Hoffman constants for systems of linear inequalities

Javier Peña, Carnegie Mellon University
(joint work with J. Vera, L. Zuluaga, and D. Gutman)

Triangle Lectures in Combinatorics
North Carolina State University
November 16, 2019

Main ideas

Error bounds: cases where "almost" implies "near"

Suppose $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

Linear equations $Ax = b$

If $A^{-1}(b) := \{x \in \mathbb{R}^n : Ax = b\} \neq \emptyset$ then for all $u \in \mathbb{R}^n$

$$\|Au - b\| \text{ small } \Rightarrow u \text{ near } A^{-1}(b)$$

Error bounds: cases where "almost" implies "near"

Suppose $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

Linear equations $Ax = b$

If $A^{-1}(b) := \{x \in \mathbb{R}^n : Ax = b\} \neq \emptyset$ then for all $u \in \mathbb{R}^n$

$$\|Au - b\| \text{ small } \Rightarrow u \text{ near } A^{-1}(b)$$

Linear inequalities (Hoffman's error bound) $Ax \leq b$

If $P_A(b) := \{x \in \mathbb{R}^n : Ax \leq b\} \neq \emptyset$ then for all $u \in \mathbb{R}^n$

$$\|(Au - b)_+\| \text{ small } \Rightarrow u \text{ near } P_A(b).$$

Error bounds: cases where "almost" implies "near"

Suppose $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

Linear equations $Ax = b$

If $A^{-1}(b) := \{x \in \mathbb{R}^n : Ax = b\} \neq \emptyset$ then for all $u \in \mathbb{R}^n$

$$\|Au - b\| \text{ small } \Rightarrow u \text{ near } A^{-1}(b)$$

Linear inequalities (Hoffman's error bound) $Ax \le b$

If $P_A(b) := \{x \in \mathbb{R}^n : Ax \le b\} \neq \emptyset$ then for all $u \in \mathbb{R}^n$

$$\|(Au - b)_+\| \text{ small } \Rightarrow u \text{ near } P_A(b).$$

Linear equations relative to a set $Ax = b, \; x \in R$

Suppose $R \subseteq \mathbb{R}^n$ and $b \in A(R)$, that is, $A^{-1}(b) \cap R \neq \emptyset$. Is it the case that for all $u \in R$

$$\|Au - b\| \text{ small } \Rightarrow u \text{ near } A^{-1}(b) \cap R?$$

Agenda

- Error bound for $Ax = b$
- Hoffman's error bound for $Ax \leq b$
- Error bounds for $Ax = b, \ x \in R$ and for $Ax \leq b, \ x \in R$
- Condition number relative to a reference set
- Algorithms to compute Hoffman constants

Error bound for systems of linear equations

Systems of linear equations $Ax = b$

Suppose $\mathbb{R}^n$ and $\mathbb{R}^m$ be endowed with norms and $A \in \mathbb{R}^{m\times n} \setminus \{0\}$.

Then for all $b \in A(\mathbb{R}^n) := \{Ax : x \in \mathbb{R}^n\}$ and all $u \in \mathbb{R}^n$

$$\operatorname{dist}(u, A^{-1}(b)) \leq \|A^{-1}\| \cdot \|Au - b\| \tag{1}$$

where

$$\|A^{-1}\| = \max_{\substack{y\in A(\mathbb{R}^n)\\ \|y\|\leq 1}} \min_{\substack{x\in\mathbb{R}^n\\ Ax=y}} \|x\| = \frac{1}{\displaystyle\min_{v\in A(\mathbb{R}^n),\, \|v\|^*=1} \|A^{\mathsf{T}} v\|^*}.$$

Systems of linear equations $Ax = b$

Suppose $\mathbb{R}^n$ and $\mathbb{R}^m$ be endowed with norms and $A \in \mathbb{R}^{m \times n} \setminus \{0\}$.

Then for all $b \in A(\mathbb{R}^n) := \{Ax : x \in \mathbb{R}^n\}$ and all $u \in \mathbb{R}^n$

$$\operatorname{dist}(u, A^{-1}(b)) \le \|A^{-1}\| \cdot \|Au - b\| \tag{1}$$

where

$$\|A^{-1}\| = \max_{\substack{y \in A(\mathbb{R}^n) \\ \|y\| \le 1}} \min_{\substack{x \in \mathbb{R}^n \\ Ax = y}} \|x\| = \frac{1}{\displaystyle\min_{v \in A(\mathbb{R}^n),\, \|v\|^* = 1} \|A^\mathsf{T} v\|^*}.$$

Inequality (1) is an *error bound* for $Ax = b$

If u "almost" solves $Ax = b$ then u is "near" a solution to $Ax = b$.

Inequality (1) is tight: there exist $b \in A(\mathbb{R}^n)$ and $u \in \mathbb{R}^n$

$$\operatorname{dist}(u, A^{-1}(b)) = \|A^{-1}\| \cdot \|Au - b\| > 0.$$

Recall: $\|v\|^* = \max_{\|y\| \le 1} v^\mathsf{T} y$. For instace, $\|\cdot\|_1^* = \|\cdot\|_\infty$ and $\|\cdot\|_2^* = \|\cdot\|_2$

Geometric interpretation of $\|A^{-1}\|$

Suppose $A \in \mathbb{R}^{m\times n} \setminus \{0\}$. Then

$$\frac{1}{\|A^{-1}\|} = \mathsf{dist}\,(0, \mathsf{relbdy}(\{Ax : \|x\| \le 1\}))$$

and

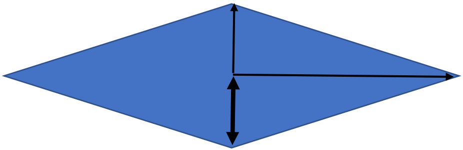
$$\frac{1}{\|A^{-1}\|} = \mathsf{dist}^*\left(0, \{A^\mathsf{T} v : v \in A(\mathbb{R}^n), \|v\|^* = 1\}\right).$$

For ℓ_2 norms $\frac{1}{\|A^{-1}\|}$ = smallest positive singular value of A.

Example

Suppose $A = \begin{bmatrix} 1 & 0 \\ 0 & \epsilon \end{bmatrix}$ for $0 \le \epsilon < 1$ and $\mathbb{R}^2$ is endowed with the ℓ_1 norm. Then

$$\frac{1}{\|A^{-1}\|} = \begin{cases} \epsilon & \text{if } \epsilon \in (0,1) \\ 1 & \text{if } \epsilon = 0. \end{cases}$$



$\{Ax : \|x\| \le 1\}$ and $1/\|A^{-1}\|$ for $0 < \epsilon < 1$.

$\{Ax : \|x\| \le 1\}$ and $1/\|A^{-1}\|$ for $\epsilon = 0$.

Hoffman's error bound for systems of linear inequalities

Systems of linear inequalities $Ax \leq b$

Key notation

For $A \in \mathbb{R}^{m\times n},\ b \in A(\mathbb{R}^n) + \mathbb{R}^m_+$, and $u \in \mathbb{R}^n$

$$P_A(b) := \{x \in \mathbb{R}^n : Ax \leq b\}.$$

Theorem (Hoffman 1952)

Let $A \in \mathbb{R}^{m\times n}$. Then there exists a constant $H(A)$ that depends only on A such that for all $b \in A(\mathbb{R}^n) + \mathbb{R}^m_+$ and all $u \in \mathbb{R}^n$

$$\mathrm{dist}(u, P_A(b)) \leq H(A) \cdot \mathrm{dist}(b - Au, \mathbb{R}^m_+).$$

For many norms: $\mathrm{dist}(b - Au, \mathbb{R}^m_+) = \|(b - Au)_-\| = \|(Au - b)_+\|$.

Characterization of $H(A)$ (special case)

Suppose $A \in \mathbb{R}^{m\times n}$ is such that $A(\mathbb{R}^n) + \mathbb{R}^m_+ = \mathbb{R}^m$. Let

$$H(A) := \max_{\substack{y\in\mathbb{R}^m \\ \|y\|\le 1}} \min_{\substack{x\in\mathbb{R}^n \\ Ax\le y}} \|x\| = \frac{1}{\displaystyle\min_{v\in\mathbb{R}^m_+,\, \|v\|^*=1} \|A^\mathsf{T} v\|^*}.$$

Characterization of $H(A)$ (special case)

Suppose $A \in \mathbb{R}^{m \times n}$ is such that $A(\mathbb{R}^n) + \mathbb{R}^m_+ = \mathbb{R}^m$. Let

$$H(A) := \max_{\substack{y \in \mathbb{R}^m \\ \|y\| \le 1}} \min_{\substack{x \in \mathbb{R}^n \\ Ax \le y}} \|x\| = \frac{1}{\displaystyle\min_{v \in \mathbb{R}^m_+,\ \|v\|^* = 1} \|A^\mathsf{T} v\|^*}.$$

Proposition (Following Renegar 1995, Ramdas & P 2015)

Suppose $A \in \mathbb{R}^{m \times n}$ is such that $A(\mathbb{R}^n) + \mathbb{R}^m_+ = \mathbb{R}^m$. Then for all $b \in \mathbb{R}^m$ and all $u \in \mathbb{R}^n$

$$\mathsf{dist}(u, P_A(b)) \le H(A) \cdot \mathsf{dist}(b - Au, \mathbb{R}^m_+).$$

This bound is tight: there exist $b \in \mathbb{R}^m$ and $u \in \mathbb{R}^n$ such that

$$\mathsf{dist}(u, P_A(b)) = H(A) \cdot \mathsf{dist}(b - Au, \mathbb{R}^m_+) > 0.$$

Observe: $A(\mathbb{R}^n) + \mathbb{R}^m_+ = \mathbb{R}^m \Leftrightarrow Ax < 0$ is feasible.

Geometric interpretation of $H(A)$ in this special case

Observe

$$
\begin{aligned}
A(\mathbb{R}^n) + \mathbb{R}^m_+ = \mathbb{R}^m &\Leftrightarrow Ax < 0 \text{ is feasible} \\
&\Leftrightarrow A^\mathsf{T} v = 0,\ v \gneq 0 \text{ infeasible.}
\end{aligned}
$$

Second step by Gordan's Theorem.

Therefore

$$A(\mathbb{R}^n) + \mathbb{R}^m_+ = \mathbb{R}^m \Leftrightarrow 0 \notin \{A^\mathsf{T} v : v \geq 0, \|v\|^* = 1\}.$$

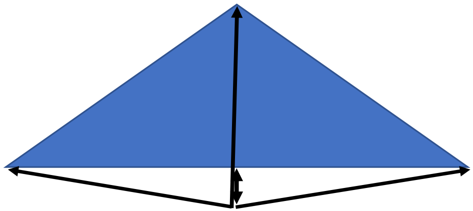
When this is the case we have

$$\frac{1}{H(A)} = \mathsf{dist}^* \left(0, \{A^\mathsf{T} v : v \geq 0, \|v\|^* = 1\}\right).$$

Example

Suppose $A = \begin{bmatrix} -1 & \epsilon \\ 1 & \epsilon \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 2}$ for $0 < \epsilon < 1$ and both $\mathbb{R}^3$ and $\mathbb{R}^2$ are endowed with the ℓ_∞ norm. Then

$$\frac{1}{H(A)} = \epsilon.$$



$\{A^\mathsf{T} v : v \geq 0,\ \|v\|^* = 1\}$ and $1/H(A)$ for $\epsilon \in (0, 1)$.

Characterization of $H(A)$ (general case)

Let $\mathcal{J}(A) := \{J \subseteq [m] : A_J(\mathbb{R}^n) + \mathbb{R}^J_+ = \mathbb{R}^J\}$ and

$$H(A) := \max_{J \in \mathcal{J}(A)} H(A_J)$$
$$= \max_{J \in \mathcal{J}(A)} \frac{1}{\min\{\|A_J^{\mathsf{T}} v\|^* : v \in \mathbb{R}^J_+, \|v\|^* = 1\}}.$$

(Here $[m]$ is shorthand for $\{1, \ldots, m\}$.)

Characterization of $H(A)$ (general case)

Let $\mathcal{J}(A) := \{J \subseteq [m] : A_J(\mathbb{R}^n) + \mathbb{R}^J_+ = \mathbb{R}^J\}$ and

$$H(A) := \max_{J \in \mathcal{J}(A)} H(A_J)$$
$$= \max_{J \in \mathcal{J}(A)} \frac{1}{\min\{\|A_J^\mathsf{T} v\|^* : v \in \mathbb{R}^J_+, \|v\|^* = 1\}}.$$

(Here $[m]$ is shorthand for $\{1, \ldots, m\}$.)

Proposition

Let $A \in \mathbb{R}^{m \times n} \setminus \{0\}$. Then for all $b \in A(\mathbb{R}^n) + \mathbb{R}^m_+$ and all $u \in \mathbb{R}^n$

$$\mathsf{dist}(u, P_A(b)) \leq H(A) \cdot \mathsf{dist}(b - Au, \mathbb{R}^m_+).$$

Furthermore, this bound is tight.

Related results by Robinson 1973, Li 1993, Klatte & Thiere 1995, Wang & Lin 2014.

Error bounds relative to a reference set

Relative error bound for equations

Consider a system of linear constraints of the form

$$\begin{aligned} Ax &= b \\ x &\in R \end{aligned}$$

where R is a "reference" set representing some easy-to-satisfy constraints, e.g., $x \geq 0$, box constraints, etc.

It is natural to consider an error bound that accounts for this kind of reference set.

Goal

For $u \in R$, bound $\text{dist}(u, A^{-1}(b) \cap R)$ in terms of $\|Au - b\|$.

Why consider relative error bounds?

Because they can be drastically different from regular error bounds.

Example

Suppose $A = \begin{bmatrix} I_{n-1} & 0 \\ 0 & \epsilon \end{bmatrix}$ for $0 < \epsilon \ll 1$ and we work with ℓ_1 norms. For $b \in A(\mathbb{R}^n)$ we have

$$\mathrm{dist}(u, A^{-1}(b)) \leq \frac{1}{\epsilon} \cdot \|Au - b\|.$$

Now suppose $R = \mathbb{R}^{n-1} \times \{0\}$. For $b \in A(R)$ and $u \in R$

$$\mathrm{dist}(u, A^{-1}(b) \cap R) \leq \|Au - b\|.$$

Why consider relative error bounds?

Because they can be drastically different from regular error bounds.

Example

Suppose $A = \begin{bmatrix} 1 & -1 & 0 \\ -\epsilon & -\epsilon & 1 \end{bmatrix}$ for $0 < \epsilon \ll 1$ and we work with ℓ_1 norms. For $b \in A(\mathbb{R}^3)$ we have

$$\mathsf{dist}(u, A^{-1}(b)) \leq \|Au - b\|.$$

Now suppose $R = \mathbb{R}^3_+$. For $b \in A(R)$ and $u \in R$

$$\mathsf{dist}(u, A^{-1}(b) \cap R) \leq \frac{1}{\epsilon} \cdot \|Au - b\|.$$

Relative error bound for equations (special case)

Suppose $A \in \mathbb{R}^{m\times n}$ and $R \subseteq \mathbb{R}^n$ is a closed convex cone such that $A(R) := \{Ax : x \in R\}$ is a linear subspace. Let

$$H(A|R) := \max_{\substack{y\in A(R)\\ \|y\|\le 1}} \min_{\substack{x\in R\\ Ax=y}} \|x\| = \frac{1}{\displaystyle\min_{\substack{v\in A(R),\, \|v\|^*=1\\ A^{\mathsf{T}}v-u\in R^*}} \|u\|^*}.$$

Relative error bound for equations (special case)

Suppose $A \in \mathbb{R}^{m\times n}$ and $R \subseteq \mathbb{R}^n$ is a closed convex cone such that $A(R) := \{Ax : x \in R\}$ is a linear subspace. Let

$$H(A|R) := \max_{\substack{y\in A(R)\\ \|y\|\le 1}} \min_{\substack{x\in R\\ Ax=y}} \|x\| = \frac{1}{\displaystyle\min_{\substack{v\in A(R),\, \|v\|^*=1\\ A^{\mathsf{T}}v-u\in R^*}} \|u\|^*}.$$

Proposition (Following Renegar 1995, Ramdas & P 2015)

Let $A \in \mathbb{R}^{m\times n}$ and $R \subseteq \mathbb{R}^n$ be a closed convex cone such that $A(R)$ is a linear subspace. Then for all $b \in A(R)$ and all $u \in R$

$$\mathrm{dist}(u, A^{-1}(b) \cap R) \le H(A|R) \cdot \|Au - b\|.$$

Furthermore, this bound is tight.

Geometric interpretation of $H(A|R)$ in this special case

Observe

Suppose $A \in \mathbb{R}^{m \times n}$ and $R \subseteq \mathbb{R}^n$ is a closed convex cone. Then

$$
\begin{aligned}
A(R) \text{ is a linear subspace} &\Leftrightarrow 0 \in \operatorname{relint}\left(\{Ax : x \in R, \|x\| \leq 1\}\right) \\
&\Leftrightarrow Ax = 0, x \in \operatorname{relint}(R) \text{ is feasible.}
\end{aligned}
$$

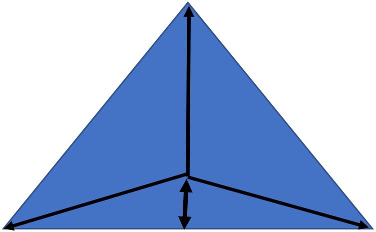
When this is the case we have

$$
\frac{1}{H(A|R)} = \operatorname{dist}\left(0, \operatorname{relbdy}(\{Ax : x \in R, \|x\| \leq 1\})\right).
$$

Example

Suppose $A = \begin{bmatrix} -1 & 1 & 0 \\ -\epsilon & -\epsilon & 1 \end{bmatrix} \in \mathbb{R}^{2\times 3}$ for $0 < \epsilon < 1/2$, both $\mathbb{R}^3$ and $\mathbb{R}^2$ are endowed with the ℓ_1 norm, and $R = \mathbb{R}^3_+$. Then

$$\frac{1}{H(A|R)} = \epsilon.$$



$\{Ax : \|x\| \leq 1\}$ and $1/H(A|R)$ for $\epsilon \in (0, 1)$.

Relative error bound for equations when R is polyhedral

Sets $\mathcal{T}(R)$ and $\mathcal{T}(A|R)$ of tangent cones

Suppose R is a polyhedron. For $u \in \mathbb{R}$, let $T_R(u)$ denote the *tangent cone* to R at $u \in R$, that is,

$$T_R(u) = \{d \in \mathbb{R}^d : u + td \in R \text{ for some } t > 0\}.$$

Let $\mathcal{T}(R) := \{T_R(u) : u \in R\}$ and

$$\mathcal{T}(A|R) := \{K \in \mathcal{T}(R) : A(K) \text{ is a linear subspace}\}.$$

Example

If $R = \mathbb{R}^n_+$ then $K \in \mathcal{T}(R)$ iff there exists $I \subseteq [n]$ such that

$$K = \{x \in \mathbb{R}^n : x_I \geq 0\}.$$

In this case $K \in \mathcal{T}(A|R)$ iff $Ax = 0, x_I > 0$ is feasible.

Relative error bound for equations when R is polyhedral

Suppose $A \in \mathbb{R}^{m \times n}$ and $R \subseteq \mathbb{R}^n$ is a reference polyhedron. Let

$$H(A|R) := \max_{K \in \mathcal{T}(A|R)} H(A|K).$$

Theorem (P, Vera, Zuluaga 2019)

Let $A \in \mathbb{R}^{m \times n}$ and $R \subseteq \mathbb{R}^n$ be a reference polyhedron. Then for all $b \in A(R)$ and all $u \in R$

$$\mathrm{dist}(u, A^{-1}(b) \cap R) \leq H(A|R) \cdot \|Au - b\|.$$

Furthermore, this bound is tight.

Relative error bound for inequalities

Consider a system of linear constraints of the form

$$\begin{aligned} Ax &\leq b \\ x &\in R \end{aligned}$$

where R is a reference set.

Recall

$P_A(b) = \{x \in \mathbb{R}^n : Ax \leq b\}$ so $P_A(b) \cap R = \{x \in R : Ax \leq b\}$.

Goal

For $u \in R$, bound $\text{dist}(u, P_A(b) \cap R)$ in terms of $\text{dist}(b - Au, \mathbb{R}^m_+)$.

Relative error bound for inequalities (special case)

Suppose $R \subseteq \mathbb{R}^n$ is a closed convex cone and $A \in \mathbb{R}^{m\times n}$ is such that $A(R) + \mathbb{R}^m_+ = \mathbb{R}^m$. Let

$$\tilde{H}(A|R) := \max_{\substack{y\in\mathbb{R}^m \\ \|y\|\le 1}} \min_{\substack{x\in R \\ Ax\le y}} \|x\| = \frac{1}{\min\limits_{\substack{v\in\mathbb{R}^m_+,\ \|v\|^*=1 \\ A^\mathsf{T} v - u \in R^*}} \|u\|^*}.$$

Proposition

Suppose $R \subseteq \mathbb{R}^n$ is a closed convex cone and $A \in \mathbb{R}^{m\times n}$ is such that $A(R) + \mathbb{R}^m_+ = \mathbb{R}^m$. Then for all $b \in \mathbb{R}^m$ and all $u \in R$

$$\mathsf{dist}(u, P_A(b) \cap R) \le \tilde{H}(A|R) \cdot \mathsf{dist}(b - Au, \mathbb{R}^m_+).$$

Furthermore this bound is tight.

Relative error bound for inequalities, R polyhedral

Suppose $A \in \mathbb{R}^{m\times n}$ and $R \subseteq \mathbb{R}^n$ is a reference polyhedron.

Recall that $[m] = \{1,\dots,m\}$ and $\mathcal{T}(R) = \{T_R(u) : u \in R\}$.

Let

$$\mathcal{J}(A|R) := \{(J,K) \in [m] \times \mathcal{T}(R) : A_J(K) + \mathbb{R}^J_+ = \mathbb{R}^J\}$$

and

$$\tilde{H}(A|R) := \max_{(J,K)\in\mathcal{J}(A|R)} \tilde{H}(A_J|K).$$

Theorem (P, Vera, Zuluaga 2019)

Let $A \in \mathbb{R}^{m\times n}$ and $R \subseteq \mathbb{R}^n$ be a reference polyhedron. Then for all $b \in A(R) + \mathbb{R}^m_+$ and all $u \in R$

$$\mathrm{dist}(u, P_A(b) \cap R) \le \tilde{H}(A|R) \cdot \mathrm{dist}(b - Au, \mathbb{R}^m_+).$$

Furthermore, this bound is tight.

Condition number relative to a reference set

Linear convergence of first-order algorithms

Consider the optimization problem

$$f^\star := \min_{x\in\mathbb{R}^n} \; f(x)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable.

Gradient descent algorithm

$$x_{k+1} = x_k - t_k \nabla f(x_k) \text{ for some } t_k > 0$$

For illustration purposes, concentrate on the least-squares function

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2$$

where $A \in \mathbb{R}^{m\times n},\; b \in \mathbb{R}^m$.

Linear convergence of first-order algorithms

Suppose $A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m$ and $f(x) = \frac{1}{2}\|Ax - b\|_2^2$. Let $f^\star := \min_{x\in\mathbb{R}^n}\ f(x)$ and $X^\star := \operatorname{argmin}_{x\in\mathbb{R}^n}\ f(x)$.

Theorem

If $t_k > 0,\ k = 0, 1, \ldots$ are judiciously chosen (e.g., via exact line-search) then the gradient descent iterates satisfy

$$\mathsf{dist}(x_k, X^\star)^2 \leq \left(1 - \frac{1}{\mathsf{Cond}(f)}\right)^k \mathsf{dist}(x_0, X^\star)^2$$

and

$$f(x_k) - f^\star \leq \left(1 - \frac{1}{\mathsf{Cond}(f)}\right)^k (f(x_0) - f^\star),$$

where

$$\mathsf{Cond}(f) = \left(\|A\| \cdot \|A^{-1}\|\right)^2.$$

Above statement holds for f convex, differentiable and suitable $\mathsf{Cond}(f)$.

Linear convergence of first-order algorithms

Suppose $A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m,\ f(x) = \frac{1}{2}\|Ax-b\|_2^2$, and $R \subseteq \mathbb{R}^n$ is a reference polyhedron.

Consider the optimization problem

$$f^\star := \min_{x\in R}\ f(x).$$

Projected gradient descent algorithm

$$x_{k+1} = \Pi_R(x_k - t_k\nabla f(x_k)).$$

Here $\Pi_R : \mathbb{R}^n \to R$ denotes the orthogonal projection on R, i.e.,

$$\Pi_R(z) = \min_{x\in R}\|z-x\|_2.$$

Linear convergence of first-order algorithms

Suppose $A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m,\ f(x) = \frac{1}{2}\|Ax-b\|_2^2$, and $R \subseteq \mathbb{R}^n$ is a polyhedron. Let $f^\star := \min_{x\in R}\ f(x),\ X^\star :=$ argmin$_{x\in R}\ f(x)$.

Proposition (Gutman-P 2019 following Necoara et al 2018)

If $t_k > 0,\ k = 0,1,\ldots$ are judiciously chosen then the projected gradient descent iterates satisfy

$$\mathsf{dist}(x_k, X^\star)^2 \le \left(1 - \frac{1}{\mathsf{Cond}(f|R)}\right)^k \mathsf{dist}(x_0, X^\star)^2$$

and

$$f(x_k) - f^\star \le \left(1 - \frac{1}{\mathsf{Cond}(f|R)}\right)^k (f(x_0) - f^\star),$$

where

$$\mathsf{Cond}(f|R) = \left(\|A|\mathsf{span}(R-R)\| \cdot H(A|R)\right)^2.$$

Gutman-P 2019: Above holds for f convex, diff and suitable $\mathsf{Cond}(f|R)$.

Cond(f) versus Cond$(f|R)$

Suppose $A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m$, and $f(x) = \frac{1}{2}\|Ax-b\|_2^2$. Recall

$$\mathsf{Cond}(f) = \left(\|A\|\cdot\|A^{-1}\|\right)^2.$$

If $R \subseteq \mathbb{R}^n$ is a reference polyhedron then

$$\mathsf{Cond}(f|R) = \left(\max_{K\in\mathcal{T}(A|R)} \|(A|K)\| \cdot \max_{K\in\mathcal{T}(A|R)} \|(A|K)^{-1}\|\right)^2,$$

where for each $K \in \mathcal{T}(A|R)$

$$\|(A|K)\| = \max_{\substack{x\in K\\ \|x\|\le 1}} \|Ax\| \ \text{ and } \ \|(A|K)^{-1}\| = \max_{\substack{y\in A(K)\\ \|y\|=1}} \min_{\substack{x\in K\\ y=Ax}} \|x\|.$$

The above expression for Cond$(f|R)$ holds because

$$\begin{aligned}\|A|\mathsf{span}(R-R)\| &= \max_{K\in\mathcal{T}(A|R)} \|(A|K)\|\\ H(A|R) &= \max_{K\in\mathcal{T}(A|R)} \|(A|K)^{-1}\|.\end{aligned}$$

Algorithms to compute $H(A)$

Computation of $H(A)$

Recall: for $A \in \mathbb{R}^{m \times n} \setminus \{0\}$

$$H(A) = \max_{J \in \mathcal{J}(A)} \frac{1}{\min\{\|A_J^\mathsf{T} v\|^* : v \in \mathbb{R}^J_+, \|v\|^* = 1\}},$$

where $\mathcal{J}(A) = \{J \subseteq [m] : A_J(\mathbb{R}^n) + \mathbb{R}^J_+ = \mathbb{R}^J\}$.

For judiciously chosen norms, the above expression for $H(A)$ can be formulated as a mixed integer linear program.

Can we do better?

Computation of $H(A)$ via the covering property

Let $\mathcal{F} \subseteq \mathcal{J}(A)$ and $\mathcal{I} \subseteq 2^{[m]} \setminus \mathcal{J}(A)$. Say that $(\mathcal{F}, \mathcal{I})$ satisfies the *covering property* if

> *For all $J \subseteq [m]$ either $J \subseteq F$ for some $F \in \mathcal{F}$ or $I \subseteq J$ for some $I \in \mathcal{I}$.*

Observation

If $(\mathcal{F}, \mathcal{I})$ satisfies the covering property then

$$H(A) = \max_{J \in \mathcal{F}} \frac{1}{\min\{\|A_J^\mathsf{T} v\|^* : v \in \mathbb{R}_+^J, \|v\|^* = 1\}}.$$

Idea for an algorithm

Gradually build $\mathcal{F} \subseteq \mathcal{J}(A)$ and $\mathcal{I} \subseteq 2^{[m]} \setminus \mathcal{J}(A)$ until $(\mathcal{F}, \mathcal{I})$ satisfies the covering property. Compute $H(A)$ via above formula.

Computation of $H(A)$ via the *covering property*

Key step: suppose J is not covered by $(\mathcal{F},\mathcal{I})$ and

$$v := \text{argmin}\{\|A_J^{\mathsf{T}}v\|^* : v \in \mathbb{R}_+^J, \|v\|^* = 1\}.$$

Observe that $\|A_J^{\mathsf{T}}v\|^* > 0 \Leftrightarrow J \in \mathcal{J}(A)$. Let $I(v) := \{i \in J : v_i > 0\}$.

Algorithm 1 Computation of $(\mathcal{F},\mathcal{I})$ and $H(A)$

Let $\mathcal{F} := \emptyset,\ \mathcal{I} := \emptyset, H(A) := 0$
while $(\mathcal{F},\mathcal{I})$ does not satisfy the covering property **do**
    Pick $J \in 2^{[m]}$ maximal not covered by $(\mathcal{F},\mathcal{I})$
    Let $v := \text{argmin}\{\|A_J^{\mathsf{T}}v\|^* : v \in \mathbb{R}_+^J, \|v\|^* = 1\}$.
    if $\|A_J^{\mathsf{T}}v\|^* > 0$ **then**
        $\mathcal{F} := \mathcal{F} \cup \{J\}$ and $H(A) := \max\left\{H(A), \frac{1}{\|A_J^{\mathsf{T}}v\|^*}\right\}$
    else
        Let $\mathcal{I} := \mathcal{I} \cup \{I(v)\}$
    end if
end while

Proposition

Let $A \in \mathbb{R}^{m\times n} \setminus \{0\}$ *and let* $\overline{\mathcal{J}}(A) \subseteq \mathcal{J}(A)$ *denote the maximal sets in* $\mathcal{J}(A)$ *and let* $\underline{\mathcal{J}}(A) \subseteq 2^{[m]} \setminus \mathcal{J}(A)$ *denote the minimal sets in* $2^{[m]} \setminus \mathcal{J}(A)$*. Algorithm 1 terminates after*

$$|\overline{\mathcal{J}}(A)| + |\underline{\mathcal{J}}(A)|$$

iterations. Upon termination Algorithm 1 returns $\mathcal{F} = \overline{\mathcal{J}}(A)$ *and* $\mathcal{I} = \underline{\mathcal{J}}(A)$*.*

Observe

Algorithm 1 terminates quickly if $\overline{\mathcal{J}}(A)$ has few and large sets.
Most favorable case: $\overline{\mathcal{J}}(A) = \{[m]\} \rightsquigarrow$ one iteration.
Next most favorable case: $\underline{\mathcal{J}}(A) = \{[m]\} \rightsquigarrow m+1$ iterations.

Examples

For convenience, suppose $\mathbb{R}^n$ and $\mathbb{R}^m$ are endowed with the ℓ_1-norm and ℓ_∞-norm respectively.

Example 1 (box)

For $A = \begin{bmatrix} \mathsf{I}_n \\ -\mathsf{I}_n \end{bmatrix} \in \mathbb{R}^{2n\times n}$ we have $H(A) = n$.

Example 2 (simplex)

For $A = \begin{bmatrix} \mathbf{1}_n^\mathsf{T} \\ -\mathsf{I}_n \end{bmatrix} \in \mathbb{R}^{(n+1)\times n}$ we have $H(A) = 2n + 1$.

Example 3 (ℓ_1-unit ball)

Let $A \in \mathbb{R}^{2^n\times n}$ be the matrix whose rows are the vectors with entries in $\{1, -1\}$. We computed the following values of $H(A)$:

n	1	2	3	4	5	6	7
$H(A)$	1	1	3	5	9	17	33

$\rightsquigarrow H(A) = 2^{n-2} + 1$ for $n \geq 3$?

Conclusions

- Error bounds ("almost" implies "near") for
  - linear equations $Ax = b$
  - linear inequalities $Ax \leq b$
  - linear equations/inequalities relative to a reference set
- Error bound constant in all cases is something like $\|A^{-1}\|$.
- Condition number of a function relative to a set.
- Algorithms to compute error bound constants.

Main references

- Peña, Vera, and Zuluaga (2019), New characterizations of Hoffman constants for systems of linear constraints, *ArXiv.*
- Gutman and Peña (2019), The condition number of a function relative to a set, *ArXiv.*
- Necoara, Nesterov, and Glineur (2018), Linear convergence of first-order methods for non-strongly convex optimization, *Math Prog.*
- Hoffman (1952), On approximate solutions of systems of linear inequalities, *J. of Research of National Bureau of Standards.*