Intermezzo II: The Condition of the Condition

How costly is it to compute a condition number? This question presents two aspects: computational cost and accuracy. We begin by briefly discussing the first of these aspects. To do so, we recall a few of the condition numbers we have met thus far.

Take matrix–vector multiplication. We analyzed this problem in Sect. O.4, where we proved (Proposition O.8) that the normwise condition number $\mathsf{cond}(A,x)$ for this problems satisfies

$$\mathsf{cond}(A,x) = \frac{\|A\|_\infty \|x\|_\infty}{\|Ax\|_\infty}.$$

The denominator on the right-hand side indicates that to compute $\mathsf{cond}(A,x)$, we need at least to compute Ax, that is, to solve the problem for which (A,x) is the input data.

Consider now matrix inversion. Its normwise condition number (for the norms $\|\ \|_{rs}$ and $\|\ \|_{sr}$ in data and solution space, respectively) is, as we proved in Theorem 1.5,

$$\kappa_{rs}(A) = \|A\|_{rs} \left\|A^{-1}\right\|_{sr}.$$

Again, it is apparent that for computing $\kappa_{rs}(A)$ one needs to solve the problem for which A is the data, i.e., inverting A.

Finally, consider the condition number $\mathscr{C}(A)$ for PCFP. All its characterizations, in Sect. 6.5 via smallest including caps, in Sect. 6.6 via images of balls, and in Sect. 6.7 via well-conditioned solutions, turn into computations of $\mathscr{C}(A)$ that require, among other things, the solution of PCFP for input A.

It would seem that invariably, to compute $\mathsf{cond}^\varphi(a)$ we need to compute $\varphi(a)$. This is not true. The function $\varphi(a) = a^k$ satisfies $\mathsf{cond}^\varphi(a) = k$ for all $a \neq 0$; it is thus trivially computed. Yet the cost of computing $\varphi(a)$ can be bounded by the cost of computing $\mathsf{cond}^\varphi(a)$ plus a constant. The emerging picture can be thus summarized as follows:

The cost of computing $\mathsf{cond}^\varphi(a)$ *is, modulo an additive constant, at least the cost of computing* $\varphi(a)$. *That is,* $\mathsf{cost}(\mathsf{cond}^\varphi) \geq \mathsf{cost}(\varphi) + \mathcal{O}(1)$.

P. Bürgisser, F. Cucker, *Condition*,
Grundlehren der mathematischen Wissenschaften 349,
DOI 10.1007/978-3-642-38896-5,

The nature of this statement makes it difficult to formally prove it. We will therefore refrain from continuing and leave the statement as an empirical conclusion.

We can now proceed with the second aspect mentioned above. The accuracy in the computation of $\mathsf{cond}^\varphi(a)$ depends on the algorithm used to compute $\mathsf{cond}^\varphi(a)$ as well as on the condition of a for the function $\mathsf{cond}^\varphi : \mathcal{D} \subseteq \mathbb{R}^m \to [0, \infty)$. Disregarding the former, the question is posed, what is the condition number of condition number computation? This "condition of the condition" is called *level-2 condition number*.

In this intermezzo we give an answer for a large class of condition numbers. We say that a condition number cond^φ is *à la Renegar* when there exists a $\Sigma \subset \mathbb{R}^m$, $\Sigma \neq \emptyset$, such that for all $a \in \mathcal{D} \subseteq \mathbb{R}^m$,

$$\mathsf{cond}^\varphi(a) = \frac{\|a\|}{\mathsf{dist}(a, \Sigma)}. \qquad \text{(II.1)}$$

Here $\| \ \|$ is an arbitrary norm in $\mathbb{R}^m$ and dist is the distance induced by that norm. As we have seen, several condition numbers have this form (or are well approximated by expressions of this form). Furthermore (cf. Sect. 6.1), expression (II.1) is the definition of choice for condition numbers of discrete-valued problems (e.g., $\mathscr{C}(A)$) when the set of ill-posed inputs is clear.

Denote by $\mathsf{cond}^\varphi_{[2]}(a)$ the normwise (for the norm $\| \ \|$) condition number of the function cond^φ. Our main result is the following.

Theorem II.1 *Let φ be any problem and let cond^φ be given by* (II.1). *Then*

$$\mathsf{cond}^\varphi(a) - 1 \le \mathsf{cond}^\varphi_{[2]}(a) \le \mathsf{cond}^\varphi(a) + 1.$$

Proof To simplify notation, let $\varrho(a) = \mathsf{dist}(a, \Sigma)$. For all input data a,

$$\begin{aligned}
\mathsf{cond}^\varphi_{[2]}(a) &= \lim_{\delta\to 0} \sup_{\|\Delta a\|\le\delta\|a\|} \frac{|\mathsf{cond}^\varphi(a+\Delta a) - \mathsf{cond}^\varphi(a)|\|a\|}{\mathsf{cond}^\varphi(a)\|\Delta a\|} \\
&= \lim_{\delta\to 0} \sup_{\|\Delta a\|\le\delta\|a\|} \frac{|\frac{\|a+\Delta a\|}{\varrho(a+\Delta a)} - \frac{\|a\|}{\varrho(a)}|\|a\|}{\frac{\|a\|}{\varrho(a)}\|\Delta a\|} \\
&= \lim_{\delta\to 0} \sup_{\|\Delta a\|\le\delta\|a\|} \left| \frac{\|a+\Delta a\|\varrho(a) - \|a\|\varrho(a+\Delta a)}{\varrho(a+\Delta a)\|\Delta a\|} \right|. \qquad \text{(II.2)}
\end{aligned}$$

To prove the upper bound, note that for every perturbation Δa,

$$\big| \|a+\Delta a\| - \|a\| \big| \le \|\Delta a\|$$

and

$$\big| \varrho(a+\Delta a) - \varrho(a) \big| \le \|\Delta a\|.$$

Therefore,

$$\big| \|a+\Delta a\|\varrho(a) - \|a\|\varrho(a) \big| \le \|\Delta a\|\varrho(a)$$

and

$$\big|\|a\|\varrho(a+\Delta a)-\|a\|\varrho(a)\big|\leq\|a\|\|\Delta a\|.$$

It follows that

$$\big|\|a+\Delta a\|\varrho(a)-\|a\|\varrho(a+\Delta a)\big|\leq\|\Delta a\|\varrho(a)+\|a\|\|\Delta a\|$$

and consequently that for sufficiently small Δa,

$$\left|\frac{\|a+\Delta a\|\varrho(a)-\|a\|\varrho(a+\Delta a)}{\varrho(a+\Delta a)\|\Delta a\|}\right|\leq\frac{\|\Delta a\|\varrho(a)+\|a\|\|\Delta a\|}{(\varrho(a)-\|\Delta a\|)\|\Delta a\|}=\frac{\varrho(a)+\|a\|}{\varrho(a)-\|\Delta a\|}.$$

Now use this inequality together with (II.2) to obtain

$$\begin{aligned}\mathsf{cond}^{\varphi}_{[2]}(a)&=\lim_{\delta\to 0}\sup_{\|\Delta a\|\leq\delta\|a\|}\left|\frac{\|a+\Delta a\|\varrho(a)-\|a\|\varrho(a+\Delta a)}{\varrho(a+\Delta a)\|\Delta a\|}\right|\\&\leq\lim_{\delta\to 0}\sup_{\|\Delta a\|\leq\delta\|a\|}\frac{\varrho(a)+\|a\|}{\varrho(a)-\|\Delta a\|}\\&=\frac{\varrho(a)+\|a\|}{\varrho(a)}=1+\frac{\|a\|}{\varrho(a)}=1+\mathsf{cond}^{\varphi}(a).\end{aligned}$$

This proves the upper bound. We now proceed with the lower bound.

Let Δa^* be such that $\varrho(a)=\|\Delta a^*\|$ and $a+\Delta a^*\in\Sigma$. For any $\varepsilon\in\mathbb{R}$ satisfying $0<\varepsilon<\|\Delta a^*\|$ let

$$\Delta a^*_\varepsilon=\frac{\varepsilon}{\varrho(a)}\Delta a^*.$$

Then, $\|\Delta a^*_\varepsilon\|=\varepsilon$ and $\varrho(a+\Delta a^*_\varepsilon)=\varrho(a)-\|\Delta a^*_\varepsilon\|=\varrho(a)-\varepsilon$ and therefore

$$\begin{aligned}\left|\frac{\|a+\Delta a^*_\varepsilon\|\varrho(a)-\|a\|\varrho(a+\Delta a^*_\varepsilon)}{\varrho(a+\Delta a^*_\varepsilon)\|\Delta a^*_\varepsilon\|}\right|&=\left|\frac{\|a+\Delta a^*_\varepsilon\|\varrho(a)-\|a\|(\varrho(a)-\varepsilon)}{(\varrho(a)-\varepsilon)\varepsilon}\right|\\&\geq\frac{(\|a\|-\|\Delta a^*_\varepsilon\|)\varrho(a)-\|a\|(\varrho(a)-\varepsilon)}{(\varrho(a)-\varepsilon)\varepsilon}\\&=\frac{(\|a\|-\varepsilon)\varrho(a)-\|a\|(\varrho(a)-\varepsilon)}{(\varrho(a)-\varepsilon)\varepsilon}\\&=\frac{-\varepsilon\varrho(a)+\|a\|\varepsilon}{(\varrho(a)-\varepsilon)\varepsilon}=\frac{\|a\|-\varrho(a)}{\varrho(a)-\varepsilon}.\end{aligned}$$

Again, use this inequality together with (II.2) to obtain

$$\begin{aligned}\mathsf{cond}^{\varphi}_{[2]}(a)&=\lim_{\delta\to 0}\sup_{\|\Delta a\|\leq\delta\|a\|}\left|\frac{\|a+\Delta a\|\varrho(a)-\|a\|\varrho(a+\Delta a)}{\varrho(a+\Delta a)\|\Delta a\|}\right|\\&\geq\lim_{\delta\to 0}\frac{\|a\|-\varrho(a)}{\varrho(a)-\delta\|a\|}=\frac{\|a\|-\varrho(a)}{\varrho(a)}=\mathsf{cond}^{\varphi}(a)-1.\end{aligned}$$

This proves the lower bound. □

Remark II.2 The bounds in Theorem II.1 are sharp, as shown by the following toy example. Consider φ to be the problem of deciding whether a point $x \in \mathbb{R}$ is greater than a fixed value $\xi > 0$. Then $\Sigma = \{\xi\}$, and for $x \in \mathbb{R}$, $x > 0$, Eq. (II.1) yields

$$\mathsf{cond}^{\varphi}(x) = \begin{cases} \frac{x}{x-\xi} & \text{if } x > \xi, \\ \frac{x}{\xi-x} & \text{if } x < \xi, \\ \infty & \text{if } x = \xi. \end{cases}$$

Since cond^{φ} is differentiable at x for $x \neq \xi$, we have (compare Proposition 14.1)

$$\mathsf{cond}_{[2]}(x) = \left| \frac{d}{dx} \mathsf{cond}^{\varphi}(x) \right| \frac{x}{|\mathsf{cond}^{\varphi}(x)|} = \begin{cases} \frac{\xi}{x-\xi} & \text{if } x > \xi, \\ \frac{\xi}{\xi-x} & \text{if } x < \xi. \end{cases}$$

Now note that $\frac{x}{x-\xi} = \frac{\xi}{x-\xi} + 1$ and $\frac{x}{\xi-x} = \frac{\xi}{\xi-x} - 1$.

Another simple example shows that a result like Theorem II.1 (actually, even a version with multiplicative constants) may fail to hold for condition numbers not having a characterization of the form (II.1). Consider the problem $\varphi : \mathbb{R} \to \mathbb{R}$ given by $\varphi(x) = x^2 + x + c$, for some $c \in \mathbb{R}$. For $x \in \mathbb{R}$, let $\mathsf{cond}^{\varphi}(x)$ be its condition number, as defined in (O.1). Since φ is differentiable on $\mathbb{R}$, we have

$$\mathsf{cond}^{\varphi}(x) = \frac{|x\varphi'(x)|}{|\varphi(x)|}$$

and, assuming $x\varphi'(x), \varphi(x) > 0$,

$$\mathsf{cond}^{\varphi}_{[2]}(x) = \left| \left(\frac{x\varphi'(x)}{\varphi(x)} \right)' \right| \frac{|\varphi(x)|}{|\varphi'(x)|} = \frac{|x\varphi''(x)\varphi(x) + \varphi'(x)\varphi(x) - x(\varphi'(x))^2|}{|\varphi(x)\varphi'(x)|}.$$

Now take $x = 1$ and $c > -2$ (so that $x, \varphi(x), \varphi'(x) > 0$). Then

$$\mathsf{cond}^{\varphi}(a) = \frac{3}{2+c}$$

and

$$\mathsf{cond}^{\varphi}_{[2]}(a) = \frac{|5c+1|}{3(2+c)}.$$

When $c \to \infty$ we have $\mathsf{cond}^{\varphi}(a) \to 0$ and $\mathsf{cond}^{\varphi}_{[2]}(a) \to \frac{5}{3}$, while for $c = -\frac{1}{5}$ we have $\mathsf{cond}^{\varphi}(a) = \frac{5}{3}$ and $\mathsf{cond}^{\varphi}_{[2]}(a) = 0$.