B}\mathcal{B}^*, (\tau\sigma)^{-1}\mathcal{I}\right) + s_\tau\sigma\mathcal{E}\mathcal{E}^*$$

$$\mathcal{H} := \mathrm{Diag}\left(\mathcal{S} + \frac{1}{2}\Sigma_g, \mathcal{T} + \frac{1}{2}\Sigma_h + \tau\sigma\mathcal{B}\mathcal{B}^*, 4t_\tau(\tau^2\sigma)^{-1}\mathcal{I}\right) + t_\tau\sigma\mathcal{E}\mathcal{E}^*.$$

Proposition 1

Let $\tau \in (0, (1+\sqrt{5})/2)$. *Then*

$$\Sigma_g + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \succ 0 \ \& \ \Sigma_h + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* \succ 0$$

$$\Updownarrow$$

$$\mathcal{M} \succ 0 \Longleftrightarrow \mathcal{H} \succ 0.$$

Proposition 2

Let $\tau \in (0, (1+\sqrt{5})/2)$ *and* $\{(y^k, z^k, x^k)\}$ *be an infinite sequence generated by the sPADMM. Then for any* $\bar{u} = (\bar{y}, \bar{z}, \bar{x}) \in \overline{\Omega}$ *and any* $k \geq 1$*,*

$$\begin{aligned} &\|u^{k+1} - \bar{u}\|_{\mathcal{M}}^2 + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 \\ &\leq \left(\|u^k - \bar{u}\|_{\mathcal{M}}^2 + \|z^k - z^{k-1}\|_{\mathcal{T}}^2\right) - \|u^{k+1} - u^k\|_{\mathcal{H}}^2. \end{aligned} \tag{1}$$

Consequently, we have for all $k \geq 1$*,*

$$\begin{aligned} &\mathrm{dist}_{\mathcal{M}}^2(u^{k+1}, \overline{\Omega}) + \|z^{k+1} - z^k\|_{\mathcal{T}}^2 \\ &\leq \left(\mathrm{dist}_{\mathcal{M}}^2(u^k, \overline{\Omega}) + \|z^k - z^{k-1}\|_{\mathcal{T}}^2\right) - \|u^{k+1} - u^k\|_{\mathcal{H}}^2. \end{aligned} \tag{2}$$

For establishing the linear rate of convergence of the sPADMM, we need the following error bound condition with respect to $\bar{u} \in \overline{\Omega}$.

Assumption 2

For some given $\bar{u} \in \overline{\Omega}$, there exist positive constants δ and $\eta > 0$ such that

$$\operatorname{dist}(u, \overline{\Omega}) \leq \eta \|R(u)\|, \quad \forall\, u \in \{u \in \mathcal{U} : \|u - \bar{u}\| \leq \delta\}. \tag{3}$$

Theorem 1

Let $\tau \in (0, (1+\sqrt{5})/2)$. *Suppose that Assumptions 1 and 2 hold. Assume* $\Sigma_g + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \succ 0$ & $\Sigma_h + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* \succ 0$. *Then for all* k *sufficiently large,*

$$\begin{aligned} &\operatorname{dist}^2_{\mathcal{M}}(u^{k+1}, \overline{\Omega}) + \|z^{k+1} - z^k\|^2_{\mathcal{T}} \\ &\leq \mu \left[\operatorname{dist}^2_{\mathcal{M}}(u^k, \overline{\Omega}) + \|z^k - z^{k-1}\|^2_{\mathcal{T}}\right], \end{aligned} \tag{4}$$

where $\mu := (1 + 2\kappa_4)^{-1}(1 + \kappa_4) < 1$ *with*

$$\kappa_4 := \min\{\tau, 4t_\tau\} \left(\eta^2 \kappa \lambda_{\max}(\mathcal{M})\right)^{-1} > 0,$$

$$\kappa := \max\{\kappa_1, \kappa_2, \kappa_3\},$$

$$\kappa_1 := 3\|\mathcal{S}\|, \quad \kappa_2 := \max\{3\sigma\lambda_{\max}(\mathcal{A}\mathcal{A}^*), 2\|\mathcal{T}\|\},$$

$$\kappa_3 := 3(1-\tau)^2\sigma\lambda_{\max}(\mathcal{A}\mathcal{A}^*) + 2(1-\tau)^2\sigma\lambda_{\max}(\mathcal{B}\mathcal{B}^*) + \sigma^{-1}.$$

(cont.) Moreover, there exists a positive number $\varsigma \in [\mu, 1)$ such that for all $k \geq 1$,

$$\begin{aligned} &\operatorname{dist}^2_{\mathcal{M}}(u^{k+1}, \overline{\Omega}) + \|z^{k+1} - z^k\|^2_{\mathcal{T}} \\ &\leq \varsigma \left[\operatorname{dist}^2_{\mathcal{M}}(u^k, \overline{\Omega}) + \|z^k - z^{k-1}\|^2_{\mathcal{T}}\right]. \end{aligned} \tag{5}$$

Corollary 2

Let $\tau \in (0, (1+\sqrt{5})/2)$. *Suppose that* $\overline{\Omega} \neq \emptyset$ *and that* $\Sigma_g + \mathcal{S} + \sigma \mathcal{A}\mathcal{A}^* \succ 0$ & $\Sigma_h + \mathcal{T} + \sigma \mathcal{B}\mathcal{B}^* \succ 0$. *Assume that the mapping* $R : \mathcal{U} \to \mathcal{U}$ *is piecewise polyhedral. Then there exists a constant* $\varsigma \in (0,1)$ *such that*

$$\begin{aligned} &\operatorname{dist}^2_{\mathcal{M}}(u^{k+1}, \overline{\Omega}) + \|z^{k+1} - z^k\|^2_{\mathcal{T}} \\ &\leq \varsigma \left[\operatorname{dist}^2_{\mathcal{M}}(u^k, \overline{\Omega}) + \|z^k - z^{k-1}\|^2_{\mathcal{T}}\right], \quad \forall\, k \geq 1. \end{aligned}$$

If [$\mathcal{S} \succ 0$ and $\mathcal{T} \succ 0$] or if [one of them is positive definite and the other is zero], one can use PPA or Ha's partial PPA (1990) to derive the linear convergence with $\tau = 1$ though in forms different from the above.

The convex composite quadratic conic programming

$$
\begin{array}{ll}
\min & \frac{1}{2}\langle x, \mathcal{Q}x\rangle + \langle c, x\rangle + \phi(x) \\
\text{s.t.} & \mathcal{A}x = b, \quad x \in \mathcal{K},
\end{array} \tag{6}
$$

where $c \in \mathcal{X}$, $b \in \Re^m$, $\mathcal{Q} : \mathcal{X} \to \mathcal{X}$ is a self-adjoint positive semi-definite linear operator, $\mathcal{A} : \mathcal{X} \to \Re^m$ is a linear operator which is surjective, $\mathcal{K}$ is a closed convex cone in $\mathcal{X}$ and $\phi : \mathcal{X} \in (-\infty, \infty]$ is a proper closed convex function whose epigraph is convex polyhedral, i.e., ϕ is a closed proper convex polyhedral function.

The Lagrange dual of problem (6) takes the form of

$$\begin{array}{ll} \max & \inf_{x\in\mathcal{X}} \left\{ \frac{1}{2}\langle x, \mathcal{Q}x\rangle + \langle v, x\rangle \right\} + \langle b, y\rangle - \phi^*(-z) \\ \text{s.t.} & s + \mathcal{A}^* y + v + z = c, \quad s \in \mathcal{K}^*, \end{array}$$

which is equivalent to

$$\begin{array}{ll} \min & \delta_{\mathcal{K}^*}(s) - \langle b, y\rangle + \frac{1}{2}\langle w, \mathcal{Q}w\rangle + \phi^*(-z) \\ \text{s.t.} & s + \mathcal{A}^* y - \mathcal{Q}w + z = c, \quad w \in \mathcal{W} := \text{Range}\,\mathcal{Q}. \end{array} \tag{7}$$

One may call problem (7) the restricted Wolfe dual to problem (6).

sGS-ADMM: For (7).

Step 0. Input $(s^0, y^0, w^0, z^0, x^0) \in \mathcal{K}^* \times \Re^m \times \mathcal{W} \times (-\text{dom}\,\phi^*) \times \mathcal{X}$. Let $\tau \in (0, \infty)$ be a positive parameter (e.g., $\tau \in (0, (1+\sqrt{5})/2)$). Set $k := 0$.

Step 1. Set

$$\begin{cases} w^{k+\frac{1}{2}} = \arg\min \mathcal{L}_\sigma(s^k, y^k, w, z^k; x^k) \\ y^{k+\frac{1}{2}} = \arg\min \mathcal{L}_\sigma(s^k, y, w^{k+\frac{1}{2}}, z^k; x^k) \\ s^{k+1} = \arg\min \mathcal{L}_\sigma(s, y^{k+\frac{1}{2}}, w^{k+\frac{1}{2}}, z^k; x^k) \\ y^{k+1} = \arg\min \mathcal{L}_\sigma(s^{k+1}, y, w^{k+\frac{1}{2}}, z^k; x^k) \\ w^{k+1} = \arg\min \mathcal{L}_\sigma(s^{k+1}, y^{k+1}, w, z^k; x^k) \\ z^{k+1} = \arg\min \mathcal{L}_\sigma(s^{k+1}, y^{k+1}, w^{k+1}, z; x^k) \\ x^{k+1} = x^k + \tau\sigma(s^{k+1} + \mathcal{A}^* y^{k+1} - \mathcal{Q} w^{k+1} + z^{k+1} - c). \end{cases}$$

Step 2. If a termination criterion is not met, set $k := k + 1$ and go to Step 1.

The global convergence of Algorithm sGS-ADMM is to convert it into an equivalent sPADMM scheme (3) with $\mathcal{S} \succeq 0$ but $\mathcal{S} \not\succ 0$ and $\mathcal{T} = 0$.

By using the same connection, one can use Theorem 1 to derive the linear rate convergence of the infinite sequence $\{(s^k, y^k, w^k, z^k, x^k)\}$ generated by Algorithm sGS-ADMM if Assumptions 1 and 2 hold for problem (7) and $\tau \in (0, (1+\sqrt{5})/2)$. Assumption 2 holds automatically if $\mathcal{K}$ is convex polyhedral.

Error bounds for convex quadratic SDP

For convex quadratic SDP, the error bound condition is valid if the second order sufficient conditions for both the primal and dual problems hold (almost best possible), one of 8 equivalent conditions.

For multi-block SDPs, the sGS-ADMM is not only convergent but also is more efficient than the naive direct extension. This is different from others.

Essentially, a desirable ADMM for many core multi-block convex optimization problems has been designed and analyzed (using dual).

Reference:

Deren Han, Defeng Sun, Liwei Zhang, "Linear Rate Convergence of the Alternating Direction Method of Multipliers for Convex Composite Quadratic and Semi-Definite Programming". arXiv:1508.02134