Error Bounds for Parametric Polynomial Systems with Applications to Higher-Order Stability Analysis and Convergence Rate

Guoyin Li

The University of New South Wales

November 2016, WoMBaT 2016

Joint work with B.S. Mordukhovich, T. A. Nghia and T.S. Pham

Outline

Outline

Outline

Outline

Outline

For $f : \mathbb{R}^n \to \mathbb{R}$, we consider the following inequality system

$$(S) \quad f(z) \leq 0.$$

- To judge whether x is an approximate solution of (S), we want to know $d(x, [f \leq 0]) := \inf\{\|x - z\| : f(z) \leq 0\}$.
- However, we often measure $[f(x)]_+ := \max\{f(x), 0\}$.
- So, we seek an error bound: there exist $\tau, \delta > 0$ such that

$$d(x, [f \leq 0]) \leq \tau\left([f(x)]_+ + [f(x)]_+^{\delta}\right)$$

either locally or globally.

For $f : \mathbb{R}^n \to \mathbb{R}$, we consider the following inequality system

$$(S) \quad f(z) \leq 0.$$

- To judge whether x is an approximate solution of (S), we want to know $d(x, [f \leq 0]) := \inf\{\|x - z\| : f(z) \leq 0\}$.
- However, we often measure $[f(x)]_+ := \max\{f(x), 0\}$.
- So, we seek an error bound: there exist $\tau, \delta > 0$ such that

$$d(x, [f \leq 0]) \leq \tau\big([f(x)]_+ + [f(x)]_+^{\delta}\big)$$

either locally or globally.

Definition

We say f has a
(1) global error bound with exponent δ if there exist $\tau > 0$ such that

$$d(x, [f \leq 0]) \leq \tau\big([f(x)]_+ + [f(x)]_+^{\delta}\big) \text{ for all } x \in \mathbb{R}^n$$

(2) local error bound with exponent δ around $\overline{x}$ if there exist $\tau, \epsilon > 0$ such that

$$d(x, [f \leq 0]) \leq \tau\big([f(x)]_+ + [f(x)]_+^{\delta}\big) \text{ for all } x \in \mathbb{B}(\overline{x}; \epsilon).$$

If $\delta = 1$ in (1) (resp. (1)), we say f has a Lipschitz type global (resp. local) error bound.

Error bound is useful in

- analyzing the convergence properties of algorithms (e.g. Luo 2000, Fukushima 2005, Attouch etal. 2009, Tseng 2010 and Izmailov & Solodov 2014);
- sensitivity analysis of optimization problem/variational inequality problem (e.g. Jourani 2000, Ye 2002)
- identifying the active constraints (e.g. Facchinei etal. 1998 and Pang 1997)
- studying maximal monotone operator (Borwein & Dutta 2015) and mixed integer programming problem (Stein, 2016)

Some Known Results

- Lipschitz type global error bound holds when f is maximum of finitely many affine functions (Hoffman 1951)
- Global error bound can fail even when f is convex and continuous (e.g. $f(x_1, x_2) = x_1 + \sqrt{x_1^2 + x_2^2}$).
- Has close link with metric sub-regularity and calmness.
- Many further developments (e.g. Aze, Ioffe, Klatte, Kummer, Kruger, Lewis, Li, López, Ng, Ngai, Outrata, Pang, Robinson, Thera, Ye etc...)

Quadratic cases

- Global error bound with exponent $1/2$ holds when f is a convex quadratic function. (Luo and Luo, 1994); extended to convex quadratic system, (Wang and Pang 1994).
- Local error bound with exponent $1/2$ holds when f is a (nonconvex) quadratic function. (Luo and Sturm, 1998).
- Open questions raised by Luo and Sturm: what happens for the case f can be expressed as finitely many (nonconvex) quadratic functions?

Motivating Example: go beyond quadratic

Consider $f(x) = x^2$. Then, $[f \leq 0] = \{0\}$ and so,

$$d(x, [f \leq 0]) = |x| \leq (x^2)^{\frac{1}{2}} = [f(x)]_+^{\frac{1}{2}}.$$

More generally, consider $f(x) = x^d$ with d is an even number. Then,

$$d(x, [f \leq 0]) = |x| \leq (x^d)^{\frac{1}{d}} = [f(x)]_+^{\frac{1}{d}}.$$

Motivating Example: go beyond quadratic (cont.)

Let d be an even number. Consider $f(x) = \max\{f_1(x), \ldots, f_n(x)\}$ where

$$f_1(x) := x_1^d \text{ and } f_i(x) := x_i^d - x_{i-1},\ i = 2, \cdots, n.$$

Then, $[f \le 0] = \{0\}$. Consider $x(t) = (t^{d^{n-1}}, t^{d^{n-2}}, \cdots, t^d, t) \in \mathbb{R}^n$, $t \in (0,1)$. Then

- $d(x(t), [f \le 0]) = O(t)$;
- $[f(x(t))]_+ = f(x(t)) = t^{d^n}$

So,

$$d(x(t), [f \le 0]) = O\left([f(x(t))]_+^{\frac{1}{d^n}}\right).$$

Recent development for polynomial systems

- Global error bound with exponent $\tau_0 = \frac{1}{(d-1)^n+1}$ holds when f is a convex polynomial with degree d on $\mathbb{R}^n$ (L. SIOPT 2010).
- Local error bound with exponent $\tau_1 = \max\left\{\frac{2}{(2d-1)^n+1}, \frac{1}{\beta(n-1)d^n}\right\}$ if f is maximum of finitely many convex polynomials with degree d on $\mathbb{R}^n$, where $\beta(s)$ is the central binomial coefficient $\binom{s}{[s/2]}$ (Borwein, L. & Yao, SIOPT 2014 and Ngai, SIOPT 2015).
- Local error bound with exponent $\tau_2 = \max\left\{\frac{1}{(d+1)(3d)^{n+r}}, \frac{1}{d(6d-3)^{n+r-1}}\right\}$ when f is maximum of r many (nonconvex) polynomials with degree d on $\mathbb{R}^n$ (L., Mordukhovich, Pham, MP, 2014).


WILD
and
INFINITE
You always were and always will be.
The Infinite Spark of Being

Main Problem

- Can we extend the error bound results to parametric polynomial system:
$$\phi_l(x,y) \le 0 \text{ for all } y \in \Omega, l = 1,\dots,L,$$
where ϕ_l are polynomials on $\mathbb{R}^n \times \mathbb{R}^m$ with degree d and $\Omega := \{y \in \mathbb{R}^m | \ g_i(y) \le 0,\ i = 1,\dots,r;\ h_j(y) = 0,\ j = 1,\dots,s\}$ with g_i and h_j are all polynomials on $\mathbb{R}^n$ with degree d.
- System of this form arises in semi-infinite programming problem and bilevel programming problem with polynomial data .

Why do we care?

- Many important nonlinear conic programs can be covered in this framework:
  - (1) second order cone constraint with polynomial data:

$$\|(f_1(x),\ldots,f_m(x))\| \le f_0(x)$$
$$\iff \sum_{j=1}^{m} y_j f_j(x) - f_0(x) \le 0 \ \text{ for all } \ \|y\|^2 = 1$$

  - (2) polynomial matrix inequality constraint:

$$P(x) \preceq 0 \iff \sum_{i,j=1}^{m} y_i \left(P(x)\right)_{i,j} y_j \le 0 \ \text{ for all } \ \|y\|^2 = 1.$$

Main Results: Assumptions

We suppose that the set of parameters

$$\Omega = \{y \in \mathbb{R}^m | \ g_i(y) \leq 0, \ i = 1, \ldots, r; \ h_j(y) = 0, \ j = 1, \ldots, s\}$$

be bounded and regular.

What is a regular set?

Ω is regular if, for all $y \in \Omega$, the following MFCQ holds:

$$\left.\begin{array}{l} \sum_{i=1}^{r} \lambda_i \nabla g_i(y) + \sum_{j=1}^{s} \kappa_j \nabla h_j(y) = 0, \\ \lambda_i \geq 0, \ \lambda_i g_i(y) = 0, \ \text{ and } \ \kappa_j \in \mathbb{R} \end{array}\right\} \Longrightarrow \lambda_i = 0, \ \kappa_j = 0 \qquad (3.0)$$

Examples of Regular sets

- Discrete set $\Omega := \{1, \ldots, p\}$, $p \in \mathbb{N}$. Write it as

$$\Omega = \{y \in \mathbb{R} |\ h(y) = 0\} \text{ with } h(y) := (y-1)(y-2)\cdots(y-p).$$

  and observe that $\nabla h(y) \neq 0$ for all $y \in \Omega$.

- The algebraic set

$$\Omega := \{y \in \mathbb{R}^m |\ h_1(y) = 0, \ldots, h_s(y) = 0\}, \quad s \leq m,$$

  is regular provided that $\mathrm{rank}\big(\nabla h_1(y), \ldots, \nabla h_s(y)\big) = s$ for all $y \in \Omega$. (e.g. Sphere under l^p-norm where $p > 1$ is an integer).

Error bounds for parametric polynomial system

Let $R(n,d) := d(3d-3)^{n-1}$.

Theorem

Let $\phi(x) = \max_{\substack{y\in\Omega \\ 1\le l\le L}} f_l(x,y)$ where $f_l\colon \mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$, are polynomials of degrees at most d and

$$\Omega = \left\{y\in\mathbb{R}^m \,\middle|\, g_i(y)\le 0,\ i=1,\dots,r;\ h_j(y)=0,\ j=1,\dots,s\right\}$$

is bounded and regular, where g_i, h_j are polynomials on $\mathbb{R}^m$ with degree d. Then, for any $\bar{x}\in\mathbb{R}^n$ there exist constants $c,\varepsilon>0$ such that

$$d(x,[\phi\le 0]) \le c\,[\phi(x)]_+^{\tau} \quad \text{whenever } \|x-\bar{x}\|\le\varepsilon, \tag{3.0}$$

where $\tau = \frac{1}{R(2n+(m+r+s+2)(n+1),\,d+L+1)}$.

Remark: Can be extended to more general cases where Ω is replaced by a set-valued mapping $Y(x)$ and under weaker regularity assumption.

What is behind the proof?

Łojasiewicz gradient inequality and its variants

- (Łojasiewicz inequality) Let f be an analytic function on $\mathbb{R}^n$ with $f(0) = 0$ and $\nabla f(0) = 0$. Then, exists a rational number $\tau \in (0, 1]$ and $c, \delta > 0$ s.t. $\|\nabla f(x)\| \geq c|f(x)|^{1-\tau}$ for all $\|x\| \leq \delta$.
- (Gwoździewicz 1999 and Kollar 2002) In addition, if f is a polynomial with degree d and 0 is a strict local minimizer, then, $\tau = \frac{1}{(d-1)^n+1}$;
- Dropping the strict minimizer assumption in Gwoździewicz's result, we have a new estimate $\tau = R(n, d)^{-1} = \frac{1}{d(3d-3)^{n-1}}$ (Kurdyka 2012, and L., Mordukhovich and Pham 2014).

Outline of the proof

Goal: Error bound for $\phi(x) = \max_{\substack{y\in\Omega \\ 1\le l\le L}} f_l(x,y)$.

- **Case 1:** $L = 1$. Then $\phi(x) = \max\{f(x,y) : g_i(y) \le 0,\ h_j(y) = 0\}$ and its Lagrangian-type function is

$$F(x,y,\mu,\kappa) := -f(x,y) + \sum_{i=1}^{r} \mu_i^2 g_i(y) + \sum_{j=1}^{s} \kappa_j h_j(y)$$

  (1) Reduce the error bound for ϕ to F;
  (2) F is a single polynomial and so, Łojasiewicz's gradient inequality for single polynomial applies.

- **Case 2:** $L \ge 2$. Reduce it to the case $L = 1$ by using

$$\max_{y\in\Omega} \max_{1\le l\le L} f_l(x,y) = \max_{(y,t)\in\Omega\times\{1,\ldots,L\}} \sum_{l=1}^{L} \gamma_l(t) f_l(x,y).$$

where $\gamma_l : \mathbb{R} \to \mathbb{R}$ is the *Lagrange interpolation polynomials*

Polynomial matrix inequalities: regularity free

Let P be an $(m \times m)$ *polynomial matrix* of n variables with degree d, and let $S_{PMI} = \{x : P(x) \preceq 0\}$.

Corollary

For a compact set $K \subset \mathbb{R}^n$ there is $c > 0$ such that

$$\mathrm{dist}(x, S_{PMI}) \leq c \left(\left[\lambda_{\max}(P(x)) \right]_+ \right)^{\tau} \quad \textit{whenever} \quad x \in K,$$

where $\tau = R(2n + (m+1)(n+1), d+3)^{-1}$ and $\lambda_{\max}$ denotes the maximum eigenvalue of a symmetric matrix.

Recall that:

$$P(x) \preceq 0 \Longleftrightarrow \sum_{i,j=1}^{m} y_i \left(P(x)\right)_{i,j} y_j \leq 0 \;\text{ for all }\; \|y\|^2 = 1.$$

Hunting for the true exponent

Example

Let d be an even number. For any $x = (x_1, \ldots, x_n)$ define $A_1(x) := x_1^d$ and then, for any $i = 2, \ldots, n$,

$$A_i(x) := \begin{pmatrix} -1 & x_i^d \\ x_i^d & -x_{i-1} \end{pmatrix},$$

and the polynomial matrix inequality $P(x) := \mathrm{diag}\big(A_1(x), \ldots, A_n(x)\big)$ Then, $S = \{x : P(x) \preceq 0\} = \{0\}$. For $x(t) = (t^{(2d)^{n-1}}, t^{(2d)^{n-2}}, \cdots, t^{2d}, t) \in \mathbb{R}^n$,

- $d(x(t), S) = O(t)$;
- $\lambda_{\max}\big(P(x(t))\big) = t^{d(2d)^{n-1}}$

Thus, $\tau \leq \frac{1}{d(2d)^{n-1}}$ while our estimate gives $\tau = \frac{1}{(d+3)(3d+6)^{4n-1}}$.

Cyclic Projection Algorithm

(1) Initialization: Given $x_0 \in \mathbb{R}^n$ and finite many closed convex sets $C_1, C_2, \cdots, C_L$ in $\mathbb{R}^n$ with $\bigcap_{l=1}^{L} C_i \neq \emptyset$.

(2) Algorithm: The sequence of *cyclic projections*, $(x_k)_{k\in\mathbb{N}}$, is defined by

$$x_1 := P_1 x_0, x_2 := P_2 x_1, \cdots, x_L := P_L x_{L-1}, x_{L+1} := P_1 x_L \ldots$$

where P_l denotes the Euclidean projection to the set C_l.

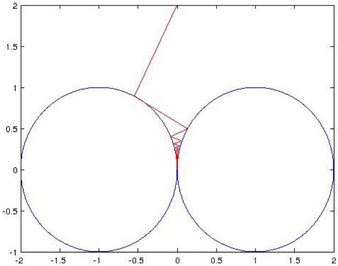
(3) Output: A point in the intersections of C_l.

When $L = 2$, it reduces to alternating projection method.

Let $x_0 = (0, 2)$ and

$$C_1 := \left\{(a, b) \in \mathbb{R}^2 \mid (a+1)^2 + b^2 - 1 \leq 0\right\}$$
$$C_2 := \left\{(a, b) \in \mathbb{R}^2 \mid (a-1)^2 + b^2 - 1 \leq 0\right\}.$$

The following figure depicts the algorithm's trajectory:



This algorithm is easy to implement and was studied by a lot of researchers:

- always convergent to $x_\infty \in \bigcap_{l=1}^{L} C_l$ (Bregman, 1950)
- linearly convergence whenever $\mathrm{int} \bigcap_{l=1}^{L} C_l \neq \emptyset$ (cf. Bauschke & Borwein, 1996).

What is the convergence rate in the degenerate cases e.g. when $\mathrm{int} \bigcap_{l=1}^{L} C_l = \emptyset$?

A partial answer can be given when each C_l has suitable polynomial structures.

Definition

We say C_l, $l = 1, \ldots, L$ are polynomial matrix inequality representable convex sets if C_l is convex and

$$C_l := \left\{x \in \mathbb{R}^n \,\middle|\, A^{(l)}(x) \preceq 0\right\}, \quad l = 1, \ldots, L,$$

where every $A^{(l)} : \mathbb{R}^n \to S^m$ is a polynomial matrix mapping such that each $\left(A^{(l)}(x)\right)_{ij}$ is a real polynomial with degree at most d.

Theorem (L. Mordukhovich, Nghia, Pham 2016)

Let $x_0 \in \mathbb{R}^n$ and let $(x_k)_{k \in \mathbb{N}}$ be the sequence generated by the cyclic projection algorithm for the above C_l. Then $x_k \to x_\infty \in \bigcap_{l=1}^{L} C_l$, and there exists $M > 0$ such that

$$\|x_k - x_\infty\| \leq M \frac{1}{k^\rho}, \quad \forall k \in \mathbb{N},$$

where $\rho := \frac{1}{[2\, R(2n+(m+3)(n+1), d+L) - 2]^{-1}}$.

Further recent advance on cyclic (alternating) projection algorithm

- For the linear matrix inequality case, sharper convergence rate for was obtained in " Drusvyatskiy, L., & Wolkowicz, Alternating projections for ill-posed semi-definite feasibility problems, MP 2016"

Other algorithms

In the semi-algebraic setting, similar techniques can also be used to analyzing the convergence rate of

- proximal point algorithm (Bolte & Attouch, 2013;L. & Mordukhovich, 2012)
- Douglas-Rachford algorithm and one of its variant (L. and Pong, MP, 2016; Borwein, L., Tam, to appear in SIOPT).

High-order stability for polynomial problems

Consider the following parameterized problem:

$$(PMI_u) \qquad \text{minimize} \;\; f(x,u) \;\; \text{subject to} \;\; P(x) \preceq 0,$$

where $u \in \mathbb{R}^l$ is the perturbation parameter, and

- $f(\cdot, u)$ is a polynomial with degree d and $f(x, \cdot)$ is locally Lipschitzian.
- $P : \mathbb{R}^n \to S^m$ is such that each (i,j)th element $\big(P(x)\big)_{ij}$, $1 \leq i,j \leq m$, is a real polynomial with degree d.
- the feasible set is compact.

For each $u \in \mathbb{R}^l$ denote the solution set of $(POP)_u$ by $S(u)$.

Example:

$$\max_{x\in\mathbb{R}^n} \quad f(x,u) := p_0(x) + \sum_{i=1}^{l} u_i p_i(x)$$

$$\text{subject to} \quad A_0 + \sum_{j=1}^{m} x_j A_j + \sum_{j,k=1}^{m} x_j x_k B_{jk} \preceq 0,$$

where all p_i are polynomials.

Hölder continuity of solution maps

Theorem (L., Mordukhovich, Nghia, & Pham, 2016)

For (POP_u), let $\overline{u} \in \mathbb{R}^l$. Then, there are constants $c, \delta > 0$ such that

$$S(u) \subset S(\overline{u}) + c\,\|u - \overline{u}\|^{\tau}\, \overline{\mathbb{B}}(0,1) \quad \textit{whenever} \quad \|u - \overline{u}\| \leq \delta$$

with the explicit exponent

$$\tau = R\big(2n + (m+3)(n+1), d+4\big)^{-1}.$$

- Similar Hölder continuity of solution maps for nonlinear 2nd-order cone program and generalized semi-infinite program with polynomial data were also provided.
- Has important application in deriving high-order semismooth property for maximum eigenvalue of a tensor (L., Qi & Yu 2013 and L., Mordukhovich & Pham 2014).

Conclusion

- Error bound is an interesting research topic and has many important applications;
- Variational analysis and semi-algebraic techniques could shed some light on how to improve error bound results from quadratic to polynomial cases.

Future Works

Still very preliminary development. A lot of interesting questions, e.g.

(1) How to sharpen the derived exponent?

(2) How to estimate/compute the error bound constant c and the radius constant δ (for local cases)?

(3) For the stability result, what happens if we also perturb the constraint functions ?

Want to know more?

(1) J. Borwein, G. Li and L. Yao, Analysis of the convergence rate for the cyclic projection algorithm applied to semi-algebraic convex sets, *SIAM J. Optim.*, 24(1), 498-527, 2014

(2) G. Li, On the asymptotically well behaved functions and global error bound for convex polynomials, *SIAM J. Optim.*, 20 (2010), No. 4, 1923-1943.

(3) V. Jeyakumar, J.B. Lasserre, G. Li and T.S. Pham, Convergent semidefinite programming relaxations for global bilevel polynomial optimization problems, to appear in SIAM J. Optim (2016), doi: DOI:10.1137/15M1017922.

(4) G. Li, B.S. Mordukhovich, T.T.A. Nghia, T.S. Pham, Error bounds for parametric polynomial systems with applications to higher-order stability analysis and convergence rates, to appear in Math. Program. (2016), DOI 10.1007/s10107-016-1014-6.

(5) G. Li, B.S. Mordukhovich and T.S. Pham, New fractional error bounds for nonconvex polynomial systems with applications to Hölderian stability in optimization and spectral theory of tensors, *Math. Program.*, 153 (2015), no. 2, Ser. A, 333-362.

(6) G. Li and B.S. Mordukhovich, Hölder metric subregularity with applications to proximal point method, *SIAM J. Optim.*, 22 (2012), No. 4, 1655-1684.

(7) G. Li, L.Q. Qi, G. Yu, Semismoothness of the maximum eigenvalue function of a symmetric tensor and its application. Linear Algebra Appl. 438 (2013), no. 2, 813-833.

Thanks !

$f(x_1, x_2) = x_1 + \sqrt{x_1^2 + x_2^2}$. $[f \leq 0] = \{(x_1, x_2) : x_1 \leq 0, x_2 = 0\}$.
Consider $x^n = (-n, 1)$. Then $d(x^n, [f \leq 0]) = 1$ and
$f(x^n) = -n + \sqrt{n^2 + 1} = \frac{1}{\sqrt{n^2+1}+n} \to 0$.