Error bounds and the superlinear convergence rates of the augmented Lagrangian methods

Defeng Sun

The Hong Kong Polytechnic University

Joint work with Ying Cui and Kim-Chuan Toh

Error bounds

Given two subsets S and T and a nonnegative valued residual function $r : S \cup T \to \mathbb{R}^+$ satisfies

$$r(x) = 0 \Longleftrightarrow x \in S, \quad \forall\, x \in T.$$

An **error bound** of the pair (S, T) in terms of $r(\cdot)$ is of the form

$$\mathrm{dist}(x, S) \leq \underbrace{c\, r(x)^{\rho}}_{\text{a surrogate measure of } \mathrm{dist}(x,S)}, \quad \forall\, x \in T$$

for some positive constants c and ρ.

We focus on the case that $\rho = 1$.

Error bounds

In optimization, the existence of error bounds is closely related to

- the (upper) Lipschitz continuity / isolated calmness / calmness of the solution mappings
- the strong metric regularity / metric regularity / strong metric sub-regular / metric subregularity of the subdifferentials of the essential objective functions
- quadratic growth conditions of the optimization problems

Error bounds

Applications of the error bounds:

- the stopping rules for iterative algorithms
- the convergence rates of iterative algorithms
- exact penalty functions

Error bounds for convex composite optimization problems

Consider the convex composite optimization problems

$$
\begin{aligned}
\min \quad & h(\mathcal{A}x) + \langle c, x \rangle + p(x) \\
\text{s.t.} \quad & \mathcal{B}x \in b + \mathcal{Q},
\end{aligned}
$$

- h: a smooth convex function
- p: a proper closed convex function, may not be smooth
- $\mathcal{A}, \mathcal{B}$: linear operators
- $\mathcal{Q}$: a convex polyhedral set
- c, b: given data

The perturbed problem:

$$P(u,v) \qquad \begin{array}{ll} \min & h(\mathcal{A}x) + \langle c, x\rangle + p(x) - \langle x, u\rangle \\ \text{s.t.} & \mathcal{B}x + v \in b + \mathcal{Q}, \end{array}$$

where u and v are two perturbation parameters

Three types of error bounds

For some positive constants ε and κ:

- **Primal type error bounds:**

$$\mathrm{dist}(x, \mathrm{SOL}_P) \leq \kappa \|u\|, \quad \forall\, x \text{ solves } P(u, 0),\ \forall\, u \in \mathbb{B}_\varepsilon(0)$$

- **Dual type error bounds:**

$$\mathrm{dist}(x, \mathrm{SOL}_D) \leq \kappa \|v\|, \quad \forall\, y \text{ solves } P(0, v),\ \forall\, v \in \mathbb{B}_\varepsilon(0)$$

- **KKT type error bounds:**

$$\mathrm{dist}((x, y),\ \mathrm{SOL}_{\mathrm{KKT}}) \leq \kappa \|(u, v)\|,$$

$$\forall\, (x, y) \text{ being the KKT solution of } P(u, v),\ \forall\, (u, v) \in \mathbb{B}_\varepsilon(0)$$

Three types of error bounds

For convex optimization problems, the linear convergence rate of the iteration sequence can be derived from the error bounds:

- The primal type error bounds: the proximal point algorithm
- The dual type error bounds: **the dual sequence** of the augmented Lagrangian method
- The KKT type error bounds: the proximal augmented Lagrangian method; the alternating direction method of multipliers

Sufficient conditions of error bounds

- A set-valued mapping G is called metrically subregular at $\bar{u}$ for $\bar{v}$ if $(\bar{u}, \bar{v}) \in \operatorname{gph} G$ and there exist $\delta > 0$, $\varepsilon > 0$ and $\kappa > 0$ such that

$$\operatorname{dist}(u, G^{-1}(\bar{v})) \leq \kappa \operatorname{dist}(\bar{v}, G(u) \cap \mathbb{B}_\delta(\bar{v})) \quad \forall\, u \in \mathbb{B}_\varepsilon(\bar{u}).$$

- Let $\mathcal{Q} \subseteq \mathbb{U}$ be a pointed convex closed cone (a cone is said to be pointed if $z \in \mathcal{Q}$ and $-z \in \mathcal{Q}$ implies that $z = 0$). The closed convex set $\mathcal{K} \subseteq \mathbb{V}$ is said to be $\mathcal{C}^2$-cone reducible at $\overline{X} \in \mathcal{K}$ to the cone $\mathcal{Q}$, if there exist an open neighborhood $\mathcal{W} \subseteq \mathbb{V}$ of $\overline{X}$ and a twice continuously differentiable mapping $\Xi : \mathcal{W} \to \mathbb{U}$ such that: (i) $\Xi(\overline{X}) = 0 \in \mathbb{U}$; (ii) the derivative mapping $\Xi'(\overline{X}) : \mathbb{V} \to \mathbb{U}$ is onto; (iii) $\mathcal{K} \cap \mathcal{W} = \{X \in \mathcal{W} \mid \Xi(X) \in \mathcal{Q}\}$. A function p is called $\mathcal{C}^2$-cone reducible if $\operatorname{epi} p$ is a $\mathcal{C}^2$-cone reducible set.

Examples of $\mathcal{C}^2$-cone reducible sets: convex polyhedral sets; positive semidefinite cone; epigraph of Ky Fan k-norm functions

Sufficient conditions of error bounds

The primal type error bounds hold under one of the following two conditions:

- $\partial p(\cdot)$ (subdifferential) is metrically subregular and there exists a KKT point satisfying the partial strict complementarity condition with respect to the complementarity condition $s \in \partial p(x)$
- $p(\cdot)$ is $\mathcal{C}^2$-cone reducible and the primal second order sufficient condition holds (the solution is unique)

Sufficient conditions of error bounds

Consider the case that p is a spectral function, i.e.,

$$p(\cdot) = g \circ \sigma(\cdot)$$

for some absolutely symmetric function g, or

$$p(\cdot) = g \circ \lambda(\cdot)$$

for some symmetric function g, where $\sigma(\cdot)$ and $\lambda(\cdot)$ are singular value and eigenvalue functions of a given matrix, respectively.

Examples of spectral functions:

- $g(x) = \delta_{\mathcal{R}^n_+}(x) \longrightarrow p(X) = \delta_{\mathcal{S}^n_+}(X)$ (the indicator function over the positive semidefinite cone)
- $g(x) = \|x\|_1 \longrightarrow p(X) = \|X\|_*$ (the nuclear norm function)
- $g(x) = \sum_{i=1}^{n} \log x_i \longrightarrow p(X) = \log\det X$

Sufficient conditions of error bounds

Let p be a spectral function. Then

- the metrically subregular of $\partial g \Longrightarrow$ the metrically subregular of ∂p
- the $\mathcal{C}^2$-cone reducibility of $g \Longrightarrow$ the $\mathcal{C}^2$-cone reducibility of p

[Cui, Ding and Zhao, SIOPT (2017)]

If g is a piecewise linear quadratic function, then ∂g is metrically subregular [Robinson (1981)$+$ J. Sun (1986)]

This implies the metric subregularity of $\partial \delta_{\mathcal{S}^n_+}(\cdot)$ (which is the normal cone of $\mathcal{S}^n_+$) and $\partial \| \cdot \|_*$

Sufficient conditions of error bounds

For the convex quadratic semidefinite programming

$$\begin{array}{ll} \min & \frac{1}{2}\langle X, \mathcal{Q}X\rangle + \langle C, X\rangle \\ \text{s.t.} & \mathcal{A}X = b, \quad l \leq \mathcal{B}X \leq u, \quad X \in \mathcal{S}^n_+, \end{array}$$

the primal error bound holds if there exists a partial strict complementarity KKT solution satisfying

$$\operatorname{rank}(\overline{X}) + \operatorname{rank}(\overline{S}) = n.$$

Do not need the strict complementarity with respect to $l \leq \mathcal{B}X \leq u$.

Sufficient conditions of error bounds

The KKT type error bounds are much more difficult to be satisfied.

Example 1

Consider the following SDP problem and its dual:

$$\min_{(x_1,x_2)\in\mathbb{S}^2\times\mathbb{R}} \ \delta_{\mathbb{S}^2_+}(x_1) \qquad\qquad \max_{s\in\mathbb{S}^2} \ s_{22}-\delta_{\mathbb{S}^2_-}(s)$$

$$\text{s.t.} \quad x_1+\begin{pmatrix} 0 & x_2 \\ x_2 & -x_2 \end{pmatrix}=\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}; \qquad\qquad \text{s.t.} \quad 2s_{12}-s_{22}=0.$$

$$\mathrm{SOL_P}=\left\{\bar{x}_1=\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},\ \bar{x}_2=0\right\},\ \mathrm{SOL_D}=\left\{\bar{s}=\begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix}\ \middle|\ t\le 0\right\}.$$

Sufficient conditions of error bounds

For the above example:

- there exists a KKT point satisfying the strict complementary condition (so that both the primal and the dual type error bounds hold at every solution point)
- the primal SOSC holds at the unique solution $\bar{x} = (\bar{x}_1, \bar{x}_2)$
- the dual SOSC holds at $\bar{s} = \begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix}$ with $t > 0$.
- the KKT type error bound fails at $(\bar{x}, \bar{s})$ with $\bar{s} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

Recall the convex optimization problem

$$\begin{array}{ll} \min & f^0(x) := h(\mathcal{A}x) + \langle c, x\rangle + p(x) \\ \text{s.t.} & \mathcal{B}x \in b + \mathcal{Q} \end{array}$$

Let $\sigma > 0$ be a given penalty parameter. The augmented Lagrangian function:

$$L_\sigma(x, y) := f^0(x) + \frac{1}{2\sigma}\left(\|\Pi_{\mathcal{Q}^\circ}[y + \sigma(\mathcal{B}x - b)]\|^2 - \|y\|^2\right)$$

The augmented Lagrangian method (ALM):

$$\left\{ \begin{array}{l} x^{k+1} \approx \arg\min \left\{\zeta_k(x) := L_{\sigma_k}(x, y^k)\right\}, \\ y^{k+1} = \Pi_{\mathcal{Q}^\circ}[y^k + \sigma_k(\mathcal{B}x^{k+1} - b)], \quad k \geq 0. \end{array} \right.$$

error bounds and convergence rates of the ALM

The (super)linear convergence rates of the ALM:

- Powell (equality constrained problem): assume the SOSC and the LICQ ("arbitrarily fast linear convergence")
- Rockafellar (convex nonlinear programming): assume the Lipschitz continuity of the dual solution mapping at the origin
- Bertsekas (nonlinear programming): assume the strict complementarity, the SOSC and the LICQ

error bounds and convergence rates of the ALM

For solving the convex composite optimization problems, a direct extension of [Rockafellar 1976, Luque 1984] shows that

- under the dual type error bounds, the dual sequence $\{y^k\}$ generated by the ALM convergences asymptotically Q-superlinearly
- under the KKT type error bounds, the primal sequence $\{x^k\}$ generated by the ALM convergences asymptotically R-superlinearly

If the KKT type error bounds fail, what about the convergence rates of the primal sequence or KKT residues?

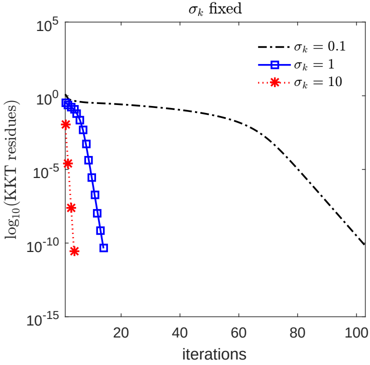

Figure: The KKT residual norm of the sequence generated by the ALM for solving Example 1 with different values of the penalty parameter σ_k.

error bounds and convergence rates of the ALM

Stopping criteria for the global convergence and local convergence rates [Rockafellar 1976]:

$$(A) \quad \zeta_k(x^{k+1}) - \inf \zeta_k \leq \varepsilon_k^2/2\sigma_k, \quad \sum_{k=0}^{\infty} \varepsilon_k < \infty,$$

$$(B) \quad \zeta_k(x^{k+1}) - \inf \zeta_k \leq (\eta_k^2/2\sigma_k)\|y^{k+1} - y^k\|^2, \quad \sum_{k=0}^{\infty} \eta_k < \infty,$$

Under the dual type error bound (with modulus κ):

- dist $(y^{k+1}, \mathrm{SOL_D}) \leq \mu_k$ dist $(y^k, \mathrm{SOL_D})$, $\quad \mu_k \to \kappa/\sqrt{\kappa^2 + \sigma_\infty^2}$ dual sequence
- $\|\Pi_{\mathcal{Q}^\circ}(\mathcal{B}x^{k+1} - b)\| \leq \mu_k'$ dist $(y^k, \mathrm{SOL_D})$, $\quad \mu_k' \to 1/\sigma_\infty$ primal feasibility
- $|\langle y^{k+1}, \mathcal{B}x^{k+1} - b\rangle| \leq \mu_k''$ dist $(y^k, \mathrm{SOL_D})$, $\quad \mu_k'' \to \|y^\infty\|/\sigma_\infty$ complementarity
- $f^0(x^{k+1}) - \inf\,(\mathrm{P}) \leq \mu_k'''$ dist $(y^k, \mathrm{SOL_D})$, $\quad \mu_k''' \to \|y^\infty\|/\sigma_\infty$ primal objectives

Implementable criteria

For any given $k \geq 0$ and $y^k \in \mathbb{Y}$, let

$$\begin{cases} y^{k+1} := \Pi_{\mathcal{Q}^\circ}[y^k + \sigma_k(\mathcal{B}x^{k+1} - b)] \\ w^{k+1} := \nabla h(\mathcal{A}x^{k+1}) \\ s^{k+1} := \mathrm{Prox}_{p^*}[x^{k+1} - (\mathcal{A}^*\tilde{w}^k(x^{k+1}) + \mathcal{B}^*\tilde{y}^k(x^{k+1}) + c)] \\ z^{k+1} := (w^{k+1}, y^{k+1}, s^{k+1}) \\ e^{k+1} := x^{k+1} - \mathrm{Prox}_p[x^{k+1} - (\mathcal{A}^*\tilde{w}^k(x^{k+1}) + \mathcal{B}^*\tilde{y}^k(x^{k+1}) + c)] \end{cases}$$

Note that $e^{k+1} = 0 \Longleftrightarrow x^{k+1} = \arg\min \zeta_k(x)$

If the Slater condition holds, then (A) and (B) can be implemented via

$$(A')\ \|e^{k+1}\| \leq \frac{\widehat{\varepsilon}_k^2/\sigma_k}{1+\|x^{k+1}\|+\|z^{k+1}\|} \min\left\{ \frac{1}{\|\nabla h^*(w^{k+1})\|+\|y^{k+1}-y^k\|/\sigma_k+1/\sigma_k},\ 1 \right\}$$

$$(B')\ \|e^{k+1}\| \leq \frac{(\widehat{\eta}_k^2/\sigma_k)\|y^{k+1}-y^k\|^2}{1+\|x^{k+1}\|+\|z^{k+1}\|} \min\left\{ \frac{1}{\|\nabla h^*(w^{k+1})\|+\|y^{k+1}-y^k\|/\sigma_k+1/\sigma_k},\ 1 \right\}$$

Solving the subproblems via the semismooth Newton-CG method

Given the semismooth equation

$$F(x) = 0$$

The semismooth Newton method:

$$x^{k+1} = x^k - V_k^{-1} F(x^k), \quad V^k \in \partial F(x^k)$$

($\partial F(x^k)$: the Clarke generalized Jacobian of F at x^k)

The nonsingularity of $\partial F(x^*) \Longrightarrow$ the superlinear convergence of $\{x^k\}$

Examples

Lasso problem:

$$\min \frac{1}{2}\|\mathcal{A}x - b\|^2 + \lambda\|x\|_1$$

- The dual SOSC holds (nonlinear programming: the KKT type error bounds hold) $\Longrightarrow$ both primal and dual ALMs have the superlinear convergence rates
- The dual constraint nondegeneracy fails (the primal problem may have multiple solutions) $\Longrightarrow$ primal semismooth Newton $\times$
- The primal constraint nondegeneracy holds $\Longrightarrow$ dual ALM + semismooth Newton ✓

Examples

Sparse estimation of a Gaussian graphical model:

$$\begin{array}{ll} \min_{X \succ 0} & -\log\det X + \langle S, X\rangle + \rho\|X\|_1, \\ \text{s.t.} & \mathcal{A}X = b, \end{array}$$

where S is a given sample covariance matrix.

- The strict complementarity with respect to $-\log\det X$ holds $\Longrightarrow$ both primal and dual ALMs have the superlinear convergence rates
- The primal constraint nondegeneracy fails $\Longrightarrow$ dual ALM + semismooth Newton $\times$
- The dual constraint nondegeneracy holds $\Longrightarrow$ primal ALM + semismooth Newton $\checkmark$

References

Y. Cui, D.F. Sun and K.C. Toh, *On the R-superlinear convergence of the KKT residues generated by the augmented Lagrangian method for convex composite conic programming*, arXiv:1706.08800, 2017.

Thank you!