Chapter 4

Norms and Matrix Conditioning

4.1 Matrix Norms

Let us review a few facts about vector norms that we will need.

Theorem 4.1.1 (Reverse Triangle Inequality)**.** *Let* $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$. *Then*

$$\big|\, \|\boldsymbol{x}\| - \|\boldsymbol{y}\| \,\big| \le \|\boldsymbol{x} - \boldsymbol{y}\|\,,$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$.

Proof. By the usual triangle inequality,

$$\|\boldsymbol{x}\| = \|\boldsymbol{x} - \boldsymbol{y} + \boldsymbol{y}\| \le \|\boldsymbol{x} - \boldsymbol{y}\| + \|\boldsymbol{y}\|\,.$$

Thus

$$\|\boldsymbol{x}\| - \|\boldsymbol{y}\| \le \|\boldsymbol{x} - \boldsymbol{y}\|\,. \tag{4.1.1}$$

Reversing the roles of $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$\|\boldsymbol{y}\| - \|\boldsymbol{x}\| \le \|\boldsymbol{y} - \boldsymbol{x}\| = \|\boldsymbol{x} - \boldsymbol{y}\|\,.$$

The last equation implies

$$-\|\boldsymbol{x} - \boldsymbol{y}\| \le \|\boldsymbol{x}\| - \|\boldsymbol{y}\| \tag{4.1.2}$$

Estimates (4.1.1) and (4.1.2) combine to give

$$-\|\boldsymbol{x} - \boldsymbol{y}\| \le \|\boldsymbol{x}\| - \|\boldsymbol{y}\| \le \|\boldsymbol{x} - \boldsymbol{y}\|\,,$$

which is equivalent to

$$\big|\, \|\boldsymbol{x}\| - \|\boldsymbol{y}\| \,\big| \le \|\boldsymbol{x} - \boldsymbol{y}\|\,.$$

$\square$

Theorem 4.1.2. *Every vector norm $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$ is a uniformly continuous function with respect to the metric $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_\infty$.*

Proof. Let $\varepsilon > 0$ be given. Suppose $\boldsymbol{x}, \boldsymbol{h} = [h_i] \in \mathbb{C}^n$ are arbitrary. Then, using the reverse triangle inequality and the triangle inequality

$$\begin{aligned} \big| \|\boldsymbol{x}+\boldsymbol{h}\| - \|\boldsymbol{x}\| \big| &\leq \|\boldsymbol{h}\| \\ &= \left\| \sum_{i=1}^n h_i \boldsymbol{e}_i \right\| \\ &\leq \sum_{i=1}^n |h_i| \, \|\boldsymbol{e}_i\| \\ &\leq \max_{1 \leq i \leq n} |h_i| \sum_{i=1}^n \|\boldsymbol{e}_i\| \\ &= M \|\boldsymbol{h}\|_\infty , \end{aligned}$$

where $\boldsymbol{e}_j$ is the j^{th} canonical basis vector and

$$M := \sum_{i=1}^n \|\boldsymbol{e}_i\| .$$

Set $\delta = \frac{\varepsilon}{M}$. Then, if $\|\boldsymbol{h}\|_\infty < \delta$, then

$$\big| \|\boldsymbol{x}+\boldsymbol{h}\| - \|\boldsymbol{x}\| \big| < \varepsilon,$$

which proves that $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$ is continuous at $\boldsymbol{x}$. Since δ is independent of $\boldsymbol{x}$, it is, in fact, uniformly continuous. ☐

Theorem 4.1.3. *Suppose that $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$ is a vector norm. There exist positive constants $0 < M_1 \leq M_2$ such that*

$$M_1 \|\boldsymbol{x}\|_\infty \leq \|\boldsymbol{x}\| \leq M_2 \|\boldsymbol{x}\|_\infty , \quad \forall\, \boldsymbol{x} \in \mathbb{C}^n.$$

In this case, we say that $\|\cdot\|$ and $\|\cdot\|_\infty$ are equivalent.

Proof. Set

$$S_\infty^{n-1} := \{\boldsymbol{x} \in \mathbb{C}^n \mid \|\boldsymbol{x}\|_\infty = 1\} .$$

S_∞^{n-1} is closed and bounded and, therefore, compact. Since $\|\cdot\| : S_\infty^{n-1} \to \mathbb{R}$ is continuous function defined on a compact set, there exist points $\boldsymbol{x}_i \in S_\infty^{n-1}$, $i = 1, 2$, such that (the minimum is achieved)

$$M_1 := \inf_{x \in S_\infty^{n-1}} \|\boldsymbol{x}\| = \|\boldsymbol{x}_1\| > 0,$$

and (the maximum is achieved)

$$M_2 := \sup_{x \in S_\infty^{n-1}} \|\boldsymbol{x}\| = \|\boldsymbol{x}_2\| > 0,$$

and, clearly, $M_1 \le M_2$. For any arbitrary $\boldsymbol{y} \in \mathbb{C}_\star^n$,

$$\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|\infty} \in S_\infty^{n-1}.$$

It follows that

$$M_1 \le \left\| \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_\infty} \right\| \le M_2,$$

or, equivalently,

$$M_1 \|\boldsymbol{y}\|_\infty \le \|\boldsymbol{y}\| \le M_2 \|\boldsymbol{y}\|_\infty .$$

Since the result is clearly true for $\boldsymbol{y} = \boldsymbol{0}$, the proof is complete. $\square$

Exercise 4.1.4. *Since any two vector norms $\|\cdot\|_a, \|\cdot\|_b : \mathbb{C}^n \to \mathbb{R}$ are equivalent to the infinity norm $\|\cdot\|_\infty : \mathbb{C}^n \to \mathbb{R}$, they are, in fact, equivalent to each other. In other words, given any two vector norms $\|\cdot\|_a, \|\cdot\|_b$, there are positive constants $0 < C_1 \le C_2$ such that*

$$C_1 \|\boldsymbol{x}\|_a \le \|\boldsymbol{x}\|_b \le C_2 \|\boldsymbol{x}\|_a , \quad \forall\, \boldsymbol{x} \in \mathbb{C}^n.$$

The reader should prove this.

Recall the following definition,

Definition 4.1.5. *A function $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}$ is called a* matrix norm *iff it satisfies*

1. $\|\mathsf{A}\| = 0$ *implies that* $\mathsf{A} = \mathsf{O} \in \mathbb{C}^{m\times n}$*, the zero matrix.*
2. $\|\alpha\mathsf{A}\| = |\alpha| \cdot \|\mathsf{A}\|$*, for all* $\alpha \in \mathbb{C}$ *and* $\mathsf{A} \in \mathbb{C}^{m\times n}$*.*
3. $\|\mathsf{A} + \mathsf{B}\| \le \|\mathsf{A}\| + \|\mathsf{B}\|$*, for all* $\mathsf{A}, \mathsf{B} \in \mathbb{C}^{m\times n}$*.*

Definition 4.1.6. *Suppose that $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}$ is a matrix norm. Then $\|\cdot\|$ is called* consistent *with respect to the norms $\|\cdot\|_{\mathbb{C}^m} : \mathbb{C}^m \to \mathbb{R}$ and $\|\cdot\|_{\mathbb{C}^n} : \mathbb{C}^n \to \mathbb{R}$ iff*

$$\|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m} \le \|\mathsf{A}\| \, \|\boldsymbol{x}\|_{\mathbb{C}^n} ,$$

for all $\mathsf{A} \in \mathbb{C}^{m\times n}$ and $\boldsymbol{x} \in \mathbb{C}^n$.

Definition 4.1.7. *Suppose that* $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ *is a matrix norm. Then* $\|\cdot\|$ *is called* sub-multiplicative *iff, for all* $\mathsf{A}, \mathsf{B} \in \mathbb{C}^{n\times n}$

$$\|\mathsf{AB}\| \leq \|\mathsf{A}\| \cdot \|\mathsf{B}\| .$$

Example 4.1.8. *For any* $\mathsf{A} = [a_{i,j}] \in \mathbb{C}^{m\times n}$ *define matrix max norm via*

$$\|\mathsf{A}\|_{\max} := \max_{\substack{1\leq i\leq m\\ 1\leq j\leq n}} |a_{i,j}|,$$

and the Frobenius norm via

$$\|\mathsf{A}\|_F := \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{i,j}|^2}.$$

These functions are bona *fide norms, as the reader should confirm. The matrix max norm is not sub-multiplicative; the Frobenius norm is sub-multiplicative. It can also be shown that the Frobenius norm is consistent with respect to the* $\ell^2(\mathbb{C}^m)$ *and* $\ell^2(\mathbb{C}^n)$ *norms. The reader should prove this.*

Definition 4.1.9. *Suppose that* $\|\cdot\|_{\mathbb{C}^m} : \mathbb{C}^m \to \mathbb{R}$ *and* $\|\cdot\|_{\mathbb{C}^n} : \mathbb{C}^n \to \mathbb{R}$ *are vector norms. The function* $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}$ *defined by*

$$\|\mathsf{A}\| := \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \frac{\|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}}, \quad \forall\, \mathsf{A} \in \mathbb{C}^{m\times n},$$

is called the induced matrix norm *with respect to* $\|\cdot\|_{\mathbb{C}^m}$ *and* $\|\cdot\|_{\mathbb{C}^n}$.

Whenever $n = m$*, it is understood that the vector norms are the same. We say that* $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ *is the induced matrix norm with respect the vector norm* $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$ *iff*

$$\|\mathsf{A}\| := \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \frac{\|\mathsf{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|}, \quad \forall\, \mathsf{A} \in \mathbb{C}^{n\times n},$$

dropping the superfluous subscripts on the vector norms.

Remark 4.1.10. *We also use the terms* operator norm *and* subordinate matrix norm *interchangeably with the term induced matrix norm.*

Example 4.1.11. *For any* $p \in [1, \infty]$ *we define the induced matrix p-norm via*

$$\|\mathsf{A}\|_p := \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \frac{\|\mathsf{A}\boldsymbol{x}\|_{\ell^p(\mathbb{C}^m)}}{\|\boldsymbol{x}\|_{\ell^p(\mathbb{C}^n)}}, \quad \mathsf{A} \in \mathbb{C}^{m\times m}.$$

We will use these frequently in the book.

Theorem 4.1.12. *Consider the induced matrix 2-norm* $\|\cdot\|_2 : \mathbb{C}^{m\times n} \to \mathbb{R}$*, defined via*

$$\|\mathsf{A}\|_2 = \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \frac{\|\mathsf{A}\boldsymbol{x}\|_{\ell^2(\mathbb{C}^m)}}{\|\boldsymbol{x}\|_{\ell^2(\mathbb{C}^n)}} ,$$

where, recall, $\mathbb{C}^n_\star := \mathbb{C}^n \setminus \{\mathbf{0}\}$*. Then*

$$\|\mathsf{A}\|_2 = \max_{1\le i\le n} \sqrt{\lambda_i} ,$$

where the λ_i *are the eigenvalues of the matrix* $\mathsf{B} = \mathsf{A}^H\mathsf{A}$*.*

Proof. The eigenvalues of $\mathsf{B} = \mathsf{A}^H\mathsf{A}$ are real and non-negative. To see this, let $(\lambda, \boldsymbol{w})$ be an eigen-pair. Then

$$0 \le \|\mathsf{A}\boldsymbol{w}\|^2_{\ell^2(\mathbb{C}^m)} = \boldsymbol{w}^H\mathsf{A}^H\mathsf{A}\boldsymbol{w} = \boldsymbol{w}^H(\lambda\boldsymbol{w}) = \lambda\boldsymbol{w}^H\boldsymbol{w} = \lambda\,\|\boldsymbol{w}\|^2_{\ell^2(\mathbb{C}^n)} .$$

Hence

$$\lambda = \frac{\|\mathsf{A}\boldsymbol{w}\|^2_{\ell^2(\mathbb{C}^m)}}{\|\boldsymbol{w}\|^2_{\ell^2(\mathbb{C}^n)}} \ge 0 .$$

As $\mathsf{B} = \mathsf{A}^H\mathsf{A}$ is Hermitian, it has an orthonormal basis of eigenvectors

$$S = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\} \subset \mathbb{C}^n.$$

Let $\boldsymbol{x} \in \mathbb{C}^n_\star$ be arbitrary. There exist unique constants $c_1, \dots, c_n \in \mathbb{C}$, not all zero, such that $\boldsymbol{x} = \sum_{i=1}^n c_i\boldsymbol{w}_i$. Then

$$\frac{\|\mathsf{A}\boldsymbol{x}\|^2_{\ell^2(\mathbb{C}^m)}}{\|\boldsymbol{x}\|^2_{\ell^2(\mathbb{C}^n)}} = \frac{\boldsymbol{x}^H\mathsf{A}^H\mathsf{A}\boldsymbol{x}}{\boldsymbol{x}^H\boldsymbol{x}} = \frac{\sum_{i=1}^n \lambda_i|c_i|^2}{\sum_{i=1}^n |c_i|^2} \le \lambda_{\max}\frac{\sum_{i=1}^n |c_i|^2}{\sum_{i=1}^n |c_i|^2} = \lambda_{\max} ,$$

where $\max_{1\le i\le n}\lambda_i =: \lambda_{\max}$. Since the far right hand side is independent of $\boldsymbol{x}$

$$\|\mathsf{A}\|_2 \le \sqrt{\max_{1\le i\le n}\lambda_i} = \max_{1\le i\le n}\sqrt{\lambda_i} .$$

Now let $\boldsymbol{w}_{\max} \in \mathbb{C}^n_\star$ be an eigenvector associated to $\lambda_{\max}$. Then note

$$\frac{\|\mathsf{A}\boldsymbol{w}_{\max}\|_{\ell^2(\mathbb{C}^m)}}{\|\boldsymbol{w}_{\max}\|_{\ell^2(\mathbb{C}^n)}} = \sqrt{\frac{\boldsymbol{w}^H_{\max}\mathsf{A}^H\mathsf{A}\boldsymbol{w}_{\max}}{\boldsymbol{w}^H_{\max}\boldsymbol{w}_{\max}}} = \sqrt{\lambda_{\max}}\sqrt{\frac{\boldsymbol{w}^H_{\max}\boldsymbol{w}_{\max}}{\boldsymbol{w}^H_{\max}\boldsymbol{w}_{\max}}} = \sqrt{\lambda_{\max}} .$$

Thus, in fact, the upper bound is achieved, and we have

$$\|\mathsf{A}\|_2 = \max_{1\le i\le n}\sqrt{\lambda_i}.$$

□

Theorem 4.1.13. *Suppose that $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}$ is the induced norm with respect to the vector norms $\|\cdot\|_{\mathbb{C}^m}$ and $\|\cdot\|_{\mathbb{C}^n}$. Then $\|\cdot\|$ is a* bona fide *matrix norm. It is consistent with respect to the vector norms vector norms $\|\cdot\|_{\mathbb{C}^m}$ and $\|\cdot\|_{\mathbb{C}^n}$ used to define it. If $m = n$, the induced norm is sub-multiplicative. Furthermore,*

$$\|\mathsf{A}\| \le C\sqrt{\rho(\mathsf{A}^H\mathsf{A})} < \infty,$$

for some $C > 0$, for all $\mathsf{A} \in \mathbb{C}^{m\times n}$, where, recall,

$$\rho(\mathsf{B}) := \max_{\lambda\in\sigma(\mathsf{B})} |\lambda|, \quad \mathsf{B} \in \mathbb{C}^{p\times p},$$

is the spectral radius. Finally,

$$\|\mathsf{A}\| = \sup_{\boldsymbol{x}\in S^{n-1}_{\mathbb{C}^n}} \|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m},$$

where

$$S^{n-1}_{\mathbb{C}^n} = \{\boldsymbol{x} \in \mathbb{C}^n \mid \|\boldsymbol{x}\|_{\mathbb{C}^n} = 1\}.$$

Proof. (Positive Definiteness:) Suppose that $\|\mathsf{A}\| = 0$. Then, $\|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m} = 0$, for all $\boldsymbol{x} \in \mathbb{C}^n$. This implies that $\mathsf{A}\boldsymbol{x} = \boldsymbol{0}$, for all $\boldsymbol{x} \in \mathbb{C}^n$. Thus $\mathsf{A}\boldsymbol{e}_i = \boldsymbol{0}$, for all $i = 1, \cdots, n$. This implies that the columns of A are all zero, which implies that A is the zero matrix.

(Non-negative Homogeneity:) Let $\alpha \in \mathbb{C}$ and $\mathsf{A} \in \mathbb{C}^{m\times n}$ be arbitrary. Then

$$\|\alpha\mathsf{A}\| = \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \frac{\|\alpha\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} = \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} |\alpha|\frac{\|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} = |\alpha| \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \frac{\|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} = |\alpha|\,\|\mathsf{A}\|,$$

where we have used a familiar property of the supremum.

(Consistency:) It helps to prove the consistency next. For any $\boldsymbol{x} \in \mathbb{C}^n_\star$ and $\mathsf{A} \in \mathbb{C}^{m\times n}$,

$$\|\mathsf{A}\| = \sup_{\boldsymbol{w}\in\mathbb{C}^n_\star} \frac{\|\mathsf{A}\boldsymbol{w}\|_{\mathbb{C}^m}}{\|\boldsymbol{w}\|_{\mathbb{C}^n}} \ge \frac{\|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}},$$

since the supremum (i.e., the least upper bound) is an upper bound. Hence,

$$\|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m} \le \|\mathsf{A}\| \cdot \|\boldsymbol{x}\|_{\mathbb{C}^n}.$$

Of course the last inequality holds trivially for $\boldsymbol{x} = \boldsymbol{0}$.

(Triangle Inequality:) Let $\mathsf{A}, \mathsf{B} \in \mathbb{C}^{m\times n}$ be arbitrary. Then, using the triangle inequality for vector norms and consistency,

$$\begin{aligned}
\frac{\|(\mathsf{A}+\mathsf{B})\boldsymbol{x}\|_{\mathbb{C}^m}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} &\leq \frac{\|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m} + \|\mathsf{B}\boldsymbol{x}\|_{\mathbb{C}^m}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} \\
&= \frac{\|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} + \frac{\|\mathsf{B}\boldsymbol{x}\|_{\mathbb{C}^m}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} \\
&\leq \frac{\|\mathsf{A}\|\,\|\boldsymbol{x}\|_{\mathbb{C}^n}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} + \frac{\|\mathsf{B}\|\,\|\boldsymbol{x}\|_{\mathbb{C}^n}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} \\
&= \|\mathsf{A}\| + \|\mathsf{B}\|,
\end{aligned}$$

for all $\boldsymbol{x} \in \mathbb{C}^n_\star$. Hence

$$\|\mathsf{A}+\mathsf{B}\| = \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \frac{\|(\mathsf{A}+\mathsf{B})\boldsymbol{x}\|_{\mathbb{C}^m}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} \leq \|\mathsf{A}\| + \|\mathsf{B}\|.$$

(Sub-Multiplicativity:) Suppose that $\mathsf{A}, \mathsf{B} \in \mathbb{C}^{n\times n}$ are arbitrary. Then, using consistency twice,

$$\begin{aligned}
\frac{\|\mathsf{AB}\boldsymbol{x}\|_{\mathbb{C}^n}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} &\leq \frac{\|\mathsf{A}\| \cdot \|\mathsf{B}\boldsymbol{x}\|_{\mathbb{C}^n}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} \\
&\leq \frac{\|\mathsf{A}\| \cdot \|\mathsf{B}\|\,\|\boldsymbol{x}\|_{\mathbb{C}^n}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} \\
&= \|\mathsf{A}\| \cdot \|\mathsf{B}\|,
\end{aligned}$$

for all $\boldsymbol{x} \in \mathbb{C}^n_\star$. Hence

$$\|\mathsf{AB}\| = \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \frac{\|\mathsf{AB}\boldsymbol{x}\|_{\mathbb{C}^n}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} \leq \|\mathsf{A}\| \cdot \|\mathsf{B}\|.$$

(Finiteness:) By equivalence of vector norms, there are constants $0 < M_1 \leq M_2$ and $0 < C_1 \leq C_2$ such that

$$M_1 \|\boldsymbol{x}\|_{\ell^2(\mathbb{C}^m)} \leq \|\boldsymbol{x}\|_{\mathbb{C}^m} \leq M_2 \|\boldsymbol{x}\|_{\ell^2(\mathbb{C}^m)}, \quad \forall\, \boldsymbol{x} \in \mathbb{C}^m,$$

and, similarly,

$$C_1 \|\boldsymbol{x}\|_{\ell^2(\mathbb{C}^n)} \leq \|\boldsymbol{x}\|_{\mathbb{C}^n} \leq C_2 \|\boldsymbol{x}\|_{\ell^2(\mathbb{C}^n)}, \quad \forall\, \boldsymbol{x} \in \mathbb{C}^n.$$

Therefore, for any $\mathsf{A} \in \mathbb{C}^{m\times n}$ and any $\boldsymbol{x} \in \mathbb{C}^n_\star$,

$$\frac{\|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} \leq \frac{M_2}{C_1} \frac{\|\mathsf{A}\boldsymbol{x}\|_{\ell^2(\mathbb{C}^m)}}{\|\boldsymbol{x}\|_{\ell^2(\mathbb{C}^n)}}.$$

Therefore,

$$\|\mathsf{A}\| \le \frac{M_2}{C_1} \|\mathsf{A}\|_2 ,$$

and the previous theorem gives the desired result.

(Equivalent Definition:) Using non-negative homogeneity,

$$\begin{aligned}
\|\mathsf{A}\| &= \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \frac{\|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} \\
&= \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \left\| \frac{1}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} \mathsf{A}\boldsymbol{x} \right\|_{\mathbb{C}^m} \\
&= \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \left\| \mathsf{A}\left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}} \right) \right\|_{\mathbb{C}^m} \\
&= \sup_{\boldsymbol{w}\in S^{n-1}_{\mathbb{C}^n}} \|\mathsf{A}\boldsymbol{w}\|_{\mathbb{C}^m} .
\end{aligned}$$

□

Theorem 4.1.14. *Suppose that* $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}$ *is the induced norm with respect to the vector norms* $\|\cdot\|_{\mathbb{C}^m}$ *and* $\|\cdot\|_{\mathbb{C}^n}$*. For any* $\mathsf{A} \in \mathbb{C}^{m\times n}$*, there is an element* $\boldsymbol{x} \in S^{n-1}_{\mathbb{C}^n}$ *such that*

$$\|\mathsf{A}\| = \|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m} .$$

Proof. The function $\|\mathsf{A}(\,\cdot\,)\|_{\mathbb{C}^m} : \mathbb{C}^n \to \mathbb{R}$ is uniformly continuous. The proof of this fact is left to reader. If this is the case, $\|\mathsf{A}(\,\cdot\,)\|_{\mathbb{C}^m} : S^{n-1}_{\mathbb{C}^n} \to \mathbb{R}$ is a uniformly continuous function defined on a compact set. Therefore, there are vectors $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S^{n-1}_{\mathbb{C}^n}$, such that

$$\sup_{\boldsymbol{x}\in S^{n-1}_{\mathbb{C}^n}} \|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m} = \|\mathsf{A}\boldsymbol{x}_1\|_{\mathbb{C}^m} \quad \text{and} \quad \inf_{\boldsymbol{x}\in S^{n-1}_{\mathbb{C}^n}} \|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m} = \|\mathsf{A}\boldsymbol{x}_2\|_{\mathbb{C}^m} .$$

Since

$$\|\mathsf{A}\| = \sup_{\boldsymbol{w}\in S^{n-1}_{\mathbb{C}^n}} \|\mathsf{A}\boldsymbol{w}\|_{\mathbb{C}^m} ,$$

the result follows. □

Theorem 4.1.15. *Suppose that* $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ *is the induced matrix norm with respect to the vector norm* $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$*. Then*

$$\|\mathsf{I}_n\| = 1$$

and, for every $\mathsf{A} \in \mathbb{C}^{n\times n}$

$$\rho(\mathsf{A}) \le \|\mathsf{A}\| .$$

Proof. The proof is an exercise. □

4.2 Conditioning

Definition 4.2.1 (condition number)**.** *Suppose that $A \in \mathbb{C}^{n\times n}$ is invertible. The* condition number *of* A *with respect to the matrix norm $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is*

$$\kappa(A) := \|A\|\|A^{-1}\|.$$

Before we get on to the meaning and utility of the condition number, let us get some elementary properties out of the way.

Proposition 4.2.2. *Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is an induced matrix norm with respect to the vector norm $\|\cdot\| : \mathbb{C}^{n} \to \mathbb{R}$, and $\mathsf{A} \in \mathbb{C}^{n\times n}$ is invertible. Then*

$$\|\mathsf{A}\| \left\|\mathsf{A}^{-1}\right\| =: \kappa(\mathsf{A}) \geq 1.$$

Furthermore,

$$\frac{1}{\|\mathsf{A}^{-1}\|} \leq \|\mathsf{A} - \mathsf{B}\| \,,$$

for any $\mathsf{B} \in \mathbb{C}^{n\times n}$ that is singular. Consequently,

$$\frac{1}{\kappa(\mathsf{A})} \leq \inf_{\det(\mathsf{B})=0} \frac{\|\mathsf{A} - \mathsf{B}\|}{\|\mathsf{A}\|}.$$

Proof. Exercise. $\square$

There are some nice formulae for and estimates of the condition number with respect to the induced matrix 2-norm:

Proposition 4.2.3. *Suppose that $\mathsf{A} \in \mathbb{C}^{n\times n}$ is invertible and $\|\cdot\|_2 : \mathbb{C}^{n\times n} \to \mathbb{R}$ is the induced matrix 2-norm.*

1. *Then, if the singular values of* A *are $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$,*

$$\kappa_2(\mathsf{A}) := \|\mathsf{A}\|_2\|\mathsf{A}^{-1}\|_2 = \frac{\sigma_1}{\sigma_n}.$$

2. *If the eigenvalues of $\mathsf{B} = \mathsf{A}^H\mathsf{A}$ are $0 < \mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$, then*

$$\kappa_2(\mathsf{A}) = \sqrt{\frac{\mu_n}{\mu_1}}.$$

3. *Denote $\kappa_p(\mathsf{A}) = \|\mathsf{A}\|_p \cdot \left\|\mathsf{A}^{-1}\right\|_p$, for $p \in [1, +\infty]$, where $\|\cdot\|_p$ is the induced matrix norm with respect to the vector norm $\|\cdot\|_{\ell^p(\mathbb{C}^n)}$. Then,*

$$\kappa_2(\mathsf{A}) \leq \sqrt{\kappa_1(\mathsf{A})\kappa_\infty(\mathsf{A})}.$$

4.
$$\frac{1}{\kappa_2(\mathsf{A})} = \inf_{\det(\mathsf{B})=0} \frac{\|\mathsf{A}-\mathsf{B}\|_2}{\|\mathsf{A}\|_2}.$$

5. *If, additionally, it is assumed that* A *is Hermitian, then*
$$\kappa_2(A) = \frac{\max_{\lambda\in\sigma(\mathsf{A})} |\lambda|}{\min_{\lambda\in\sigma(\mathsf{A})} |\lambda|}.$$

6. *If additionally, it is known that* A *is HPD with eigenvalues* $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ *then*
$$\kappa_2(A) = \frac{\lambda_n}{\lambda_1}.$$

Proof. Exercise. □

The first and last formulae give an easy geometric interpretation of the condition number. It is the ratio of the *maximal stretching* to the *minimal stretching* under the action of the matrix A.

Definition 4.2.4. *Suppose that* $\mathsf{A} \in \mathbb{C}^{n\times n}$ *is invertible and* $\boldsymbol{b} \in \mathbb{C}^n$ *is given. The* residual vector *with respect to* $\boldsymbol{x}' \in \mathbb{C}^n$ *is defined as*
$$\boldsymbol{r} = \boldsymbol{r}(\boldsymbol{x}') := \boldsymbol{b} - \mathsf{A}\boldsymbol{x}' = \mathsf{A}(\boldsymbol{x} - \boldsymbol{x}'),$$
where $\boldsymbol{x} = \mathsf{A}^{-1}\boldsymbol{b}$. *The* error vector *with respect to* $\boldsymbol{x}'$ *is defined as*
$$\boldsymbol{e} = \boldsymbol{e}(\boldsymbol{x}') := \boldsymbol{x} - \boldsymbol{x}'.$$
Consequently,
$$\mathsf{A}\boldsymbol{e} = \boldsymbol{r}.$$

It often happens that we have obtained an approximate solution $\boldsymbol{x}' \in \mathbb{C}^n$. We would like to have some measure of the error, but a direct measurement of the error would require the exact solution $\boldsymbol{x}$. The next best thing is the residual, which is an indirect measurmnent of the error, as the last definition suggests. The next theorem tells us how useful the residual is in determining the relative size of the error.

Theorem 4.2.5. *Suppose that* $\mathsf{A} \in \mathbb{C}^{n\times n}$ *is invertible,* $\boldsymbol{b} \in \mathbb{C}^n_\star$ *is given, and* $\boldsymbol{x} = \mathsf{A}^{-1}\boldsymbol{b}$. *Assume that* $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ *is the induced matrix norm with respect to the vector norm* $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$. *Then*
$$\frac{1}{\kappa(\mathsf{A})}\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} \leq \frac{\|\boldsymbol{e}\|}{\|\boldsymbol{x}\|} \leq \kappa(\mathsf{A})\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|}.$$

Proof. Since $\boldsymbol{e} = \mathsf{A}^{-1}\boldsymbol{r}$, using consistency of the induced norm

$$\|\boldsymbol{e}\| = \left\|\mathsf{A}^{-1}\boldsymbol{r}\right\| \leq \left\|\mathsf{A}^{-1}\right\| \|\boldsymbol{r}\|.$$

Likewise,

$$\|\boldsymbol{b}\| = \|\mathsf{A}\boldsymbol{x}\| \leq \|\mathsf{A}\| \|\boldsymbol{x}\|,$$

which implies

$$\frac{1}{\|\boldsymbol{x}\|} \leq \|\mathsf{A}\| \frac{1}{\|\boldsymbol{b}\|}.$$

Combining the first and third inequalities, we get half of the result:

$$\frac{\|\boldsymbol{e}\|}{\|\boldsymbol{x}\|} \leq \|\mathsf{A}\| \left\|\mathsf{A}^{-1}\right\| \frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} = \kappa(\mathsf{A})\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|}.$$

The other half is left as a homework exercise. □

Now, suppose that we wish to find $\boldsymbol{x} \in \mathbb{C}^n$, such that $\mathsf{A}\boldsymbol{x} = \boldsymbol{b}$, where $\mathsf{A} \in \mathbb{C}^{n\times n}$ is invertible and $\boldsymbol{b} \in \mathbb{C}^n$ is given. We sometimes call A and $\boldsymbol{b}$ the data of the problem. We can imagine a scenario in which the coefficient matrix A and the vector $\boldsymbol{b}$ are perturbed during the process of storing their values or solving the system of equations. Let $\delta\mathsf{A} \in \mathbb{C}^{n\times n}$ and $\delta\boldsymbol{b} \in \mathbb{C}^n$ be known (or estimable) perturbations of the data. The problem that is actually solved is

$$(\mathsf{A} + \delta\mathsf{A})(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{b} + \delta\boldsymbol{b}.$$

The question is this: how large is the relative error? In other words, how large is $\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$? Formally, the perturbation in $\delta\boldsymbol{x} \in \mathbb{C}^n$ is

$$\delta\boldsymbol{x} = (\mathsf{A} + \delta\mathsf{A})^{-1}(\boldsymbol{b} + \delta\boldsymbol{b}) - \boldsymbol{x},$$

provided $\mathsf{A} + \delta\mathsf{A}$ is invertible. Observe that since A is invertible we have

$$(\mathsf{A} + \delta\mathsf{A})^{-1} = (\mathsf{A}(\mathsf{I}_n + \mathsf{A}^{-1}\delta\mathsf{A}))^{-1} = (\mathsf{I}_n + \mathsf{A}^{-1}\delta\mathsf{A})^{-1}\mathsf{A}^{-1}.$$

Therefore, we have reduced the question of the invertibility of $\mathsf{A} + \delta\mathsf{A}$ to a more general question: given $\mathsf{M} \in \mathbb{C}^{n\times n}$, when is $\mathsf{I}_n \pm \mathsf{M}$ invertible?

Theorem 4.2.6 (Neumann Series)**.** *Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is an induced matrix norm with respect to the vector norm $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$. Let $\mathsf{M} \in \mathbb{C}^{n\times n}$ with $\|\mathsf{M}\| < 1$, then $\mathsf{I}_n - \mathsf{M}$ is invertible,*

$$\|(\mathsf{I}_n - \mathsf{M})^{-1}\| \leq \frac{1}{1 - \|\mathsf{M}\|},$$

and

$$(\mathsf{I}_n - \mathsf{M})^{-1} = \sum_{k=0}^{\infty} \mathsf{M}^k.$$

Proof. Using the reverse triangle inequality and consistency, since $\|\mathsf{M}\| < 1$, for any $\boldsymbol{x} \in \mathbb{C}^n$,

$$\|(\mathsf{I}_n - \mathsf{M})\boldsymbol{x}\| \geq \big|\|\boldsymbol{x}\| - \|\mathsf{M}\boldsymbol{x}\|\big| \geq (1 - \|\mathsf{M}\|)\|\boldsymbol{x}\|.$$

This inequality implies that, if $(\mathsf{I}_n - \mathsf{M})\boldsymbol{x} = \mathbf{0}$ then $\boldsymbol{x} = \mathbf{0}$. Therefore, $\mathsf{I}_n - \mathsf{M}$ is invertible.

To obtain the norm estimate notice that

$$\begin{aligned} 1 = \|\mathsf{I}_n\| &= \|(\mathsf{I}_n - \mathsf{M})(\mathsf{I}_n - \mathsf{M})^{-1}\| \\ &= \left\|(\mathsf{I}_n - \mathsf{M})^{-1} - \mathsf{M}(\mathsf{I}_n - \mathsf{M})^{-1}\right\| \\ &\geq \|(\mathsf{I}_n - \mathsf{M})^{-1}\| - \|\mathsf{M}\|\|(\mathsf{I}_n - \mathsf{M})^{-1}\|, \end{aligned}$$

where we have used the reverse triangle inequality for the matrix norm, which holds as for vector norms, and sub-multiplicativity. The upper bound of the quantity $\|(\mathsf{I}_n - \mathsf{M})^{-1}\|$ now follows.

Finally, consider

$$\mathsf{S}_\ell := \sum_{k=0}^{\ell} \mathsf{M}^k.$$

We will show that $\mathsf{S}_\ell(\mathsf{I}_n - \mathsf{M}) \to \mathsf{I}_n$ as $\ell \to \infty$. Indeed,

$$\mathsf{S}_\ell(\mathsf{I}_n - \mathsf{M}) = \sum_{k=0}^{\ell} \mathsf{M}^k(\mathsf{I}_n - \mathsf{M}) = \sum_{k=0}^{\ell} \mathsf{M}^k - \sum_{k=0}^{\ell} \mathsf{M}^{k+1} = \mathsf{I}_n - \mathsf{M}^{\ell+1},$$

which shows that, as $\ell \to \infty$,

$$\|\mathsf{S}_\ell(\mathsf{I}_n - \mathsf{M}) - \mathsf{I}_n\| = \|\mathsf{M}^{\ell+1}\| \leq \|\mathsf{M}\|^{\ell+1} \to 0,$$

using the sub-multiplicativity of the induced norm and the fact that $\|\mathsf{M}\| < 1$. □

Corollary 4.2.7. *Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is an induced matrix norm with respect to the vector norm $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$. Let $\mathsf{M} \in \mathbb{C}^{n\times n}$ with $\|\mathsf{M}\| < 1$, then $\mathsf{I}_n + \mathsf{M}$ is invertible and*

$$\|(\mathsf{I}_n + \mathsf{M})^{-1}\| \leq \frac{1}{1 - \|\mathsf{M}\|}.$$

Corollary 4.2.8. *Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is an induced matrix norm with respect to the vector norm $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$. If $\mathsf{S} \in \mathbb{C}^{n\times n}$ is invertible and $\mathsf{T} \in \mathbb{C}^{n\times n}$ satifies*

$$\left\|\mathsf{S}^{-1}\right\| \|\mathsf{S} - \mathsf{T}\| < 1,$$

then T is invertible.

Proof. Notice that
$$\mathsf{T} = \mathsf{S}(\mathbf{I}_n - (\mathbf{I}_n - \mathsf{S}^{-1}\mathsf{T})),$$
and, therefore, T will be invertible provided $\mathbf{I}_n - (\mathbf{I}_n - \mathsf{S}^{-1}\mathsf{T})$ is invertibble. Defining $\mathsf{M} = \mathbf{I}_n - \mathsf{S}^{-1}\mathsf{T}$, we realize that we need $\|\mathsf{M}\| < 1$. Observe that
$$\|\mathsf{M}\| = \|\mathbf{I}_n - \mathsf{S}^{-1}\mathsf{T}\| = \|\mathsf{S}^{-1}(\mathsf{S} - \mathsf{T})\| \leq \|\mathsf{S}^{-1}\|\|\mathsf{S} - \mathsf{T}\| < 1.$$
□

Let us first begin by assuming that $\delta\boldsymbol{b} = 0$.

Theorem 4.2.9. *Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is an induced matrix norm with respect to the vector norm $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$. Suppose that $\boldsymbol{x} \in \mathbb{C}^n$ solves $\mathsf{A}\boldsymbol{x} = \boldsymbol{b}$, where $\mathsf{A} \in \mathbb{C}^{n\times n}$ is invertible and $\boldsymbol{b} \in \mathbb{C}^n$ is known. Assume that $\delta\mathsf{A} \in \mathbb{C}^{n\times n}$ satisfies $\|\mathsf{A}^{-1}\delta\mathsf{A}\| < 1$ and that $\delta\boldsymbol{x} \in \mathbb{C}^n$ satisfies the perturbed problem*
$$(\mathsf{A} + \delta\mathsf{A})(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{b}.$$
Then $\delta\boldsymbol{x}$ is uniquely determined and
$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \frac{\kappa(\mathsf{A})}{1 - \kappa(\mathsf{A})\frac{\|\delta\mathsf{A}\|}{\|\mathsf{A}\|}}\frac{\|\delta\mathsf{A}\|}{\|\mathsf{A}\|}$$

Proof. Since $\mathsf{M} := -\mathsf{A}^{-1}\delta\mathsf{A}$ satisfies $\|\mathsf{M}\| < 1$ we know that $\mathsf{A} + \delta\mathsf{A}$ is invertible. Therefore, $\delta\boldsymbol{x}$ exists and is unique. In addition, we have
$$(\mathsf{A} + \delta\mathsf{A})^{-1} = (\mathsf{A}(\mathbf{I}_n - \mathsf{M}))^{-1} = (\mathbf{I}_n - \mathsf{M})^{-1}\mathsf{A}^{-1}$$
and $\|(\mathbf{I}_n - \mathsf{M})^{-1}\| \leq \frac{1}{1 - \|\mathsf{M}\|}$. Moreover,
$$\|\mathsf{M}\| \leq \|\mathsf{A}^{-1}\|\|\delta\mathsf{A}\|$$
implies
$$\frac{1}{1 - \|\mathsf{M}\|} \leq \frac{1}{1 - \|\mathsf{A}^{-1}\|\|\delta\mathsf{A}\|}.$$

Now
$$\begin{aligned}
\delta\boldsymbol{x} &= (\mathsf{A} + \delta\mathsf{A})^{-1}\boldsymbol{b} - \mathsf{A}^{-1}\boldsymbol{b} \\
&= (\mathbf{I}_n - \mathsf{M})^{-1}\mathsf{A}^{-1}\boldsymbol{b} - \mathsf{A}^{-1}\boldsymbol{b} \\
&= (\mathbf{I}_n - \mathsf{M})^{-1}(\mathsf{A}^{-1}\boldsymbol{b} - (\mathbf{I}_n - \mathsf{M})\mathsf{A}^{-1}\boldsymbol{b}) \\
&= (\mathbf{I}_n - \mathsf{M})^{-1}\mathsf{M}\mathsf{A}^{-1}\boldsymbol{b} \\
&= (\mathbf{I}_n - \mathsf{M})^{-1}\mathsf{M}\boldsymbol{x}.
\end{aligned}$$

Consequently,

$$\begin{aligned}\|\delta\boldsymbol{x}\| &\le \|(\mathsf{I}_n - \mathsf{M})^{-1}\| \cdot \|\mathsf{M}\| \cdot \|\boldsymbol{x}\| \\ &\le \frac{\|\mathsf{A}^{-1}\|\|\delta\mathsf{A}\|}{1 - \|\mathsf{A}^{-1}\|\|\delta\mathsf{A}\|}\|\boldsymbol{x}\| \\ &= \frac{\kappa(\mathsf{A})}{1 - \kappa(\mathsf{A})\frac{\|\delta\mathsf{A}\|}{\|\mathsf{A}\|}} \frac{\|\delta\mathsf{A}\|}{\|\mathsf{A}\|}\|\boldsymbol{x}\|.\end{aligned}$$

The result follows. $\square$

To end the discussion, let us see what happens when we perturb both A and $\boldsymbol{b}$.

Theorem 4.2.10. *Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is an induced matrix norm with respect to the vector norm $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$. Suppose that $\boldsymbol{x} \in \mathbb{C}^n$ solves $\mathsf{A}\boldsymbol{x} = \boldsymbol{b}$, where $\mathsf{A} \in \mathbb{C}^{n\times n}$ is invertible and $\boldsymbol{b} \in \mathbb{C}^n$ is known. Assume that $\delta\mathsf{A} \in \mathbb{C}^{n\times n}$ satisfies $\|\mathsf{A}^{-1}\delta\mathsf{A}\| < 1$, $\delta\boldsymbol{b} \in \mathbb{C}^n$ is given, and $\delta\boldsymbol{x} \in \mathbb{C}^n$ satisfies the perturbed problem*

$$(\mathsf{A} + \delta\mathsf{A})(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{b} + \delta\boldsymbol{b}.$$

Then $\delta\boldsymbol{x}$ is uniquely determined and

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \le \frac{\kappa(\mathsf{A})}{1 - \kappa(\mathsf{A})\frac{\|\delta\mathsf{A}\|}{\|\mathsf{A}\|}} \left(\frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|} + \frac{\|\delta\mathsf{A}\|}{\|\mathsf{A}\|}\right).$$

Proof. Let $\mathsf{M} = -\mathsf{A}^{-1}\delta\mathsf{A}$. We then have that $\boldsymbol{x} = \mathsf{A}^{-1}\boldsymbol{b}$ and $\boldsymbol{x} + \delta\boldsymbol{x} = (\mathsf{I}_n - \mathsf{M})^{-1}\mathsf{A}^{-1}(\boldsymbol{b} + \delta\boldsymbol{b})$. Therefore,

$$\begin{aligned}\delta\boldsymbol{x} &= (\mathsf{I}_n - \mathsf{M})^{-1}\mathsf{A}^{-1}(\boldsymbol{b} + \delta\boldsymbol{b}) - \mathsf{A}^{-1}\boldsymbol{b} \\ &= (\mathsf{I}_n - \mathsf{M})^{-1}\left(\mathsf{A}^{-1}\boldsymbol{b} + \mathsf{A}^{-1}\delta\boldsymbol{b} - (\mathsf{I}_n - \mathsf{M})\mathsf{A}^{-1}\boldsymbol{b}\right) \\ &= (\mathsf{I}_n - \mathsf{M})^{-1}(\mathsf{A}^{-1}\delta\boldsymbol{b} - \mathsf{M}\mathsf{A}^{-1}\boldsymbol{b})\end{aligned}$$

This shows that

$$\|\delta\boldsymbol{x}\| \le \frac{1}{1 - \kappa(\mathsf{A})\frac{\|\delta\mathsf{A}\|}{\|\mathsf{A}\|}} \left(\|\mathsf{A}^{-1}\delta\boldsymbol{b}\| + \|\mathsf{M}\mathsf{A}^{-1}\boldsymbol{b}\|\right).$$

Notice also that

$$\|\mathsf{M}\mathsf{A}^{-1}\boldsymbol{b}\| = \|\mathsf{M}\boldsymbol{x}\| \le \|\mathsf{M}\|\|\boldsymbol{x}\| \le \|\mathsf{A}^{-1}\| \cdot \|\delta\mathsf{A}\| \cdot \|\boldsymbol{x}\| = \kappa(\mathsf{A})\frac{\|\delta\mathsf{A}\|}{\|\mathsf{A}\|}\|\boldsymbol{x}\|$$

and

$$\|\mathsf{A}^{-1}\delta\boldsymbol{b}\| \le \|\mathsf{A}^{-1}\|\|\delta\boldsymbol{b}\|\frac{\|\mathsf{A}\boldsymbol{x}\|}{\|\mathsf{A}\boldsymbol{x}\|} \le \kappa(\mathsf{A})\frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}\|\boldsymbol{x}\|.$$

The previous three inequalities, when combined, yield

$$\|\delta\boldsymbol{x}\| \le \frac{\kappa(\mathsf{A})}{1-\kappa(\mathsf{A})\frac{\|\delta\mathsf{A}\|}{\|\mathsf{A}\|}}\left(\frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}+\frac{\|\delta\mathsf{A}\|}{\|\mathsf{A}\|}\right)\|\boldsymbol{x}\|,$$

as we intended to show. □

4.3 Problems

1. Suppose that $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}$ is the induced norm with respect to the vector norms $\|\cdot\|_{\mathbb{C}^m}$ and $\|\cdot\|_{\mathbb{C}^n}$, and $\mathsf{A} \in \mathbb{C}^{m\times n}$, prove that the function $\|\mathsf{A}(\,\cdot\,)\|_{\mathbb{C}^m} : \mathbb{C}^n \to \mathbb{R}$ is uniformly continuous. Use this fact to prove that there is vector $\boldsymbol{x} \in S^{n-1}_{\mathbb{C}^n}$ such that
$$\|\mathsf{A}\| = \|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m}\,.$$

2. Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is the induced norm with respect to the vector norm $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$. Let $\mathsf{A} \in \mathbb{C}^{n\times n}$ be invertible. Prove that
$$\frac{1}{\|\mathsf{A}^{-1}\|} = \min_{\boldsymbol{y}\in\mathbb{C}^n_\star}\frac{\|\mathsf{A}\boldsymbol{y}\|}{\|\boldsymbol{y}\|}\,.$$

3. Let $\mathsf{A} \in \mathbb{C}^{n\times n}$ be nonsingular. Show that the condition numbers $\kappa_\infty(\mathsf{A})$ and $\kappa_1(\mathsf{A})$ will not change after permutation of rows or columns.

4. Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is a matrix norm and κ is the condition number defined with respect to it. Let $\mathsf{A} \in \mathbb{C}^{n\times n}$ be nonsingular and $0 \ne \alpha \in \mathbb{C}$. Show that $\kappa(\alpha\mathsf{A}) = \kappa(\mathsf{A})$.

5. Show that if $\mathsf{Q} \in \mathbb{C}^n$ is unitary, then $\kappa_2(\mathsf{Q}) = 1$.

6. Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is the induced norm with respect to the vector norm $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$ and κ is the condition number defined with respect to this norm. Let $\mathsf{A} \in \mathbb{C}^n$ be invertible. Show that
$$\kappa(\mathsf{A}) \ge \frac{\max_{\lambda\in\sigma(\mathsf{A})}|\lambda|}{\min_{\lambda\in\sigma(\mathsf{A})}|\lambda|},$$
Show that if $\mathsf{A}^H = \mathsf{A}$, then equality holds for the 2-norm.

7. Let $\mathsf{A} = \mathsf{S}^H\mathsf{S}$ with $\mathsf{S} \in \mathbb{C}^{n\times n}$ nonsingular. Give an expression for $\kappa_2(\mathsf{A})$ in terms of $\kappa_2(\mathsf{S})$.

8. Let $\mathsf{S}, \mathsf{T} \in \mathbb{C}^{n\times n}$ with S invertible and $\|\mathsf{T} - \mathsf{S}\| \cdot \|\mathsf{S}^{-1}\| < 1$. Show that T is invertible and
$$\|\mathsf{T}^{-1}\| \le \frac{1}{1-q}\|\mathsf{S}^{-1}\|, \quad q = \|\mathsf{S} - \mathsf{T}\| \cdot \|\mathsf{S}^{-1}\|.$$
Prove that the set of invertible matrices is *open* in $\mathbb{C}^{n\times n}$.

9. Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is the induced norm with respect to the vector norm $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$. Let $\mathsf{A} \in \mathbb{C}^{n\times n}$ be invertible and suppose $\boldsymbol{b} \in \mathbb{C}^n_\star$. Suppose $\boldsymbol{x} \in \mathbb{C}^n$ satisfies $\mathsf{A}\boldsymbol{x} = \boldsymbol{b}$. Let the perturbations $\delta\boldsymbol{x}$, $\delta\boldsymbol{b} \in \mathbb{C}^n$ satisfy $\mathsf{A}\delta\boldsymbol{x} = \delta\boldsymbol{b}$, so that $\mathsf{A}(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{b} + \delta\boldsymbol{b}$.

   (a) Prove the error (or perturbation) estimate
   $$\frac{1}{\kappa(\mathsf{A})}\frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|} \le \frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \le \kappa(\mathsf{A})\frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|} .$$

   (b) Show that for any invertible matrix A, the upper bound for $\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$ above can be attained for suitable choices of $\boldsymbol{b}$ and $\delta\boldsymbol{b}$.

10. Let $\mathsf{A} \in \mathbb{C}^{n\times n}$ be HPD. Define $\|\boldsymbol{x}\|_\mathsf{A} : \mathbb{C}^n \to \mathbb{R}$ via $\|\boldsymbol{x}\|_\mathsf{A} := \sqrt{\boldsymbol{x}^H\mathsf{A}\boldsymbol{x}}$. Prove that this is a vector norm, and satisfies the estimates
$$\sqrt{\lambda_1}\,\|\boldsymbol{x}\|_2 \le \|\boldsymbol{x}\|_\mathsf{A} \le \sqrt{\lambda_n}\,\|\boldsymbol{x}\|_2 ,$$
where $0 < \lambda_1 \le \cdots \le \lambda_n$ are the eigenvalues of A, with both inequalities attainable (though perhaps not simultaneously) for suitable choices of $\boldsymbol{x}$.

11. For any $\boldsymbol{x} \in \mathbb{C}^n$, prove
$$\begin{aligned}\|\boldsymbol{x}\|_\infty &\le \|\boldsymbol{x}\|_1 \le n\,\|\boldsymbol{x}\|_\infty ,\\ \|\boldsymbol{x}\|_\infty &\le \|\boldsymbol{x}\|_2 \le \sqrt{n}\,\|\boldsymbol{x}\|_\infty ,\\ \|\boldsymbol{x}\|_2 &\le \|\boldsymbol{x}\|_1 \le \sqrt{n}\,\|\boldsymbol{x}\|_2 .\end{aligned}$$

12. For any $\mathsf{A} \in \mathbb{C}^{n\times n}$, prove the following:

   (a)
   $$\frac{1}{\sqrt{n}}\|\mathsf{A}\|_2 \le \|\mathsf{A}\|_\infty \le \sqrt{n}\|\mathsf{A}\|_2.$$

(b) If A is nonsingular

$$\frac{1}{n} \le \frac{\kappa_\infty(\mathsf{A})}{\kappa_2(\mathsf{A})} \le n.$$

13. Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is the induced norm with respect to the vector norm $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$. Suppose that $\mathsf{A}, \mathsf{B} \in \mathbb{C}^{n\times n}$ and A is non-singular and B is singular. Prove that

$$\frac{1}{\kappa(\mathsf{A})} \le \frac{\|\mathsf{A}-\mathsf{B}\|}{\|\mathsf{A}\|} ,$$

where $\kappa(\mathsf{A}) = \|\mathsf{A}\| \cdot \left\|\mathsf{A}^{-1}\right\|$.

Note: This formula is useful in a couple of ways. First, it says that if A is close in norm to a singular matrix B, then $\kappa(\mathsf{A})$ will be very large. Thus, nearly singular matrices are ill-conditioned. Second, this formula gives an upper bound on $\kappa(\mathsf{A})^{-1}$.

14. Let $\mathsf{A} \in \mathbb{C}^{n\times n}$ be invertible. Prove that

$$\frac{1}{\kappa_2(\mathsf{A})} = \inf_{\det(\mathsf{B})=0} \frac{\|\mathsf{A}-\mathsf{B}\|_2}{\|\mathsf{A}\|_2} .$$

15. Suppose

$$\mathsf{A} = \begin{bmatrix} 1.0000 & 2.0000 \\ 1.0001 & 2.0000 \end{bmatrix} .$$

(a) Calculate $\kappa_1(\mathsf{A}) := \|\mathsf{A}\|_1 \cdot \left\|\mathsf{A}^{-1}\right\|_1$ and $\kappa_\infty(\mathsf{A}) := \|\mathsf{A}\|_\infty \cdot \left\|\mathsf{A}^{-1}\right\|_\infty$.

(b) Use the result of problem 13 to obtain upper bounds on $\kappa_1(\mathsf{A})^{-1}$ and also on $\kappa_\infty(\mathsf{A})^{-1}$.

(c) Suppose that you wish to solve $\mathsf{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{b} = \begin{bmatrix} 3.0000 \\ 3.0001 \end{bmatrix}$. Instead of $\boldsymbol{x}$ you obtain the approximation $\boldsymbol{x}' = \boldsymbol{x} + \delta\mathbf{x} = \begin{bmatrix} 0.0000 \\ 1.5000 \end{bmatrix}$. For this approximation you discover $\boldsymbol{b}' = \boldsymbol{b} + \delta\boldsymbol{b} = \begin{bmatrix} 3.0000 \\ 3.0000 \end{bmatrix}$, where $\mathsf{A}\boldsymbol{x}' = \boldsymbol{b}'$. Calculate $\|\delta\boldsymbol{x}\|_1 / \|\boldsymbol{x}\|_1$ exactly. (You will need the exact solution, of course). Then use the general estimate

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \le \kappa(\mathsf{A})\frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}$$

to obtain an upper bound for $\|\delta\boldsymbol{x}\|_1 / \|\boldsymbol{x}\|_1$. How good is $\|\delta\boldsymbol{b}\|_1 / \|\boldsymbol{b}\|_1$ as indicator of the size of $\|\delta\boldsymbol{x}\|_1 / \|\boldsymbol{x}\|_1$.

16. Suppose
$$\mathsf{A} = \begin{bmatrix} a & b \\ b & a \end{bmatrix},$$
where a and b are real numbers. Show that the subordinate matrix norms satisfy $\|\mathsf{A}\|_1 = \|\mathsf{A}\|_2 = \|\mathsf{A}\|_\infty$.

17. Suppose that
$$\mathsf{A} = \begin{bmatrix} a & b \\ b & -a \end{bmatrix},$$
where a and b are real numbers. Show $\|\mathsf{A}\|_2 = (a^2 + b^2)^{1/2}$.

18. Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is the induced norm with respect to the vector norm $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$. Show that if λ is an eigenvalue of $\mathsf{A}^H\mathsf{A}$, where $\mathsf{A} \in \mathbb{C}^{n\times n}$, then
$$0 \le \lambda \le \left\|\mathsf{A}^H\right\| \|\mathsf{A}\| .$$

19. Suppose that $\mathsf{A} \in \mathbb{C}^{n\times n}$ is invertible. Show that
$$\kappa_2(\mathsf{A}) = \sqrt{\frac{\lambda_n}{\lambda_1}},$$
where λ_n is the largest eigenvalue of $\mathsf{B} := \mathsf{A}^H\mathsf{A}$, and λ_1 is the smallest eigenvalue of B.

20. Suppose that A is HPD with eigenvalues $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. Prove that
$$\kappa_2(\mathsf{A}) = \frac{\lambda_n}{\lambda_1}.$$

21. Suppose that $\mathsf{A} \in \mathbb{C}^{n\times n}$ is invertible. Use the results of problems 18 and 19 to show that
$$\kappa_2(\mathsf{A}) \le \sqrt{\kappa_1(\mathsf{A})\,\kappa_\infty(\mathsf{A})} .$$