Condition numbers and conditioning in two-stage stochastic programming

W. Römisch

Humboldt-University Berlin
Institute of Mathematics

http://www.math.hu-berlin.de/~romisch

(K. Emich, R. Henrion (WIAS Berlin))

Stevens Institute of Technology, Hoboken, November 6, 2018

Contents

To our knowledge there is only one paper on conditioning in stochastic programming:
A. Shapiro, T. Homem-de-Mello and J. Kim: Conditioning of convex piecewise linear stochastic programs, Math. Progr. 94 (2002), 1–19.

General definition of a condition number

(Bürgisser-Cucker 2013)

Let a mapping $\varphi : \mathcal{D} \subseteq \mathbb{R}^m \to \mathbb{R}^q$ be given, where the (data) set $\mathcal{D}$ is open.

The condition number of φ is defined by

$$\operatorname{cond}_\varphi(d) = \lim_{\delta \to 0} \sup_{\operatorname{rel\,err}(d) \leq \delta} \frac{\operatorname{rel\,err}(\varphi(d))}{\operatorname{rel\,err}(d)}$$

or to avoid the limit by the estimate

$$\operatorname{rel\,err}(\varphi(d)) \leq \operatorname{cond}_\varphi(d) \operatorname{rel\,err}(d) + o(\operatorname{rel\,err}(d)),$$

where $\operatorname{rel\,err}(d) := \frac{\|\tilde{d} - d\|}{\|d\|}$ for some $\tilde{d} \in \mathcal{D}$ etc.

The condition number of an input is the worst possible magnification of the output error with respect to a small input perturbation.
On the other hand, it provides information on the distance to the nearest ill-posed problem.

Linear systems

We set for $r, s \in [1, \infty]$ and $A \in \mathbb{R}^{n\times m}$

$$\|A\|_{rs} = \max_{\|x\|_r=1} \|Ax\|_s.$$

For $m = n$ let Σ denote the set of ill-posed matrices, i.e.,

$$\Sigma = \{A \in \mathbb{R}^{m\times m} : A \text{ is not invertible}\},$$

and for all $A \in \mathcal{A} = \mathbb{R}^{m\times m} \setminus \Sigma$ Turing's condition number

$$\kappa_{rs} = \|A\|_{rs}\|A^{-1}\|_{sr}.$$

Distance to ill-posedness:

$$d_{sr}(A, \Sigma) = \inf\{\|A - B\|_{rs} : B \in \Sigma\}$$

Theorem: (Eckart-Young 1936)
Let $A \in \mathbb{R}^{m\times m} \setminus \Sigma$. Then it holds

$$d_{sr}(A, \Sigma) = \|A^{-1}\|_{sr}^{-1} \quad \text{and, hence,} \quad \kappa_{rs}(A) = \frac{\|A\|_{rs}}{d_{sr}(A, \Sigma)}$$

Matrices in $\mathbb{R}^{n\times m}$:

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

$$\kappa_{rs}(A) = \|A\|_{rs}\,\|A^+\|_{sr}$$

is Turing's condition number, where $A^+ \in \mathbb{R}^{m\times n}$ is the Moore-Penrose inverse of A.

Let $\Sigma = \{A \in \mathbb{R}^{n\times m} : \operatorname{rank}(A) < \min\{n, m\}\}$ be the set of ill-posed matrices.

Proposition:

For $A \in \mathbb{R}^{n\times m} \setminus \Sigma$ it holds

$$d(A, \Sigma) = \sigma_{\min}(A) = \|A^+\|^{-1} = \sup\{\delta > 0 : \delta\mathbb{B}_n \subseteq A(\mathbb{B}_m)\},$$

where $\mathbb{B}_m$ and $\mathbb{B}_n$ are the closed unit balls in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, w.r.t. $\|\cdot\|_2$ and $\sigma_{\min}(A)$ the smallest positive singular value of A.

Polyhedral conic systems

For $A \in \mathbb{R}^{n\times m}$ and a closed convex cone $K \subseteq \mathbb{R}^m$ with polar cone $K^\star$ we consider the homogeneous primal and dual feasibility problem.

$$\exists x \in \mathbb{R}^m \setminus \{0\} \qquad Ax = 0, \quad x \in K\,, \tag{PF}$$

$$\exists y \in \mathbb{R}^n \setminus \{0\} \qquad A^\top y \in K^\star\,. \tag{DF}$$

We assume $n \leq m$ and define

$$\begin{aligned} \mathcal{P} &= \{A \in \mathbb{R}^{n\times m} : A(K) = \mathbb{R}^n\}, \\ \mathcal{D} &= \{A \in \mathbb{R}^{n\times m} : A^\top \mathbb{R}^n + K^\star = \mathbb{R}^m\}, \\ \Sigma &= \mathbb{R}^{n\times m} \setminus (\mathcal{P} \cup \mathcal{D}) \text{ is the set of ill-posed matrices.} \end{aligned}$$

Proposition:

$A \in \mathcal{P}$ iff $\{x \in \mathbb{R}^m : Ax = b, x \in K\} \neq \emptyset$ for every $b \in \mathbb{R}^n$.

$A \in \mathcal{D}$ iff $\{y \in \mathbb{R}^n : c - A^\top y \in K^\star\} \neq \emptyset$ for every $c \in \mathbb{R}^m$.

If $n < m$ then both $\mathcal{P}$ and $\mathcal{D}$ are open and $\mathcal{P} \cap \mathcal{D} = \emptyset$.

Definition: (Renegar 95)

The condition number of the homogeneous conic system with respect to K given by $A \in \mathbb{R}^{n\times m} \setminus \Sigma$ is defined by

$$\mathrm{cond}(A) = \frac{\|A\|_{rs}}{d_{rs}(A,\Sigma)}.$$

Condition number of the inhomogeneous conic system with respect to K:

$$\mathrm{cond}(A,b,c) = \max\left\{\mathrm{cond}(A,-b), \mathrm{cond}\begin{pmatrix} A \\ -c^\top \end{pmatrix}\right\}.$$

Proposition: (Renegar 95)

If $A \in \mathcal{P}$ then $d_{rs}(A,\Sigma) = \sup\{\delta > 0 : \delta\mathbb{B}_n \subseteq A(\mathbb{B}_m \cap K)\}$.

If $A \in \mathcal{D}$ then $d_{rs}(A,\Sigma) = \sup\{\delta > 0 : \delta\mathbb{B}_m \subseteq A^\top\mathbb{B}_n + K^\star\}$.

Here, $\mathbb{B}_n$ and $\mathbb{B}_m$ are the unit ball w.r.t. $\|\cdot\|_s$ in $\mathbb{R}^n$ and $\|\cdot\|_r$ in $\mathbb{R}^m$, respectively.

Conditioning of set-valued mappings and equations

(Dontchev-Rockafellar 2004, 2014)

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite-dimensional normed spaces, $F : \mathcal{X} \times \mathcal{D} \rightrightarrows \mathcal{Y}$ and consider a parametric generalized equation

$$0 \in F(x, d)\,.$$

Then $F(\cdot, d)^{-1}(y)$ is the solution set of the parametric generalized equation $y \in F(x, d)$. Next we fix d and consider $F = F(\cdot, d)$.

F is metrically regular at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ if there is a constant $\kappa > 0$ together with neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that

$$(\star) \qquad d(x, F^{-1}(y)) \leq \kappa\, d(y, F(x)) \quad \text{for all } (x, y) \in U \times V\,.$$

The **condition number** of $\bar{y} \in F(\bar{x})$ is the regularity modulus defined by

$$\operatorname{cond}(F) = \operatorname{reg} F(\bar{x}|\bar{y}) = \inf\{\kappa : \kappa \text{ satisfies condition } (\star)\}.$$

F^{-1} has the Aubin property at $(\bar{y}, \bar{x}) \in \operatorname{gph} F^{-1}$ iff F is metrically regular at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ and it holds

$$\operatorname{lip} F^{-1}(\bar{y}|\bar{x}) = \operatorname{reg} F(\bar{x}|\bar{y}).$$

F^{-1} is said to be calm at $(\bar{y}, \bar{x}) \in \operatorname{gph} F^{-1}$ iff F is metrically subregular at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ iff there is a constant $\kappa > 0$ along with a neighborhood U of $\bar{x}$ such that

$$d(x, F^{-1}(\bar{y})) \leq \kappa d(\bar{y}, F(x)) \quad \text{for all } x \in U.$$

Radius of metric regularity of F at $\bar{x}$ for $\bar{y}$: (Dontchev-Lewis-Rockafellar 2003)

$$\operatorname{rad} F(\bar{x}|\bar{y}) = \inf_{\substack{G:X\to Y \\ G(\bar{x})=0}} \{\operatorname{lip} G(\bar{x}) : F + G \text{ is not metrically regular at } (\bar{x}, \bar{y} + G(\bar{x}))\},$$

where $\operatorname{lip} G(\bar{x}) = \limsup_{\substack{x,x'\to\bar{x} \\ x\neq x'}} \frac{|G(x)-G(x')|}{\|x-x'\|}$.

Proposition: (Rockafellar-Wets 98)
Let $F : \mathcal{X} \rightrightarrows \mathcal{Y}$ be locally closed at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$. Then

$$\operatorname{rad} F(\bar{x}|\bar{y}) = \frac{1}{\operatorname{reg} F(\bar{x}|\bar{y})} \quad \text{and} \quad \operatorname{reg} F(\bar{x}|\bar{y}) = \sup_{x\in\mathbb{B}} \sup_{y\in D^\star F(\bar{x}|\bar{y})^{-1}(x)} \|y\|.$$

where $D^\star F(\bar{x}|\bar{y}) : \mathcal{Y}^\star \to \mathcal{X}^\star$ is the Mordukhovich coderivative, i.e.,

$$D^\star F(\bar{x}|\bar{y})(y^\star) = \left\{x^\star : (x^\star, -y^\star) \in N_{\operatorname{gph} F}(\bar{x}, \bar{y})\right\}.$$

Parametric convex differentiable program with polyhedral constraints

$$\min\{f(x,d) : x \in X\} \quad (d \in \mathcal{D})$$

and the optimality condition in form of a parametric set-valued equation

$$0 \in F(x,d) = \nabla f(x,d) + N_X(x).$$

with the solution mapping $S(d) = \{x \in X : 0 \in \nabla f(x,d) + N_X(x)\}$ for $d \in \mathcal{D}$.

We know that the conditioning of the program is characterized by

$$\operatorname{lip} S\left(\bar{d}|\bar{x}\right) = \sup_{x^* \in \mathbb{B}} \sup_{p^* \in D^* S\left(\bar{d}|\bar{x}\right)(x^*)} \|p^*\|,$$

Proposition: (see also Mordukhovich 06)
Let $(\bar{d}, \bar{x}) \in \operatorname{gph} S$ with $\bar{d} \in \mathcal{D}$ and $\bar{x} \in X$. Assume that the multifunction

$$y \mapsto \{(d,x) : y \in \nabla f(x,d) + N_X(x)\}$$

is calm at $(0, \bar{d}, \bar{x})$. Then it holds

$$D^\star S(\bar{d}|\bar{x})(x^\star) \subseteq \big\{p^\star : \exists v^\star \text{ with } (-x^\star, p^\star) \in (D^\star \nabla) f(\bar{x}, \bar{d})(v^\star) + D^\star N_X(\bar{x}, -\nabla f(\bar{x}, \bar{d}))(v^\star)\big\}.$$

Computing $\partial^2 f = (D^\star \nabla) f$ and $D^\star N_X$

Proposition:
Let $f(v) = \sup_{z\in Z}\langle v, z\rangle - \frac{1}{2}\langle Bz, z\rangle$ $(v \in \mathbb{R}^k)$ with $B \in \mathbb{R}^{k\times k}$ symmetric and positive definite, and Z convex polyhedral. Then

$$\partial^2 f(\bar{v})(w^\star) = \left\{z^\star \in \mathbb{R}^k : Bz^\star - w^\star \in D^\star N_Z(z(\bar{v}), \bar{v} - Bz(\bar{v}))(-z^\star)\right\},$$

where $z(\bar{v})$ is the unique solution of the maximum problem defining $f(\bar{v})$.

Proposition: (Henrion-Römisch 07)
Consider the polyhedron $P = \{u : Au \leq b\}$, N_P the normal cone mapping to P and fix $(\bar{u}, \bar{w}) \in \operatorname{gph} N_P$. Denote by $I = \{i : \langle a_i, \bar{u}\rangle = b_i$ the index set of active rows of A at $\bar{u}$ and assume that these active rows are linearly independent. Moreover, let $J = \{i \in I : \lambda_i > 0\}$ be the index set of strictly positive multipliers, where λ is the unique solution of $\sum_{i\in I} \lambda_i a_i = \bar{w}$. Then

$$D^\star N_P(\bar{u}, \bar{w})(s^\star) = \begin{cases} \operatorname{pos}\{a_i : i \in I, \langle a_i, s^\star\rangle > 0\} + \operatorname{span}\{a_i : i \in I, \langle a_i, s^\star\rangle = 0\} & \text{if } s^\star \in \cap_{i\in J} a_i^\perp, \\ \emptyset & \text{else.} \end{cases}$$

Linear-quadratic two-stage stochastic optimization problems

$$\min \left\{ \langle c, x\rangle + \tfrac{1}{2}\langle x, Cx\rangle + \mathbb{E}\left(\Phi(x,\xi)\right) \middle| x \in X \right\},$$

where x is the first-stage decision and

$$\Phi(x,\xi) = \max_{z\in Z} \left\{ \langle z, h(\xi) - Tx\rangle - \tfrac{1}{2}\langle z, Bz\rangle \right\}.$$

We assume that X and Z are nonempty convex polyhedra in $\mathbb{R}^m$ and $\mathbb{R}^k$, respectively, B and C are symmetric positive definite matrices, $c \in \mathbb{R}^m$, $h(\xi)$ is a random vector in $\mathbb{R}^k$, T a $k \times m$ matrix, Z is of the form $Z = \{z \in \mathbb{R}^r : W^\top z \leq q\}$ with a $k \times r$ matrix W and $q \in \mathbb{R}^r$, and $\mathbb{E}$ denotes expectation with respect to a probability distribution P on $\mathbb{R}^s$.

Here, we assume that P is a discrete probability distribution of the form

$$P = \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi^i}$$

with scenarios $\xi^i \in \mathbb{R}^s$, $i = 1, \ldots, n$.

Aim: Conditioning of the two-stage model with respect to P.

So, we have $d = (\xi^1, \ldots, \xi^n) \in \mathbb{R}^{ns}$ and

$$f(x,d) = \langle c, x\rangle + \tfrac{1}{2}\langle x, Cx\rangle + \mathbb{E}_P\left(\Phi(x,\xi)\right).$$

Proposition:
The function $f(\cdot, d)$ is Frechet differentiable and its gradient locally Lipschitz continuous, but, in general, not twice differentiable.

Proposition: Let $(\bar{d}, \bar{x}) \in \operatorname{gph} S$, T be surjective and $h(\xi) = H\xi + \bar{h}$. Then

$$\operatorname{lip} S\left(\bar{d}|\bar{x}\right) = \sup_{x^* \in \mathbb{B}} \sup_{p^* \in D^*S\left(\bar{d}|\bar{x}\right)(x^*)} \|p^*\|,$$

where $D^*S\left(\bar{d}|\bar{x}\right)(x^*) \subseteq$

$$\left\{ p^* \left| \begin{array}{l} \exists v^*, \exists u^* \in D^*N_X(\bar{x}, -c - C\bar{x} + n^{-1}T^\top \sum_{i=1}^n z(\bar{v}_i))\,(v^*) \\ \exists z_i^* : Bz_i^* + Tv^* \in D^*N_Z(z(\bar{v}_i), \bar{v}_i - Bz(\bar{v}_i))(-z_i^*) \quad (i = 1, \ldots, n) \\ n^{-1}T^\top \sum\limits_{i=1}^n z_i^* = C^\top v^* + x^* + u^* \\ p_i^* = n^{-1}H^\top z_i^*,\ \bar{v}_i = H\bar{\xi^i} + \bar{h} - T\bar{x} \quad (i = 1, \ldots, n) \end{array} \right. \right\}$$

with $z(v) = \arg\max_{z \in Z}\{\langle z, v\rangle - \frac{1}{2}\langle z, Bz\rangle\}$.

Special case: $C = \sigma I$, $B = \tau I$ and $Z = [-q^-, q^+]$ (simple recourse), where $\sigma > 0$, $\tau > 0$.

Theorem:

Assume that strict complementarity holds at $\bar{x}$. Let T be surjective and let σ and τ satisfy

$$\sigma \tau > n^{-1} \triangle(T, \bar{d}, \bar{x}) \|T\| \,.$$

Then the condition number $\operatorname{lip} S(\bar{d}|\bar{x})$ can be estimated by

$$\operatorname{lip} S(\bar{d}|\bar{x}) \leq \frac{\|H\|}{[\triangle(T, \bar{d}, \bar{x})]^{-1} n\sigma\tau - \|T\|} \,,$$

where $\triangle(T)$ is defined by

$$\triangle(T, \bar{d}, \bar{x}) = \sum_{i=1}^{n} \triangle_i(T, \bar{\xi}^i, \bar{x}) \,, \qquad (\triangle_i(T, \bar{\xi}^i, \bar{x}))^2 = \sum_{\substack{j=1 \\ z_j(H\bar{\xi}^i + \bar{h} - T\bar{x}) \\ \text{is not active in } Z}}^{r} \|t_j\|^2$$

with t_j denoting the rows of T. Note that $n^{-1}\triangle(T, \bar{d}, \bar{x})$ refers to the mean number of non strongly active components of $z(H\bar{\xi}^i + \bar{h} - T\bar{x})$, $i = 1, \ldots, n$.

Conclusions

- Characterization of the condition number in the general two-stage case is open. Which quantities influence its size and what are the consequences of large condition numbers ? Of course, the Lipschitz constants of the second-stage solution mapping

$$v \mapsto z(v) = \arg\max_{z \in Z} \left\{ \langle z, v \rangle - \tfrac{1}{2} \langle z, Bz \rangle \right\}$$

  become important.

- The relations to the results in (Shapiro–Homem-de-Mello–Kim 02) and in the recent paper (Zolezzi 15) need to be explored.

- Extension of the results to more general linear-quadratic two-stage models and to linear two-stage models are desirable, but not straightforward. In the linear case, uniqueness of solutions and, hence, differentiability of the recourse function is lost in general.

- Extension of characterizing the conditioning by considering **metric subregularity** instead of metric regularity is of interest.

References:

P. Bürgisser and F. Cucker: *Condition*, Springer, Berlin, 2013.

A. L. Dontchev, A. S. Lewis and R. T. Rockafellar: Radius of metric regularity, *Trans. Amer. Math. Soc.* 355 (2003), 493–517.

A. L. Dontchev and R. T. Rockafellar: Regularity and conditioning of solution mappings in variational analysis, Set-Valued Analysis 12 (2004), 79–109.

A. L. Dontchev and R. T. Rockafellar: *Implicit Functions and Solution Mappings* (Second Edition), Springer, New York, 2014.

K. Emich, R. Henrion and W. Römisch: Conditioning of linear-quadratic two-stage stochastic optimization problems, *Math. Progr.* 148 (2014), 201–221.

B. S. Mordukhovich and R. T. Rockafellar: Second-order subdifferential calculus with application to tilt stability in optimization, *SIAM J. Optim.* 22 (2012), 953–986.

J. Renegar: Incorporating condition measures into the complexity theory of linear programming, *SIAM J. Optim.* 5 (1995), 506–524.

J. Renegar: Linear programming, complexity theory and elementary functional analysis, *Math. Progr.* 70 (1995), 279–351.

A. Shapiro, T. Homem-de-Mello and J. Kim: Conditioning of convex piecewise linear stochastic programs, *Math. Progr.* 94 (2002), 1–19.

T. Zolezzi: Condition number theorems in optimization, *SIAM J. Optim.* 14 (2003), 507–516.

T. Zolezzi: A condition number theorem in convex programming, *Math. Progr.* 149 (2015), 195–207.