6

A Condition Number for Polyhedral Conic Systems

Consider the problem φ that maps any pair (b, c) of real numbers to the number of real roots of the polynomial $f = X^2 + bX + c$. Since the possible values for this problem are the elements in $\{0, 1, 2\}$ the set of inputs is partitioned as $\mathcal{D}_0 \cup \mathcal{D}_1 \cup \mathcal{D}_2$ with $\mathcal{D}_i = \{(b, c) \in \mathbb{R}^2 \mid \varphi(b, c) = i\}$. We know that

$$\begin{aligned} \mathcal{D}_2 &= \{(b, c) \in \mathbb{R}^2 \mid b^2 > 4c\}, \\ \mathcal{D}_1 &= \{(b, c) \in \mathbb{R}^2 \mid b^2 = 4c\}, \\ \mathcal{D}_0 &= \{(b, c) \in \mathbb{R}^2 \mid b^2 < 4c\} \end{aligned}$$

so that $\dim(\mathcal{D}_2) = \dim(\mathcal{D}_0) = 2$ and $\dim(\mathcal{D}_1) = 1$. Actually, the boundaries $\partial\mathcal{D}_2$ and $\partial\mathcal{D}_0$ are the same and coincide with the parabola $\mathcal{D}_1$.

What is the, say normwise, condition number for this problem? If $(b, c) \in \mathcal{D}_2$ then all sufficiently small perturbations $(\tilde{b}, \tilde{c})$ of (b, c) will also be in $\mathcal{D}_2$. Hence, for these perturbations $\mathsf{RelError}(\varphi(b, c)) = 0$ and therefore $\mathsf{cond}(b, c) = 0$. A similar argument shows the same equality when $(b, c) \in \mathcal{D}_0$. In contrast, when $(b, c) \in \mathcal{D}_1$ one can find arbitrarily small perturbations $(\tilde{b}, \tilde{c})$ in $\mathcal{D}_2$ as well as arbitrarily small perturbations in $\mathcal{D}_0$. Therefore, for these perturbations the quotient $\frac{\mathsf{RelError}(\varphi(\tilde{b},\tilde{c}))}{\mathsf{RelError}(b,c)}$ can be arbitrarily large and it follows that $\mathsf{cond}(b, c) = \infty$ when $(b, c) \in \mathcal{D}_1$.

No matter whether for complexity or for finite-precision analysis it is apparent that $\mathsf{cond}(b, c)$ cannot be of any relevance.

The problem considered above has no computational mysteries. We have chosen it simple for illustration purposes. The discussion above, notwhistanding, will carry over to any counting problem (one with values in $\mathbb{N}$) and, with the appropriate modifications, to any decisional problem (one with values in $\{\mathtt{Yes}, \mathtt{No}\}$). For these problems a different development is needed.

Firstly, a different format for finite-precision analysis appears to be a must, the one discussed in Chapter 1 making no sense in this context. The relevant question is no longer how many correct significant figures are lost in the computation but rather how many do we need to start with (i.e., how small should ϵ_{mach} be) to ensure a correct output.

Secondly, a different way of measuring condition, appropriate for the goal

just described, should be devised. One also expects such a measure to be of use in complexity analyses.

In this chapter we begin the development of these ideas. We do so based on a particular problem, the feasibility of polyhedral conic systems. But, unlike the exposition of the previous chapters, we will first use condition for complexity analyses and only at the end of our development, in Section 8.6, we will discuss finite-precision analysis.

6.1 The polyhedral conic system feasibility problem

For $A \in \mathbb{R}^{m\times n}$, consider the *primal feasibility problem*

$$\text{(PF)} \qquad \exists x \in \mathbb{R}^n \setminus \{0\} \quad Ax = 0\,,\ x \geq 0$$

and the *dual feasibility problem*

$$\text{(DF)} \qquad \exists y \in \mathbb{R}^m \setminus \{0\} \quad A^{\mathrm{T}} y \leq 0.$$

We say that A is *primal feasible* or *dual feasible* when (PF), or (DF), respectively, are satisfied. In both cases we talk about *strict* feasibility when the satisfied inequality is strict. The following result shows that strict primal feasibility and strict dual feasibility are incompatible. To simplify its statement we introduce some notation. Let $\mathcal{F}_P$ and $\mathcal{F}_D$ denote the set of matrices A where (PF) and (DF) are satisfied, respectively. Moreover, let

$$\begin{array}{rcl} \mathcal{F}_P^\circ & = & \{A \in \mathbb{R}^{m\times n} \mid \exists x \in \mathbb{R}^n\ Ax = 0\,,\ x > 0\}, \\ \mathcal{F}_D^\circ & = & \{A \in \mathbb{R}^{m\times n} \mid \exists y \in \mathbb{R}^m\ A^{\mathrm{T}} y < 0\} \end{array}$$

be the sets of strictly primal and strictly dual feasible matrices. Finally, let $\mathcal{R} := \{A \in \mathbb{R}^{m\times n} \mid \mathsf{rank}\, A = m\}$ and

$$\Sigma := \mathcal{F}_P \cap \mathcal{F}_D.$$

Denote by $\mathsf{int}(M)$, $\overline{M}$, and $\partial M = \overline{M} \setminus \mathsf{int} M$, the interior, closure and boundary of a subset M of Euclidean space.

Proposition 6.1. *Both $\mathcal{F}_P$ and $\mathcal{F}_D$ are closed subsets of $\mathbb{R}^{m\times n}$. In addition, this space is partitioned as follows*

$$\mathbb{R}^{m\times n} = \mathsf{int}(\mathcal{F}_P^\circ) \cup \mathsf{int}(\mathcal{F}_D^\circ) \cup \Sigma$$

and we have

$$\Sigma = \partial\mathcal{F}_P = \partial\mathcal{F}_D.$$

Furthermore, when $n > m$, $\mathcal{F}_D^\circ = \mathsf{int}(\mathcal{F}_D)$, $\mathcal{F}_P^\circ \supseteq \mathsf{int}(\mathcal{F}_P)$, and $\mathcal{F}_P = \overline{\mathcal{F}_P^\circ \cap \mathcal{R}}$.

One can easily show that if $n \leq m$ then $\mathcal{F}_D = \mathbb{R}^{m\times n}$. The situation of interest is therefore the case where $n > m$. For this case Figure 6.1 provides a schematic picture derived from Proposition 6.1. On it, the 2-dimensional space corresponds to the set of all matrices. The curve corresponds to the set Σ which is divided into two parts. All matrices there are in $\mathcal{F}_D \setminus \mathcal{F}_D^\circ$: those on the full part of the curve correspond to matrices in $\mathcal{F}_P \setminus \mathcal{F}_P^\circ$ and those on the dashed part to (rank-defficient) matrices in $\mathcal{F}_P^\circ$. The set Σ, just as in the picture, is of dimension smaller than mn.

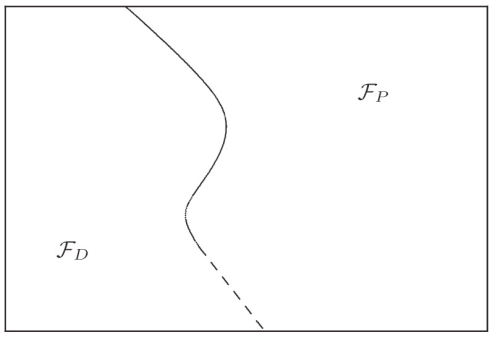

Figure 6.1: A partition of $\mathbb{R}^{m\times n}$ with respect to feasibility.

We see that for matrices in Σ arbitrary small perturbations can lead to a change with respect to feasibility. In contrast, in the set $\mathcal{D} = \mathbb{R}^{m\times n} \setminus \Sigma$ the following problem is well-defined:

$$\textit{Given } A \in \mathcal{D} \textit{ decide whether } A \in \mathcal{F}_P^\circ \textit{ or } A \in \mathcal{F}_D^\circ.$$

We call this the *polyhedral conic system feasibility problem* (and we denote it by PCSFP). For all $A \in \Sigma$ the problem is ill-posed.

The polyhedral conic system feasibility problem fits the situation described in the introduction of this chapter. The approach to condition described in Chapter 1 cannot be applied here (note that even the values of this problem — the tags "strictly primal feasible" and "strictly dual feasible"— are not elements in a Euclidean space). We need a different measure of condition. We will define this measure in the next section. Before doing so, however, we will prove Proposition 6.1 and get some understanding about the partition depicted in Figure 6.1.

We start with some preliminaries about convex sets. A subset $K \subseteq \mathbb{R}^m$ is called *convex* when

$$\forall x, y \in K\ \forall t \in [0,1]\ tx + (1-t)y \in K.$$

That is, K contains the line segment with endpoints x, y for all $x, y \in K$. The *convex hull* of a set of points $a_1, \ldots, a_n \in \mathbb{R}^m$ is defined as

$$\mathsf{conv}\{a_1, \ldots, a_n\} := \left\{ \sum_{i=1}^n t_i a_i \;\middle|\; t_1, \ldots, t_n \geq 0\,,\ \sum_{i=1}^n t_i = 1 \right\}.$$

This is easily seen to be the smallest convex set containing $a_1, \ldots, a_n$. The *affine hull* of $a_1, \ldots, a_n$ is defined as

$$\mathsf{aff}\{a_1, \ldots, a_n\} := \left\{ \sum_{i=1}^n t_i a_i \;\middle|\; t_1, \ldots, t_n \in \mathbb{R}\,,\ \sum_{i=1}^n t_i = 1 \right\}.$$

This is the smallest affine subspace of $\mathbb{R}^m$ containing $a_1, \ldots, a_n$.

We state without proof the following result due to Carathéodory.

Theorem 6.2. *Let $a_1, \ldots, a_n \in \mathbb{R}^m$ with d-dimensional affine hull. Then for any $x \in \mathsf{conv}\{a_1, \ldots, a_n\}$ there exists $I \subseteq [n]$ with $|I| \leq d+1$ such that $x \in \mathsf{conv}\{a_i \mid i \in I\}$.* □

Corollary 6.3. *Assume that I is as in Theorem 6.2 with minimal cardinality. Then the affine hull of $\{a_i \mid i \in I\}$ must have dimension $k = |I| - 1$, that is $(a_i)_{i \in I}$ are affinely independent.*

Proof. If we had $k < |I| - 1$ then Theorem 6.2 applied to the subset $\{a_i \mid i \in I\}$ would yield the existence of $J \subseteq I$ with $x \in \mathsf{conv}\{a_j \mid j \in J\}$ and $|J| \leq k + 1 < |I|$, which contradicts the minimality of I. □

We define the *relative interior* of $K = \mathsf{conv}\{a_1, \ldots, a_n\}$ by

$$\mathsf{relint}(\mathsf{conv}\{a_1, \ldots, a_n\}) := \left\{ \sum_{i=1}^n t_i a_i \;\middle|\; t_1, \ldots, t_n > 0\,,\ \sum_{i=1}^n t_i = 1 \right\}.$$

One can show that this set can be intrinsically characterized by

$$\mathsf{relint}(K) = \{a \mid \exists \varepsilon > 0\, \forall a' \in A : \|a' - a\| < \varepsilon \Rightarrow a' \in K\},$$

where $A = \mathsf{aff}\{a_1, \ldots, a_n\}$.

The *separating hyperplane theorem* is a fundamental result in convexity theory. We will use the following of its versions.

Theorem 6.4. *Let $K \subseteq \mathbb{R}^m$ be closed and convex. For $p \notin K$ there exists $y \in \mathbb{R}^m \setminus \{0\}$ and $\lambda \in \mathbb{R}$ such that*

$$\forall x \in K \;\; \langle x, y \rangle < \lambda < \langle p, y \rangle \quad \textit{(strict separation).}$$

If $p \in \partial K$ there exists $y \in \mathbb{R}^m \setminus \{0\}$ such that

$$\forall x \in K \;\; \langle x, y \rangle \leq \langle p, y \rangle \quad \textit{(supporting hyperplane).}$$ □

A *closed halfspace* $H \subset \mathbb{R}^m$ is a set $H = \{z \in \mathbb{R}^m \mid \langle z, y\rangle \leq 0\}$ for some $y \in \mathbb{R}^m \setminus \{0\}$. Similarly, we say that $H^\circ = \{z \in \mathbb{R}^m \mid \langle z, y\rangle < 0\}$ is an *open halfspace*.

We can now proceed with the analysis of the feasibility regions towards the proof of Proposition 6.1. We begin with a simple result.

Proposition 6.5. *Both $\mathcal{F}_P$ and $\mathcal{F}_D$ are closed subsets of $\mathbb{R}^{m\times n}$ and closed under multiplication with scalars $\lambda_i \geq 0$: $[a_1, \ldots, a_n] \in \mathcal{F}_P \Rightarrow [\lambda_1 a_1, \ldots, \lambda_n a_n] \in \mathcal{F}_P$ and similarly for $\mathcal{F}_D$.*

Proof. Let $\mathbb{S}^{m-1} := \{y \in \mathbb{R}^m \mid \|y\| = 1\}$ denote the $(m-1)$-dimensional unit sphere. The compactness of $\mathbb{S}^{m-1}$ easily implies that

$$\mathcal{F}_D = \{A \mid \exists y \in \mathbb{S}^{m-1} \ \langle a_1, y\rangle \leq 0, \ldots, \langle a_n, y\rangle \leq 0\}$$

is closed. Similarly, one shows that $\mathcal{F}_P$ is closed. The second statement is trivial. □

Let $A \in \mathbb{R}^{m\times n}$ and denote by $a_1, \ldots, a_n \in \mathbb{R}^m$ its columns. We have the following geometric characterizations:

$$\begin{aligned} A \in \mathcal{F}_P &\Leftrightarrow 0 \in \mathsf{conv}\{a_1, \ldots, a_n\}, \\ A \in \mathcal{F}_P^\circ &\Leftrightarrow 0 \in \mathsf{relint}(\mathsf{conv}\{a_1, \ldots, a_n\}). \end{aligned} \tag{6.1}$$

Also, by definition, we have

$$\begin{aligned} A \in \mathcal{F}_D &\Leftrightarrow \exists H \text{ closed halfspace such that } \mathsf{conv}\{a_1, \ldots, a_n\} \subseteq H, \\ A \in \mathcal{F}_D^\circ &\Leftrightarrow \exists H^\circ \text{ open halfspace such that } \mathsf{conv}\{a_1, \ldots, a_n\} \subseteq H^\circ. \end{aligned}$$

From the definition of Σ and the first line in (6.1) we obtain the following characterization

$$A \in \Sigma \Leftrightarrow A \in \mathcal{F}_D \text{ and } 0 \in \mathsf{conv}\{a_1, \ldots, a_n\}. \tag{6.2}$$

Lemma 6.6. *For $A \in \mathbb{R}^{m\times n}$ we have*

1. $A \notin \mathcal{F}_D^\circ \Leftrightarrow A \in \mathcal{F}_P$
2. $A \notin \mathcal{F}_P^\circ \Rightarrow A \in \mathcal{F}_D$. *The converse is true if* $\mathsf{rank}\, A = m$.

Proof. (1) We show the contraposition. Suppose $A \in \mathcal{F}_D^\circ$. Then there exists $y \in \mathbb{R}^m \setminus \{0\}$ such that $\langle a_i, y\rangle < 0$ for all i. If we had $\sum_i x_i a_i = 0$ for some $x_i \geq 0$ with $\sum_i x_i = 1$, then $\sum_i x_i \langle a_i, y\rangle = \langle \sum_i x_i a_i, y\rangle = 0$. Hence $x_i = 0$ for all i, which is a contradiction.

Conversely, suppose that $A \notin \mathcal{F}_P$, that is, $0 \notin \mathsf{conv}\{a_1, \ldots, a_n\}$. Theorem 6.4 (strict separation) implies that $A \in \mathcal{F}_D^\circ$.

(2) Suppose $A \notin \mathcal{F}_P^\circ$. Then $0 \notin \mathsf{relint}(\mathsf{conv}\{a_1, \ldots, a_n\})$, therefore $0 \notin \mathsf{int}(\mathsf{conv}\{a_1, \ldots, a_n\})$. Theorem 6.4 implies $A \in \mathcal{F}_D$. For the other direction

assume that $A \in \mathcal{F}_D$, say $\langle a_i, y\rangle \leq 0$ for all i and some $y \neq 0$. If we had $A \in \mathcal{F}_P^\circ$, then $\sum_i x_i a_i = 0$ for some $x_i > 0$. Therefore $\sum_i x_i \langle a_i, y\rangle = 0$, hence $\langle a_i, y\rangle = 0$ for all i. This implies $\mathsf{rank}(A) \leq m-1$. □

Remark 6.7. For the converse of part (2) of Lemma 6.6 we indeed need the rank assumption. To see this take for example $a_1, \ldots, a_n \in \mathbb{R}^{m-1}$ such that $0 \in \mathsf{relint}(\mathsf{conv}\{a_1, \ldots, a_n\})$. Then $A \in \mathcal{F}_D \cap \mathcal{F}_P^\circ$.

Lemma 6.6 implies that $\mathcal{F}_P^\circ$ and $\mathcal{F}_D^\circ$ are disjoint,

$$\mathcal{F}_D \setminus \mathcal{F}_D^\circ = \Sigma\,,\ \mathcal{F}_P \setminus \mathcal{F}_P^\circ \subseteq \Sigma$$

and the right-hand inclusion becomes an equality when restricting the matrices to be of rank m. Moreover, using Lemma 6.6,

$$\mathbb{R}^{m\times n} = \mathcal{F}_P \cup \mathcal{F}_D = \mathcal{F}_P^\circ \cup \mathcal{F}_D^\circ \cup \Sigma. \tag{6.3}$$

Since Σ is closed, $\mathcal{F}_D^\circ$ is open. It is somewhat confusing that $\mathcal{F}_P^\circ$ is not open. To see this, consider again $a_1, \ldots, a_n \in \mathbb{R}^{m-1}$ such that $0 \in \mathsf{relint}(\mathsf{conv}\{a_1, \ldots, a_n\})$. Then $A \in \mathcal{F}_P^\circ$, but there are arbitrarily small perturbations of A that lie in $\mathcal{F}_D^\circ$.

Proposition 6.8. *1.* $\mathcal{F}_D \subseteq \overline{\mathcal{F}_D^\circ}$

2. If $n > m$ *then* $\mathcal{F}_P \subseteq \overline{\mathcal{F}_P^\circ \cap \mathcal{R}}$.

Proof. (1) Let $A = [a_1, \ldots, a_n] \in \mathcal{F}_D$. Hence there exists $y \in \mathbb{S}^{m-1}$ such that $\langle a_i, y\rangle \leq 0$ for all i. For $\varepsilon > 0$ put $a_i(\varepsilon) := a_i - \varepsilon y$. Then $\langle a_i(\varepsilon), y\rangle = \langle a_i, y\rangle - \varepsilon \leq -\varepsilon$, hence $A(\varepsilon) = [a_1(\varepsilon), \ldots, a_n(\varepsilon)] \in \mathcal{F}_D^\circ$. Moreover, $\lim_{\varepsilon\to 0} A(\varepsilon) = A$.

(2) Let $A = [a_1, \ldots, a_n] \in \mathcal{F}_P$. Put $W := \mathsf{span}\{a_1, \ldots, a_n\}$ and $d := \dim W + 1$. The first line in (6.1) implies that $0 \in \mathsf{conv}\{a_1, \ldots, a_n\}$. By Carathéodory's Theorem 6.2 we may assume w.l.o.g. that $0 = x_1 a_1 + \ldots + x_k a_k$ with $x_i > 0$, $\sum_{i=1}^k x_i = 1$ and $k \leq d$. Moreover, by Corollary 6.3, we may assume that the affine hull of $a_1, \ldots, a_k$ has dimension $k-1$. The affine hull equals the linear hull due to $0 \in \mathsf{conv}\{a_1, \ldots, a_n\}$. W.l.o.g. we may assume that $a_1, \ldots, a_{k-1}$ are linearly independent and that $a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_d$ is a basis of W. Let $b_{d+1}, \ldots, b_{m+1}$ be a basis of the orthogonal complement $W^\perp$. We define now

$$v(\varepsilon) := a_{k+1} + \ldots + a_d + (a_{d+1} + \varepsilon b_{d+1}) + \ldots + (a_{m+1} + \varepsilon b_{m+1}) + a_{m+2} + \ldots + a_n.$$

(Here we used the assumption $n \geq m+1$.) Moreover, we put

$$a_i(\varepsilon) := \begin{cases} a_i - \varepsilon v(\varepsilon) & \text{for } 1 \leq i \leq k \\ a_i & \text{for } k+1 \leq i \leq d \\ a_i + \varepsilon b_i & \text{for } d+1 \leq i \leq m+1 \\ a_i & \text{for } m+2 \leq i \leq n\,. \end{cases}$$

Note that $v(\varepsilon) = \sum_{i=k+1}^{n} a_i(\varepsilon)$. It is clear that $A(\varepsilon) := [a_1(\varepsilon), \ldots, a_n(\varepsilon)]$ converges to A for $\varepsilon \to 0$. Using the fact that $W = \mathsf{span}\{a_1, \ldots, a_d\}$ it follows that $\mathsf{span}\{a_1(\varepsilon), \ldots, a_n(\varepsilon)\} = \mathbb{R}^m$, i.e., that $\mathsf{rank}(A(\varepsilon)) = m$. Finally, we have

$$0 = \sum_{i=1}^{k} x_i a_i = \sum_{i=1}^{k} x_i a_i(\varepsilon) + \varepsilon v(\varepsilon) = \sum_{i=1}^{k} x_i a_i(\varepsilon) + \sum_{j=k+1}^{n} \varepsilon\, a_j(\varepsilon).$$

Hence $A(\varepsilon) \in \mathcal{F}_P^\circ$. □

Corollary 6.9. *Suppose $n > m$. Then*

1. $\Sigma = \partial\mathcal{F}_D$, $\mathsf{int}(\mathcal{F}_D) = \mathcal{F}_D^\circ$,
2. $\Sigma = \partial\mathcal{F}_P$, $\mathsf{int}(\mathcal{F}_P) \subseteq \mathcal{F}_P^\circ$.

Proof. (1) We have $\mathcal{F}_D^\circ \subseteq \mathsf{int}(\mathcal{F}_D)$ since $\mathcal{F}_D^\circ$ is open. Hence $\partial\mathcal{F}_D = \mathcal{F}_D \setminus \mathsf{int}(\mathcal{F}_D) \subseteq \mathcal{F}_D \setminus \mathcal{F}_D^\circ = \Sigma$. Suppose $A \in \Sigma$. By Proposition 6.8 there is a sequence $A_k \to A$ such that $\mathsf{rank}\, A_k = m$ and $A_k \in \mathcal{F}_P^\circ$. Lemma 6.6 shows $A_k \notin \mathcal{F}_D$. Hence $A \in \partial\mathcal{F}_D$. It follows that $\partial\mathcal{F}_D = \Sigma$ and $\mathsf{int}(\mathcal{F}_D) = \mathcal{F}_D^\circ$.

(2) Let $A \in \Sigma$. By Proposition 6.8 there is a sequence $A_k \to A$ such that $A_k \in \mathcal{F}_D^\circ$, hence $A_k \notin \mathcal{F}_P$. Therefore $A \in \partial\mathcal{F}_P$. It follows that $\Sigma \subseteq \partial\mathcal{F}_P$. On the other hand,

$$\partial\mathcal{F}_P \subseteq \overline{\mathbb{R}^{m\times n} \setminus \mathcal{F}_P} = \overline{\mathcal{F}_D^\circ} \subseteq \mathcal{F}_D,$$

hence $\partial\mathcal{F}_P \subseteq \mathcal{F}_P \cap \mathcal{F}_D = \Sigma$. It follows that $\Sigma = \partial\mathcal{F}_P$. Finally,

$$\mathsf{int}(\mathcal{F}_P) = \mathcal{F}_P \setminus \partial\mathcal{F}_P = \mathcal{F}_P \setminus \Sigma \subseteq \mathcal{F}_P^\circ.$$ □

It may seem disturbing that $\mathsf{int}(\mathcal{F}_P)$ is properly contained in $\mathcal{F}_P^\circ$. However, the difference $\mathcal{F}_P^\circ \setminus \mathsf{int}(\mathcal{F}_P)$ lies in Σ and thus has measure zero, so that this will not harm us (see Figure 6.1).

Proof of Proposition 6.1. It immediately follows from the results in this section. □

6.2 The GCC Condition Number and Distance to Ill-posedness

We want to define a condition number for PCSFP. A way of doing so relies on the Condition Number Theorem (Corollary 2.7). This result characterized the condition number of linear equation solving, or matrix inversion, as the inverse of the relativized distance from the matrix at hand to the set of ill-posed matrices. Instead of defining condition in terms of perturbations (which we have seen is now useless) we can take the characterization of the Condition Number Theorem as definition. We have shown in the previous section that for PCSFP the set of ill-posed instances is the boundary between feasible and infeasible instances. This motivates the following definition.

Definition 6.10. Let $A \in \mathbb{R}^{m\times n}$ be given with nonzero columns a_i. Suppose $A \notin \Sigma$ and $A \in \mathcal{F}_S^\circ$ for $S \in \{P, D\}$. We define

$$\Delta(A) := \sup\left\{\delta > 0 \;\middle|\; \forall A' \in \mathbb{R}^{m\times n} \;\left(\max_{i\le n} \frac{\|a_i' - a_i\|}{\|a_i\|} < \delta \;\Rightarrow\; A' \in \mathcal{F}_S^\circ\right)\right\},$$

where a_i' stands for the ith column of A'. The *GCC-condition number* of A is defined as

$$\mathscr{C}(A) := \frac{1}{\Delta(A)}.$$

If $A \in \Sigma$ we set $\Delta(A) = 0$ and $\mathscr{C}(A) = \infty$.

We note that the suprema are over nonempty bounded sets and hence well-defined, since $\mathcal{F}_S^\circ \setminus \Sigma = \mathsf{int}(\mathcal{F}_S)$ for $S \in \{P, D\}$ due to Corollary 6.9.

We have written the definition in a way so that it becomes clear that we measure the *relative* size of the perturbation for each row a_i, where the relativization is with respect to the norm of a_i. Also, it is clear from the definition that $\Delta(A)$ is scale invariant in the sense that

$$\Delta([\lambda_1 a_1, \ldots, \lambda_n a_n]) = \Delta([a_1, \ldots, a_n]) \text{ for } \lambda_i > 0.$$

For the analysis of Δ we may therefore assume, without loss of generality, that $\|a_i\| = 1$ for all i. Hence we can see the matrix A with columns $a_1, \ldots, a_n$ as an element in the product $(\mathbb{S}^{m-1})^n$ of spheres.

We now want to rewrite Definition 6.10 in a way that follows the ideas of Section 3.3. Let $d_{\mathbb{S}}(a, b) \in [0, \pi]$ denote the angular distance[1]

$$d_{\mathbb{S}}(a, b) := \arccos(\langle a, b\rangle).$$

It is clear that this defines a metric on $\mathbb{S}^{m-1}$

$$d_{\mathbb{S}}(A, B) := \max_{1\le i\le m} d_{\mathbb{S}}(a_i, b_i)$$

on $(\mathbb{S}^{m-1})^n$. For a nonempty subset $M \subseteq (\mathbb{S}^{m-1})^n$ we write

$$d_{\mathbb{S}}(A, M) := \inf\{d_{\mathbb{S}}(A, B) \mid B \in M\}.$$

For simplicity of notation, we shall denote $\mathcal{F}_P \cap (\mathbb{S}^{m-1})^n$ also by the symbol $\mathcal{F}_P$ and similarly for $\mathcal{F}_P^\circ, \mathcal{F}_D, \mathcal{F}_D^\circ$, and Σ. This should not lead to any confusion.

The fact that $\Sigma = \partial\mathcal{F}_P = \partial\mathcal{F}_D$ (cf. Corollary 6.9) immediately tells us that

$$\begin{array}{rcll} d_{\mathbb{S}}(A, (\mathbb{S}^{m-1})^n \setminus \mathcal{F}_P^\circ) & = & d_{\mathbb{S}}(A, \Sigma) & \text{for } A \in \mathcal{F}_P^\circ \\ d_{\mathbb{S}}(A, (\mathbb{S}^{m-1})^n \setminus \mathcal{F}_D^\circ) & = & d_{\mathbb{S}}(A, \Sigma) & \text{for } A \in \mathcal{F}_D^\circ. \end{array} \tag{6.4}$$

We postpone the proof of the following result (compare Theorem 6.16).

[1] The distance $d_{\mathbb{P}}$ defined in §3.1.4 is related to $d_{\mathbb{S}}$ by $d_{\mathbb{P}}(a, b) = \sin d_{\mathbb{S}}(a, b)$.

Lemma 6.11. *For $A \in (\mathbb{S}^{m-1})^n$ we have $d_{\mathbb{S}}(A, \Sigma) \leq \frac{\pi}{2}$. Moreover, $d_{\mathbb{S}}(A, \Sigma) = \frac{\pi}{2}$ iff $A = (a, a, \ldots, a)$ for some $a \in \mathbb{S}^{m-1}$.*

We can now give a geometric characterization of the GCC-condition number.

Proposition 6.12. *For $A \in (\mathbb{S}^{m-1})^n$ we have $\Delta(A) = \sin d_{\mathbb{S}}(A, \Sigma)$. Hence $\mathscr{C}(A) = \frac{1}{\sin d_{\mathbb{S}}(A,\Sigma)}$ or, equivalently,*

$$\mathscr{C}(A) = \frac{1}{d_{\mathbb{P}}(A, \Sigma)}.$$

Proof. Without loss of generality $A \notin \Sigma$. Suppose $A \in \mathcal{F}_P^\circ$. It suffices to show that

$$\begin{aligned} &(1) \quad \sin d_{\mathbb{S}}(A, \Sigma) = 1 \quad \Rightarrow \quad \Delta(A) = 1 \\ &(2) \quad \sin d_{\mathbb{S}}(A, \Sigma) < d \quad \Leftrightarrow \quad \Delta(A) < d \quad \text{for all } 0 < d < 1. \end{aligned}$$

The first case is easily established with the second part of Lemma 6.11. Thus, let $0 < d < 1$ such that $\sin d_{\mathbb{S}}(A, \Sigma) < d$. Lemma 6.11 tells us that $d_{\mathbb{S}}(A, \Sigma) \leq \frac{\pi}{2}$, hence $d_{\mathbb{S}}(A, \Sigma) < \arcsin d$. By (6.4) there exists $B = (b_1, \ldots, b_n) \notin \mathcal{F}_P^\circ$ such that $d_{\mathbb{S}}(A, B) < \arcsin d$. Additionally, we may assume that $\|b_i\| = 1$. Let $\theta_i = d_{\mathbb{S}}(a_i, b_i)$ (cf. Figure 6.2).

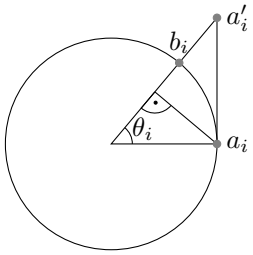

Figure 6.2: The definition of b_i.

By definition, $d_{\mathbb{S}}(A, B) = \max_i \theta_i$, hence $\theta_i < \arcsin d$ for all i and therefore

$$\|(\cos \theta_i) b_i - a_i\| = \sin \theta_i < d.$$

It follows from the definition of $\Delta(A)$ that $\Delta(A) < d$ (consider the matrix A' with the columns $(\cos \theta_i) b_i$).

Conversely, assume $\Delta(A) < d$ for $d < 1$. Then there exists $A' \notin \mathcal{F}_P^\circ$ such that $\max_i \|a_i' - a_i\| < d$. In particular, $a_i' \neq 0$. For $b_i := \frac{a_i'}{\|a_i'\|}$ we have $\theta_i := d_{\mathbb{S}}(a_i, b_i) < \frac{\pi}{2}$ and for all i

$$\sin \theta_i = \min_{\lambda > 0} \|\lambda b_i - a_i\| \leq \|a_i' - a\| < d$$

(cf. Figure 6.2). Hence $d_{\mathbb{S}}(A, B) < \arcsin d$ and therefore we have $d_{\mathbb{S}}(A, \Sigma) = d_{\mathbb{S}}(A, (\mathbb{S}^{m-1})^n \setminus \mathcal{F}_P^\circ) < \arcsin d$.

The case where $A \in \mathcal{F}_D^\circ$ is shown analogously. □

6.3 The GCC Condition Number and Spherical Caps

For $p \in \mathbb{S}^{m-1}$ and $\alpha \in [0, 2\pi]$ recall that

$$\mathsf{cap}(p, \alpha) := \{y \in \mathbb{S}^{m-1} \mid \langle p, y\rangle \geq \cos\alpha\}$$

denotes the spherical cap in $\mathbb{S}^{m-1}$ with center p and angular radius α.

Definition 6.13. A *smallest including cap* (SIC) for $A = (a_1, \ldots, a_n) \in (\mathbb{S}^{m-1})^n$ is a spherical cap of minimal radius containing the points $a_1, \ldots, a_n$. If p denotes its center, then its *blocking set* is defined as $\{i \in [n] \mid \langle a_i, p\rangle = \cos\alpha\}$ (which can be seen as the set of "active rows").

We remark that by a compactness argument, a SIC always exists. However, there may be several SICs (consider for instance three equidistant points on the circle). While a SIC for A might not be uniquely determined, its radius certainly is and will be denoted by $\rho(A)$.

Lemma 6.14. *We have $\rho(A) < \frac{\pi}{2}$ iff $A \in \mathcal{F}_D^o$. Moreover, $\rho(A) = \frac{\pi}{2}$ iff $A \in \Sigma$.*

Proof. We have $\rho(A) < \frac{\pi}{2}$ iff $a_1, \ldots, a_n$ are contained in a spherical cap of radius less than $\frac{\pi}{2}$. This means that there exists $p \in \mathbb{S}^{m-1}$ such that $\langle a_1, -p\rangle < 0, \ldots, \langle a_n, -p\rangle < 0$. This is equivalent to $A \in \mathcal{F}_D^o$. By the same reasoning, $\rho(A) \leq \frac{\pi}{2}$ is equivalent to $A \in \mathcal{F}_D$. This proves the lemma. □

Lemma 6.15. *Let $\mathsf{cap}(p, \rho)$ be a SIC for $A = (a_1, \ldots, a_n)$ with blocking set $[k+1]$. Write $t := \cos\rho$ so that*

$$\langle a_1, p\rangle = \cdots = \langle a_{k+1}, p\rangle = t\ , \quad \langle a_{k+2}, p\rangle > t\ , \ldots,\ \langle a_n, p\rangle > t.$$

Then $tp \in \mathsf{conv}\{a_1, \ldots, a_{k+1}\}$.

Proof. Suppose first that A is feasible, i.e., that $t \geq 0$. It suffices to show that $p \in \mathsf{cone}\{a_1, \ldots, a_{k+1}\}$. Indeed, if $p = \sum_{i=1}^{k+1} \lambda_i a_i$, $\lambda_i \geq 0$, then $tp = \sum_{i=1}^{k+1} t\lambda_i a_i$. Furthermore,

$$\sum_{i=1}^{k+1} t\lambda_i = \sum_{i=1}^{k+1} \lambda_i \langle a_i, p\rangle = \left\langle \sum_{i=1}^{k+1} \lambda_i a_i, p \right\rangle = \langle p, p\rangle = 1\ .$$

We argue by contradiction. If $p \notin \mathsf{cone}\{a_1, \ldots, a_{k+1}\}$, then by the separation theorem there would exist a vector $v \in \mathbb{S}^{m-1}$ such that $\langle p, v\rangle < 0$ and $\langle a_i, v\rangle > 0$ for all i. For $\delta > 0$ we set

$$p_\delta := \frac{p + \delta v}{\|p + \delta v\|} = \frac{p + \delta v}{\sqrt{1 + 2\delta\langle p, v\rangle + \delta^2}}.$$

Then for $1 \leq i \leq k+1$ and sufficiently small δ we have

$$\langle a_i, p_\delta \rangle \;=\; \frac{t + \delta \langle a_i, v \rangle}{\sqrt{1 + 2\delta \langle p, v \rangle + \delta^2}} \;>\; t.$$

Moreover, by continuity we have $\langle a_i, p_\delta \rangle > t$ for all $i > k+1$ and δ sufficiently small. We conclude that for sufficiently small $\delta > 0$ there exists $t_\delta > 0$ such that $\langle a_i, p_\delta \rangle > t_\delta$ for all $i \in [n]$. Hence $\mathsf{cap}(p_\delta, \alpha_\delta)$ is a spherical cap containing all the a_i that has angular radius $\alpha_\delta = \arccos t_\delta < \alpha$, contradicting the minimality assumption.

In the case where A is infeasible ($t < 0$) one can argue analogously. □

Theorem 6.16. *We have*

$$d_{\mathbb{S}}(A, \Sigma) \;=\; \begin{cases} \frac{\pi}{2} - \rho(A) & \text{if } A \in \mathcal{F}_D \\ \rho(A) - \frac{\pi}{2} & \text{if } A \in (\mathbb{S}^{m-1})^n \setminus \mathcal{F}_D \end{cases}.$$

In particular, $d_{\mathbb{S}}(A, \Sigma) \leq \frac{\pi}{2}$ *and*

$$\mathscr{C}(A)^{-1} = \sin d_{\mathbb{S}}(A, \Sigma) \;=\; |\cos \rho(A)| \,.$$

Proof. We first assume that $A \in \mathcal{F}_D$. Let $\mathsf{cap}(p, \rho)$ be a SIC for A and put $t := \cos \rho$. Thus $\rho \leq \frac{\pi}{2}$ and hence $t \geq 0$. Let $A' \in (\mathbb{S}^{m-1})^n$ such that $d_{\mathbb{S}}(A', A) \leq \frac{\pi}{2} - \rho$. Since $d_{\mathbb{S}}(p, a_i) \leq \rho$ for all i, we get

$$d_{\mathbb{S}}(p, a_i') \leq d_{\mathbb{S}}(p, a_i) + d_{\mathbb{S}}(a_i, a_i') \leq \rho + \frac{\pi}{2} - \rho = \frac{\pi}{2}.$$

Hence $\langle p, a_i' \rangle \geq 0$ for all i, which implies $A' \in \mathcal{F}_D$. We have thus shown the implication

$$\forall A' \quad d_{\mathbb{S}}(A', A) \leq \frac{\pi}{2} - \rho \Rightarrow A' \in \mathcal{F}_D.$$

This implies

$$d_{\mathbb{S}}(A, \Sigma) = d_{\mathbb{S}}(A, (\mathbb{S}^{m-1})^n \setminus \mathcal{F}_D) \geq \frac{\pi}{2} - \rho.$$

For the other direction, without loss of generality, let $[k+1]$ be the blocking set of $\mathsf{cap}(p, \rho)$. We have $\langle a_i, p \rangle = t$ for $i \leq k+1$, $\langle a_i, p \rangle > t$ for $i > k+1$, and $tp \in \mathsf{conv}\{a_1, \ldots, a_{k+1}\}$ by Lemma 6.15 (see Figure 6.3).

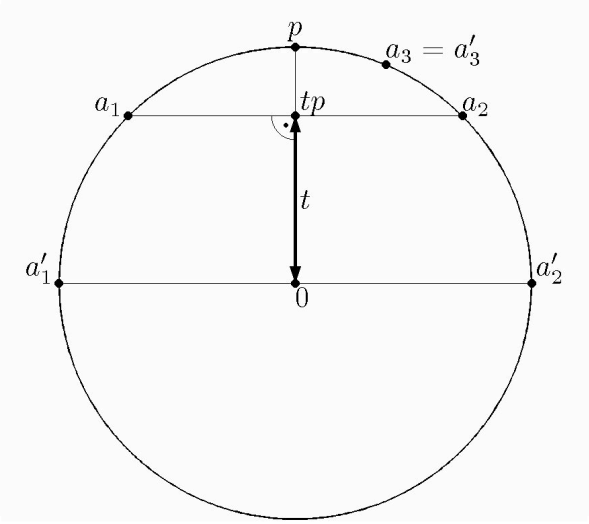

Figure 6.3: $A = (a_1, a_2, a_3) \in \mathcal{F}_D$, $A' = (a_1', a_2', a_3') \in \Sigma$, and $t = t(A)$

We assume that $a_i \neq tp$ for $i \in [k+1]$ since otherwise $a_i = tp = p$ for all $i \in [n]$ and for this case the claim is easily established. Put

$$a_i' := \begin{cases} \frac{a_i - tp}{\|a_i - tp\|} & \text{for } i \leq k+1, \\ a_i & \text{for } i > k+1. \end{cases}$$

Then $\langle a_i', p\rangle \geq 0$ for all $i \in [n]$, $\langle a_i', p\rangle = 0$ for $i \leq k{+}1$ and $0 \in \mathsf{conv}\{a_1', \ldots, a_{k+1}'\}$. Characterization (6.2) (p. 95) implies that $A' = (a_1', \ldots, a_n') \in \Sigma$. Hence

$$d_{\mathbb{S}}(A, \Sigma) \leq d_{\mathbb{S}}(A, A') \leq \frac{\pi}{2} - \rho.$$

Altogether, we have shown that $d_{\mathbb{S}}(A, \Sigma) = \frac{\pi}{2} - \rho$, which proves the assertion in the case $A \in \mathcal{F}_D$.

We assume now $A \notin \mathcal{F}_D$. Let $\mathsf{cap}(p, \rho)$ be a SIC for A. Note that for all $i \in [n]$ with $\langle a_i, p\rangle < 0$ we have $a_i \neq \langle a_i, p\rangle \cdot p$ since equality would yield a contradiction to the minimality of ρ, which is easily seen. We set

$$a_i' := \begin{cases} \frac{a_i - \langle a_i, p\rangle \cdot p}{\|a_i - \langle a_i, p\rangle \cdot p\|} & \text{if } a_i - \langle a_i, p\rangle \cdot p < 0 \\ a_i & \text{otherwise.} \end{cases}$$

As in the proof of the case $A \in \mathcal{F}_D$ we see that $A' = (a_1', \ldots, a_n') \in \Sigma$ and $d_{\mathbb{S}}(A', A) \leq \rho - \frac{\pi}{2}$. Hence

$$d_{\mathbb{S}}(A, \Sigma) \;\leq\; \rho - \frac{\pi}{2}.$$

For the other direction we need to prove that

$$\forall A' \quad \left(A' \in \mathcal{F}_D \;\Rightarrow\; d_{\mathbb{S}}(A', A) \geq \rho - \frac{\pi}{2}\right).$$

So let $A' \in \mathcal{F}_D$ and $q \in \mathbb{S}^{m-1}$ such that $A'q \leq 0$. Consider the cap of smallest angular radius α with center $-q$ that contains all the points a_i. Then $\alpha \geq \rho$. Choose i_0 such that (see Figure 6.4)

$$d_{\mathbb{S}}(a_{i_0}, q) = \max_{1 \leq i \leq n} d_{\mathbb{S}}(a_i, q) = \alpha.$$

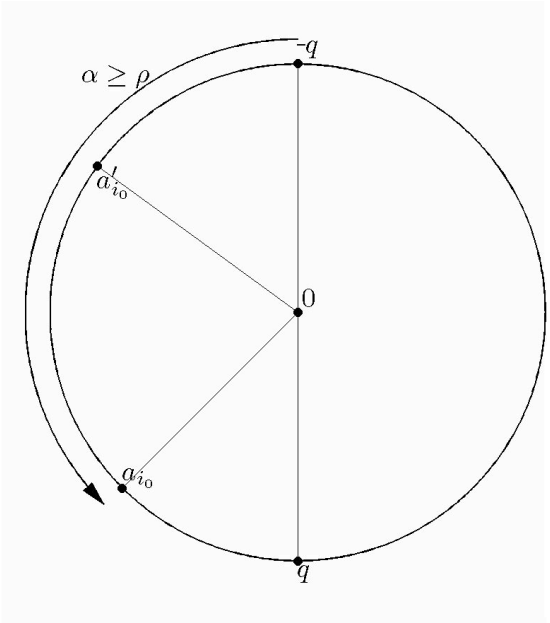

Figure 6.4: $A'q \leq 0$, $A \notin \mathcal{F}_D$, and $d_{\mathbb{S}}\left(a_{i_0}, a'_{i_0}\right) \geq \alpha - \frac{\pi}{2}$

It follows that

$$d_{\mathbb{S}}(A, A') \geq d_{\mathbb{S}}(a_{i_0}, a'_{i_0}) \geq d_{\mathbb{S}}(a_{i_0}, -q) - d_{\mathbb{S}}(-q, a'_{i_0}) \geq \alpha - \frac{\pi}{2} \geq \rho - \frac{\pi}{2}.$$

Therefore $d_{\mathbb{S}}(A, \Sigma) \geq \rho - \frac{\pi}{2}$, which completes the proof. □

6.4 The GCC Condition Number and Images of Balls

The goal of this section is to exhibit a characterization of $\mathscr{C}(A)$ in the spirit of Proposition 2.9. The positive orthant will have to play a role alongside the

balls and, unlike Proposition 2.9, the statement of the corresponding result now, Proposition 6.17 below, is far from apparent.

Proposition 6.17. *Let* $A = [a_1, \dots, a_n] \in (\mathbb{S}^{m-1})^n$.

1. *If* $A \in \mathcal{F}_D$ *then*

$$\Delta(A) = \sup\big\{\delta \mid \|\bar{x}\|_\infty \leq \delta \Rightarrow \bar{x} \in \{A^{\mathrm{T}}v + \mathbb{R}^n_+ : \|v\| \leq 1\}\big\}.$$

2. *If* $A \in \mathcal{F}_P$ *then*

$$\Delta(A) = \sup\big\{\delta \mid \|\bar{y}\| \leq \delta \Rightarrow \bar{y} \in \{Au : u \geq 0, \|u\|_1 \leq 1\}\big\}.$$

Proof. (1) Let $\Omega := \big\{\delta \mid \|\bar{x}\|_\infty \leq \delta \Rightarrow \bar{x} \in \{A^{\mathrm{T}}v + \mathbb{R}^n_+ : \|v\| \leq 1\}\big\}$. Since $A \in \mathcal{F}_D$ we have $\rho(A) \leq \frac{\pi}{2}$. Theorem 6.16 then implies $\Delta(A) = \sin d_{\mathbb{S}}(A, \Sigma) = \sin(\frac{\pi}{2} - \rho(A)) = \cos\rho(A)$. Let $-v$ be the center of a SIC of A. Then $\langle a_i, -v\rangle \geq \cos\rho(A) = \Delta(A)$, hence $\langle a_i, v\rangle \leq -\Delta(A)$ for all i.

For any $\bar{x} \in \mathbb{R}^n$ with $\|\bar{x}\|_\infty < \Delta(A)$ we have $-\Delta(A) < \bar{x}_i$ and therefore $A^{\mathrm{T}}v < \bar{x}$. That is, $\bar{x} \in \{A^{\mathrm{T}}v + \mathbb{R}^n_+ : \|v\| \leq 1\}$. This shows that $\Delta(A) \in \Omega$ and hence $\Delta(A) \leq \sup\Omega$.

To see the reversed inequality let $E \in \mathbb{R}^{m\times n}$ such that $A + E \notin \mathcal{F}_D$. Then, there exists $x \in \mathbb{R}^n$, $x \geq 0$, such that $(A+E)x = 0$. Without loss of generality, we may assume $\|x\|_1 = 1$. This implies that $Ex = -Ax$ and therefore, that

$$\text{for all } v \in \mathbb{R}^m \qquad x^{\mathrm{T}}E^{\mathrm{T}}v = -x^{\mathrm{T}}A^{\mathrm{T}}v. \tag{6.5}$$

Consider now any $\delta \in \Omega$. By (2.3), there exists $\bar{x} \in \mathbb{R}^n$ such that $\|\bar{x}\|_\infty = \delta$ and $\bar{x}^{\mathrm{T}}x = -\delta$. Since $\delta \in \Omega$ there exists $v \in \mathbb{R}^m$, $\|v\| \leq 1$, such that $A^{\mathrm{T}}v \leq \bar{x}$. Using that $x \geq 0$ we deduce

$$v^{\mathrm{T}}Ax = x^{\mathrm{T}}A^{\mathrm{T}}v \leq x^{\mathrm{T}}\bar{x} = -\delta$$

which implies using (6.5),

$$\|E\|_{12} \geq \|Ex\| \geq \|Ex\|\|v\| \geq \big|v^{\mathrm{T}}Ex\big| = \big|v^{\mathrm{T}}Ax\big| = \delta.$$

This shows that $\Delta(A) \geq \sup\Omega$.

(2) Let now $\Omega = \big\{\delta \mid \|\bar{y}\| \leq \delta \Rightarrow \bar{y} \in \{Au : u \geq 0, \|u\|_1 \leq 1\}\big\}$. Since $A \in \mathcal{F}_P$ we have $\rho(A) \geq \frac{\pi}{2}$ and Theorem 6.16 implies $\Delta(A) = \sin d_{\mathbb{S}}(A, \Sigma) = \sin(\rho(A) - \frac{\pi}{2}) = -\cos\rho(A)$.

Let $\bar{y} \in \mathbb{R}^m$ such that $\|\bar{y}\| < \Delta(A)$. Suppose that $\bar{y}$ is not contained in the closed convex set

$$\{Au \mid u \geq 0\,,\ \|u\|_1 \leq 1\}.$$

The separating hyperplane Theorem 6.4 shows that there exists $u \in \mathbb{R}^m$ with $\|u\| = 1$ and $\lambda \in \mathbb{R}$ such that, for all u with $u \geq 0$, $\|u\|_1 \leq 1$,

$$u^{\mathrm{T}}\bar{y} < \lambda < u^{\mathrm{T}}Au.$$

For $i = 1, \ldots, n$ take $u = (0, \ldots, 0, 1, 0 \ldots, 0)$. This yields

$$\forall i \in [n] \quad \lambda < \langle u, a_i \rangle .$$

By the Cauchy-Schwarz inequality

$$-\lambda < -u^{\mathrm{T}} \bar{y} \leq \|\bar{y}\| < \Delta(A).$$

Hence $\lambda > -\Delta(A) = \cos \rho(A)$. From the inequalities

$$\forall i \in [n] \ \cos \rho(A) < \langle u, a_i \rangle$$

it follows that there is a spherical cap centered at u containing all the a_i, that has a radius strictly smaller than $\rho(A)$. This is a contradiction. We conclude that if $\delta < \Delta(A)$ then $\delta \in \Omega$. This shows $\Delta(A) \leq \sup \Omega$.

To show the reversed inequality let now $E \in \mathbb{R}^{m \times n}$ be such that $A + E \notin \mathcal{F}_P$. Then, there exists $y \in \mathbb{R}^m$ such that $(A+E)^{\mathrm{T}} y \geq 0$. Without loss of generality, we can assume $\|y\| = 1$. This implies that $E^{\mathrm{T}} y \geq -A^{\mathrm{T}} y$ and hence, that

$$\text{for all } u \in \mathbb{R}^n,\ u \geq 0, \qquad u^{\mathrm{T}} E^{\mathrm{T}} y \geq -u^{\mathrm{T}} A^{\mathrm{T}} y. \tag{6.6}$$

Consider now any $\delta \in \Omega$. By (2.3) there exists $\bar{y} \in \mathbb{R}^m$, $\|\bar{y}\| = \delta$, such that $\bar{y}^{\mathrm{T}} y = -\delta$. Since $\delta \in \Omega$ there exists $u \in \mathbb{R}^n$, $u \geq 0$, $\|u\|_1 = 1$, such that $Au = \bar{y}$. Hence, using (6.6),

$$y^{\mathrm{T}} E u = u^{\mathrm{T}} E^{\mathrm{T}} y \geq -u^{\mathrm{T}} A^{\mathrm{T}} y = -y^{\mathrm{T}} A u = -y^{\mathrm{T}} \bar{y} = \delta$$

which implies

$$\|E\|_{12} \geq \|Eu\| \geq \|Eu\| \|y\| \geq \left| y^{\mathrm{T}} E u \right| = \delta.$$

This shows $\Delta(A) \geq \sup \Omega$. □

6.5 The GCC Condition Number and Well-Conditioned Solutions

The definition of $\mathscr{C}(A)$ given in Section 6.2 is in terms of a relativized distance to ill-posedness. Its characterization in Section 6.3 translates the space where the geometric property defining $\mathscr{C}(A)$ occurs from the space of data $(\mathbb{S}^{m-1})^n$ —where $d_{\mathbb{S}}$ is defined— to the sphere $\mathbb{S}^{m-1}$ —where smallest including caps are. With a little extra effort we can now look at $\mathbb{S}^{m-1}$ as the space of solutions for the problem $A^{\mathrm{T}} y \leq 0$ and characterize $\mathscr{C}(A)$ in terms of the 'best conditioned solution' (at least when $A \in \mathcal{F}_D$). This is the idea.

For $A \in \mathbb{R}^{m \times n}$ with non-zero columns a_i we define

$$\Xi(A) := \min_{y \in \mathbb{S}^{m-1}} \max_{i \leq n} \frac{a_i^{\mathrm{T}} y}{\|a_i\|}.$$

Proposition 6.18. *For all $A \in \mathbb{R}^{m\times n}$ with non-zero columns $|\Xi(A)| = \Delta(A)$.*

Proof. By Theorem 6.16 it is enough to show that $\Xi(A) = -\cos\rho(A)$. To do so, in addition, we may assume $\|a_i\| = 1$ for $i \in [n]$.

Let $\rho = \rho(A)$ and $p \in \mathbb{S}^{m-1}$ such that $\mathsf{cap}(p,\rho)$ is a SIC for A. Take $\bar{y} = -p$. Then,

$$\Xi(A) \leq \max_{i\leq n} a_i^{\mathrm{T}}\bar{y} = -\min_{i\leq n} a_i^{\mathrm{T}}p \leq -\cos\rho$$

the last inequality since $a_i \in \mathsf{cap}(p,\rho)$.

To show the reversed inequality let y_* be such that $\Xi(A) = \max_{i\leq n} a_i^{\mathrm{T}}y_*$ and let $p_* = -y_*$ and $\sigma = \arccos\Xi(A)$. Then,

$$\min_{i\leq n} a_i^{\mathrm{T}}p = -\max_{i\leq n} a_i^{\mathrm{T}}y_* = -\cos\sigma = \cos\Big(\frac{\pi}{2} - \sigma\Big).$$

It follows that $a_i \in \mathsf{cap}(p, \frac{\pi}{2} - \sigma)$ and therefore, that $\rho \leq \frac{\pi}{2} - \sigma$. This implies $\Xi(A) \geq -\cos\rho$. □

Proposition 6.18 introduces a new view for condition. In our first approach in Chapter 1 we considered problems as functions $\varphi : \mathcal{D} \subseteq \mathbb{R}^m \to \mathbb{R}^q$. A number of natural problems, however, do not fit this pattern since the desired output for a data $a \in \mathcal{D}$ may not be univocaly specified. For instance, the problem of computing a complex root when given a univariate polynomial (which does not require any precise root to be returned). Or the problem of, given a matrix $A \in \mathbb{R}^{m\times n}$, decide whether $A \in \mathcal{F}_D$ and if so, return a point $y \in \mathbb{R}^m \setminus \{0\}$ such that $A^{\mathrm{T}}y \leq 0$.

For problems of this kind, we may approach condition from a different viewpoint. For an input a, let $\mathsf{Sol}(a)$ be its associated set of solutions (i.e., all the possible outputs for a). If for each $y \in \mathsf{Sol}(A)$ we have a number $\xi(a,y)$ quantifying the quality of the solution y we may define the condition $\xi(a)$ of a by taking some function on the set $\{\xi(a,y) \mid y \in \mathsf{Sol}(a)\}$. Typical choices are

$$\xi(a) := \inf_{y\in\mathsf{Sol}(a)} \xi(a,y), \quad \xi(a) := \sup_{y\in\mathsf{Sol}(a)} \xi(a,y), \quad \text{and} \quad \xi(a) := \mathop{\mathbb{E}}_{y\in\mathsf{Sol}(a)} \xi(a,y)$$

where the expectation in the last expression is for some distribution on $\mathsf{Sol}(A)$. In the case of a matrix $A \in \mathcal{F}_D$ we have $\mathsf{Sol}(A) = \{y \in \mathbb{R}^m \setminus \{0\} \mid A^{\mathrm{T}}y \leq 0\}$ and Proposition 6.18 expresses $\mathscr{C}(A)$ as $\min_{y\in\mathsf{Sol}(A)} \xi(A,y)$ for

$$\xi(A,y) := \frac{1}{-\max_{i\leq n} \frac{a_i^{\mathrm{T}}y}{\|a_i\|\|y\|}}.$$

The quantity $\xi(A,y)$ is the sinus of the angular distance from y to the boundary of the cone $\mathsf{Sol}(A)$. The larger this distance, the better conditioned is the solution y. The equality $\mathscr{C}(A)$ as $\min_{y\in\mathsf{Sol}(A)} \xi(A,y)$ thus expresses $\mathscr{C}(A)$ as the condition of the 'best conditioned' point in $\mathsf{Sol}(A)$.

We close this chapter by mentioning that we will encounter in Chapter 11 examples for the other two choices for $\xi(a)$ namely, $\xi(a) := \sup_{y \in \mathsf{Sol}(a)} \xi(a, y)$ —the 'worse conditioned' solution— as well as $\xi(a) := \mathbb{E}_{y \in \mathsf{Sol}(a)} \xi(a, y)$ —the 'average conditioned' solution.