The condition number of a function relative to a set

Javier Peña, Carnegie Mellon University
(joint work with D. Gutman, Texas Tech)

Variational Analysis and Optimization Webinar
March, 2021

Preamble: classical condition number

The condition number of a matrix

Suppose $A \in \mathbb{R}^{n \times n}$ is non-singular. The condition number of A is

$$\|A\| \cdot \|A^{-1}\| = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}.$$

This quantity is related to properties of the problem

$$Ax = b.$$

More generally, for $A \in \mathbb{R}^{m \times n}$ the condition number

$$\frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

is related to properties of the problem

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2.$$

The condition number of a function

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function and consider

$$\min_{x \in \mathbb{R}^n} f(x).$$

Condition number of f

$$\mathsf{Cond}(f) := \frac{L_f}{\mu_f}.$$

Smoothness and strong convexity constants

$$L_f := \sup_{\substack{y,x \in \mathbb{R}^n \\ y \neq x}} \frac{D_f(y,x)}{\|y-x\|^2/2}, \quad \mu_f := \inf_{\substack{y,x \in \mathbb{R}^n \\ y \neq x}} \frac{D_f(y,x)}{\|y-x\|^2/2}$$

Bregman distance

$$D_f(y,x) := f(y) - f(x) - \langle \nabla f(x), y - x \rangle.$$

Geometric intuition

Example

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

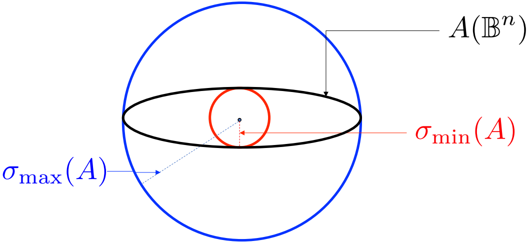
$$L_f = \sigma_{\max}(A)^2 = \min\{r : A(\mathbb{B}^n) \subseteq r\mathbb{B}^m\}^2$$

and

$$\mu_f = \sigma_{\min}(A)^2 = \max\{r : r\mathbb{B}^m \subseteq A(\mathbb{B}^n)\}^2$$

for $\mathbb{B}^d = \{u \in \mathbb{R}^d : \|u\|_2 \leq 1\}$.

Thus $\text{Cond}(f) = (\sigma_{\max}(A)/\sigma_{\min}(A))^2 = (\text{aspect ratio of } A(\mathbb{B}^n))^2$.



Linear convergence of gradient descent

Consider the optimization problem

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

Gradient descent algorithm

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

Theorem

If $\alpha_k = 1/L_f, \; k = 0, 1, \ldots$ *then gradient descent iterates satisfy*

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

and

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

Agenda

- Relative condition number
- Linear convergence of first-order methods:
  Mirror Descent and Frank-Wolfe
- Bounds and geometric intuition

Relative condition number

Our main goal

Suppose $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a differentiable convex function and $X \subseteq \text{dom}(f)$ is a convex set.

Construct a condition number for

$$\min_{x \in X} f(x).$$

Incorporate reference set X and distance $D : X \times X \to \mathbb{R}_+$.

The reference distance allows us to give a *non-Euclidean* construction.

Reference set and reference distance

Blanket assumption

The triple (f, X, D) satisfies

- $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is convex and differentiable on the *reference convex set* $X \subseteq \mathrm{dom}(f)$.
- $D : X \times X \to \mathbb{R}_+$ is a *reference distance function* such that $D(y, x) = 0$ if and only if $x = y$.

Key object

Let $Z_{f,X} : X \rightrightarrows X$ be defined as

$$Z_{f,X}(y) := \{x \in X : f(z) = f(y) \text{ for all } z \in [x, y]\}.$$

Condition number relative to a reference set and distance

Smoothness constant relative to (X, D)

$$L_{f,X,D} := \sup_{\substack{y,x \in X \\ x \neq y}} \frac{D_f(y,x)}{D(y,x)}$$

Strong convexity constant relative to (X, D)

$$\mu_{f,X,D} := \inf_{\substack{y,x \in X \\ x \notin Z_{f,X}(y)}} \frac{D_f(Z_{f,X}(y),x)}{D(Z_{f,X}(y),x)}$$

Observation

If $X = \mathbb{R}^n$, $D(y,x) = \|y-x\|^2/2$, and f is strictly convex then

$$L_{f,X,D} = L_f \text{ and } \mu_{f,X,D} = \mu_f.$$

Recover the classical condition number costruction.

Examples of relative condition numbers

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

Example 1

If $X \subseteq \mathbb{R}^n$ linear subspace and $D(y, x) = \|y - x\|_2^2/2$ then

$$L_{f,X,D} = \sigma_{\max}(A|X)^2 \text{ and } \mu_{f,X,D} = \sigma_{\min}^+(A|X)^2.$$

Here $A|L =$ restriction of A to L and $\sigma_{\min}^+(A|X) =$ its smallest *positive* singular value.

It is evident that

$$L_{f,X,D} \leq L_f \text{ and } \mu_{f,X,D} \geq \mu_f.$$

Furthermore, $L_{f,X,D}/\mu_{f,X,D}$ can be arbitrarily smaller than L_f/μ_f.

Examples of relative condition numbers (continued)

Example 2

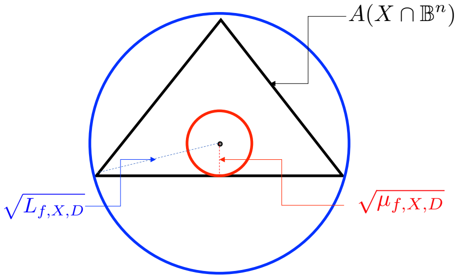
If $X \subseteq \mathbb{R}^n$ is a convex cone, $D(y,x) = \|y-x\|^2/2$, and $L := A(X)$ is a linear subspace of $\mathbb{R}^m$ then

$$L_{f,X,D} = \left(\max\{r : A(\mathbb{B}^n \cap \mathsf{span}(X)) \subseteq r\mathbb{B}^m \cap L\}\right)^2$$

and

$$\mu_{f,X,D} = \left(\min\{r : r\mathbb{B}^m \cap L \subseteq A(\mathbb{B}^n \cap X)\}\right)^2,$$

where $\mathbb{B}^d := \{u \in \mathbb{R}^d : \|u\| \leq 1\}$.



Examples of reference distance functions

Squared norm

$$D(y,x) = \frac{\|y-x\|^2}{2}.$$

Bregman distance

$$D_h(y,x) = h(y) - h(x) - \langle \nabla h(x), y - x \rangle$$

for some reference differentiable convex function $h : X \to \mathbb{R}$.

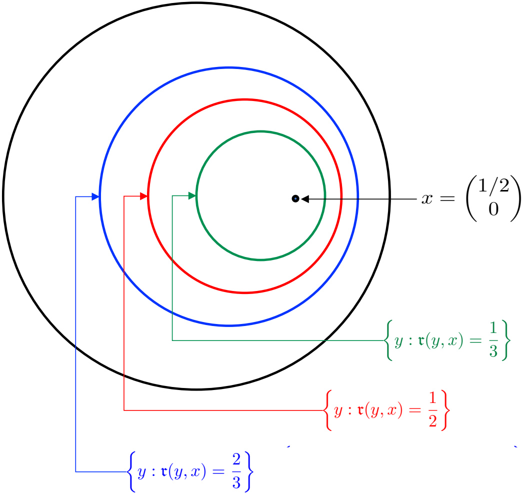
Radial distance

Suppose X is bounded. Let $\mathfrak{R} = \mathfrak{r}^2/2$ where

$$\mathfrak{r}(y,x) := \inf\{\rho > 0 : y - x = \rho(u - x) \text{ for some } u \in X\}.$$

Geometric intuition of radial distance

Level sets of $\mathfrak{r}(\cdot, x)$ for $X = \{x \in \mathbb{R}^2 : \|x\|_2 \leq 1\}$



Convergence of first-order methods

Constrained convex minimization

Suppose (f, X, D) satisfies the blanket assumption and consider the optimization problem

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

Let $X^\star := \{x \in X : f(x) = f^\star\}$.

Recall:

$$L_{f,X,D} = \sup_{\substack{y,x \in X \\ x \neq y}} \frac{D_f(y,x)}{D(y,x)} \quad \text{and} \quad \mu_{f,X,D} = \inf_{\substack{y,x \in X \\ x \notin Z_{f,X}(y)}} \frac{D_f(Z_{f,X}(y),x)}{D(Z_{f,X}(y),x)}$$

where $Z_{f,X}(y) := \{x \in X : f(z) = f(y) \text{ for all } z \in [x,y]\}$.

Mirror descent

Suppose $h : X \to \mathbb{R}$ is reference differentiable convex function.

Mirror descent algorithm

$$x_{k+1} = \operatorname*{argmin}_{y \in X} \{ \langle \nabla f(x_k), y - x_k \rangle + L_k D_h(y, x_k) \}$$

Theorem (Gutman & P 2019, following Teboulle 2018)

Suppose $L_k := L_{f,X,D_h} < \infty$ and $\mu_{f,X,D_h} > 0$. Then the mirror descent iterates satisfy

$$D_h(X^\star, x_k) \leq \left(1 - \frac{\mu_{f,X,D_h}}{L_{f,X,D_h}}\right)^k D_h(X^\star, x_0)$$

and

$$f(x_k) - f^\star \leq L_{f,X,D_h} \left(1 - \frac{\mu_{f,X,D_h}}{L_{f,X,D_h}}\right)^k D_h(X^\star, x_0).$$

Frank-Wolfe (aka conditional gradient)

Frank-Wolfe algorithm

$$s_k = \operatorname*{argmin}_{y \in X} \langle \nabla f(x_k), y \rangle$$

$$x_{k+1} = x_k + \alpha_k (s_k - x_k),\ \alpha_k \in [0,1]$$

Theorem (Gutman & P 2019)

Suppose $L_{f,X,\mathfrak{R}} < \infty$ *and* $\mu_{f,X,\mathfrak{R}} > 0$. *For judiciously chosen* $\alpha_k \in [0,1]$ *the Frank-Wolfe iterates satisfy*

$$f(x_k) - f^\star \leq \left(1 - \frac{\mu_{f,X,\mathfrak{R}}}{L_{f,X,\mathfrak{R}}}\right)^k (f(x_0) - f^\star).$$

Jaggi's curvature constant of f on X is precisely $L_{f,X,\mathfrak{R}}$.

Bounds and geometric intuition

Bounds on $L_{f,X,D}$ and $\mu_{f,X,D}$ when $f = g \circ A$

To ease exposition, consider special case $D(y, x) := \|y - x\|^2/2$.

Suppose $A \in \mathbb{R}^{m \times n}$. For $X \subseteq \mathbb{R}^n$ nonempty let

$$Z_{A,X}(y) := \{x \in X : Ax = Ay\}.$$

For a convex cone $C \subseteq \mathbb{R}^n$ let $A|C : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be defined via

$$x \mapsto (A|C)(x) := \begin{cases} \{Ax\} & \text{if } x \in C \\ \emptyset & \text{otherwise} \end{cases}$$

and let $(A|C)^{-1} : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be its inverse. Let

$$\|A|C\| := \sup_{\substack{x \in C \\ \|x\| \le 1}} \|Ax\|, \quad \|(A|C)^{-1}\| := \sup_{\substack{v \in A(C) \\ \|v\| \le 1}} \inf_{\substack{x \in C \\ Ax = v}} \|x\|.$$

Upper bound on $L_{f,X,D}$ (easy)

Suppose $f = g \circ A$ for $A \in \mathbb{R}^{m\times n}$ and $g : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$.

Proposition

Let $X \subseteq \mathrm{dom}(f)$ be convex. If g is L_g-smooth then

$$L_{f,X,D} \leq L_g \cdot \|A|\mathrm{span}(X - X)\|^2.$$

This bound is tight: if $g(v) = \|v\|_2^2/2$ then

$$L_{f,X,D} = \|A|\mathrm{span}(X - X)\|^2.$$

Recall:

$$L_{f,X,D} = \sup_{\substack{y,x\in X\\ x\neq y}} \frac{D_f(y,x)}{D(y,x)} \quad \text{and} \quad \mu_{f,X,D} = \inf_{\substack{y,x\in X\\ x\notin Z_{f,X}(y)}} \frac{D_f(Z_{f,X}(y),x)}{D(Z_{f,X}(y),x)}$$

where $Z_{f,X}(y) := \{x \in X : f(z) = f(y) \text{ for all } z \in [x,y]\}$.

Lower bound on $\mu_{f,X,D}$ (more interesting)

Suppose $f = g \circ A$ for $A \in \mathbb{R}^{m \times n}$ and $g : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$.

Theorem (Gutman & P. 2019)

Let $X \subseteq \text{dom}(f)$ be a convex cone such that $A(X)$ is a linear subspace.

If g is μ_g-strongly convex on $A(X)$ then

$$\mu_{f,X,D} \geq \frac{\mu_g}{\|(A|X)^{-1}\|^2}.$$

This bound is tight: if $g(v) = \|v\|_2^2/2$ then

$$\mu_{f,X,D} = \frac{1}{\|(A|X)^{-1}\|^2}.$$

Observe: when X is a convex cone $A(X)$ is a linear subspace iff

$$Ax = 0, \ x \in \text{ri}(X) \text{ is feasible.}$$

Geometric intuition

Suppose X is a convex cone and $A(X) = \mathbb{R}^m$. Then

$$\|A|X\| = \min\{r : A(X \cap \mathbb{B}^n) \subseteq r\mathbb{B}^m\}$$

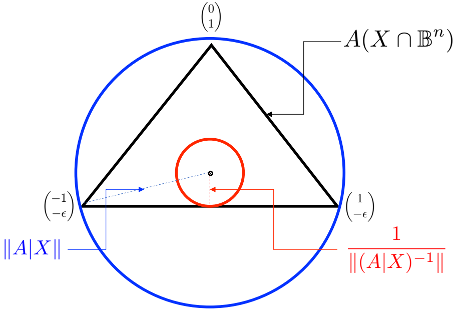
and

$$\|(A|X)^{-1}\| = \max\{r : r\mathbb{B}^m \subseteq A(X \cap \mathbb{B}^n)\}$$

Example

Let $A := \begin{bmatrix} 1 & -1 & 0 \\ -\epsilon & -\epsilon & 1 \end{bmatrix}$ for $0 < \epsilon < 1$ and $X = \mathbb{R}^3_+$.

Let $\mathbb{R}^2$ and $\mathbb{R}^3$ be endowed with the ℓ_2 and ℓ_1 norms respectively.



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

Suppose $X \subseteq \mathbb{R}^n$ is a nonempty polyhedron.

Let $\mathcal{T}(X) := \{T_X(x) : x \in X\}$, where

$$T_X(x) := \{d \in \mathbb{R}^n : x + td \in X \ \text{ for some } \ t > 0\}.$$

For $A \in \mathbb{R}^{m \times n}$ let

$$\mathcal{T}(A|X) := \{C \in \mathcal{T}(X) : A(C) \text{ is a subspace and } C \text{ is minimal}\}.$$

Example

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

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

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

Lower bound on $\mu_{f,X,D}$ for f of the form $g \circ A$ (again)

Suppose $f = g \circ A$ for $A \in \mathbb{R}^{m \times n}$ and $g : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$.

Theorem (Gutman & P. 2019)

Let $X \subseteq \text{dom}(f)$ be a nonempty polyhedron.

If g is μ_g-strongly convex on $A(X)$ then

$$\mu_{f,X,D} \geq \min_{C \in \mathcal{T}(A|X)} \frac{\mu_g}{\|(A|C)^{-1}\|^2}.$$

This bound is tight: if $g(v) = \|v\|_2^2/2$ then

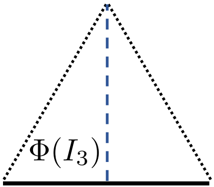
$$\mu_{f,X,D} = \min_{C \in \mathcal{T}(A|X)} \frac{1}{\|(A|C)^{-1}\|^2}.$$

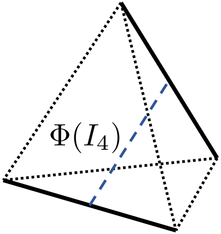
Geometric intuition

Let $X = \Delta_{n-1} := \{x \in \mathbb{R}^n_+ : \|x\|_1 = 1\}$ and endow $\mathbb{R}^n$ with ℓ_1 norm. Then

$$\min_{C \in \mathcal{T}(A|X)} \frac{1}{\|(A|C)^{-1}\|} = \frac{\Phi(A)}{2}$$

for "facial distance" $\Phi(A) = \min_{\substack{F \in \mathsf{faces}(\mathsf{conv}(A)) \\ \emptyset \neq F \neq \mathsf{conv}(A)}} \mathsf{dist}(F, \mathsf{conv}(A \setminus F))$





Conclusions

- Condition number of f relative to a reference set and distance pair (X, D) via relative constants $L_{f,X,D}$ and $\mu_{f,X,D}$.
- Convergence of first-order methods in terms of relative condition number.
- Bound when $f = g \circ A$ and X is a polyhedron:

$$\frac{L_{f,X,D}}{\mu_{f,X,D}} \leq \frac{L_g}{\mu_g} \cdot \left(\max_{C \in \mathcal{T}(X)} \|A|C\| \cdot \min_{C \in \mathcal{T}(A|X)} \|(A|C)^{-1}\| \right)^2$$

- Other related developments:
  - Frank-Wolfe algorithm with away steps
  - Refinements of relative strong convexity: relative quasi-strong convexity and relative functional growth

Main reference

Gutman and P. *"The condition number of a function relative to a set,"* Mathematical Programming, 2020.