7. Norms, map norms, and the condition of a basis

Assume that V is a basis for the nontrivial linear subspace X of the inner product space Y. The coordinate vector a for $x \in X$ is the unique solution of the equation

$$V? = x.$$

We may not be able to compute the solution exactly. Even if we know the entries of the solution exactly, as common fractions say, we may not be able to use them exactly if we use some floating-point arithmetic, as is common. It is for this reason that one is interested in gauging the effect of an erroneous coordinate vector $\widehat{a}$ on the accuracy of $V\widehat{a}$ as a representation for $x = Va$.

How to judge the error by the residual

Since, presumably, we do not know a, we cannot compute the **error**

$$\varepsilon := a - \widehat{a};$$

we can only compute the **residual**

$$r := x - V\widehat{a}.$$

Nevertheless, can we judge the error by the residual? Does a 'small' **relative residual**

$$\|r\|/\|x\|$$

imply a 'small' **relative error**

$$\|\varepsilon\|/\|a\| \ ?$$

By definition, the **condition** (or, **condition number**) $\kappa(V)$ of the basis V is the greatest factor by which the relative error, $\|\varepsilon\|/\|a\|$, can exceed the relative residual, $\|r\|/\|x\| = \|V\varepsilon\|/\|Va\|$; i.e.,

$$\kappa(V) := \sup_{a,\varepsilon} \frac{\|\varepsilon\|/\|a\|}{\|V\varepsilon\|/\|Va\|}. \tag{7.1}$$

However, by interchanging here the roles of a and ε and then taking reciprocals, this also says that

$$1/\kappa(V) = \inf_{\varepsilon,a} \frac{\|\varepsilon\|/\|a\|}{\|V\varepsilon\|/\|Va\|}.$$

Hence, altogether,

$$\frac{1}{\kappa(V)} \frac{\|r\|}{\|x\|} \ \le \ \frac{\|\varepsilon\|}{\|a\|} \ \le \ \kappa(V) \frac{\|r\|}{\|x\|}. \tag{7.2}$$

In other words, *the larger the condition number, the less information about the size of the relative error is provided by the size of the relative residual.*

For a better feel for the condition number, note that we can also write the formula (7.1) for $\kappa(V)$ in the following fashion:

$$\kappa(V) = \sup_{\varepsilon} \frac{\|\varepsilon\|}{\|V\varepsilon\|} \sup_{a} \frac{\|Va\|}{\|a\|}.$$

Also,

$$\|Va\|/\|a\| = \|V(a/\|a\|)\|,$$

with $a/\|a\|$ *normalized*, i.e., of norm 1. Hence, altogether,

$$\kappa(V) = \frac{\sup\{\|Va\| : \|a\| = 1\}}{\inf\{\|Va\| : \|a\| = 1\}}. \tag{7.3}$$

This says that we can visualize the condition number $\kappa(V)$ in the following way; see (7.5)Figure. Consider the image

$$\{Va : \|a\| = 1\} \qquad (7.4)$$

under V of the **unit sphere**

$$\{a \in \mathbb{F}^n : \|a\| = 1\}$$

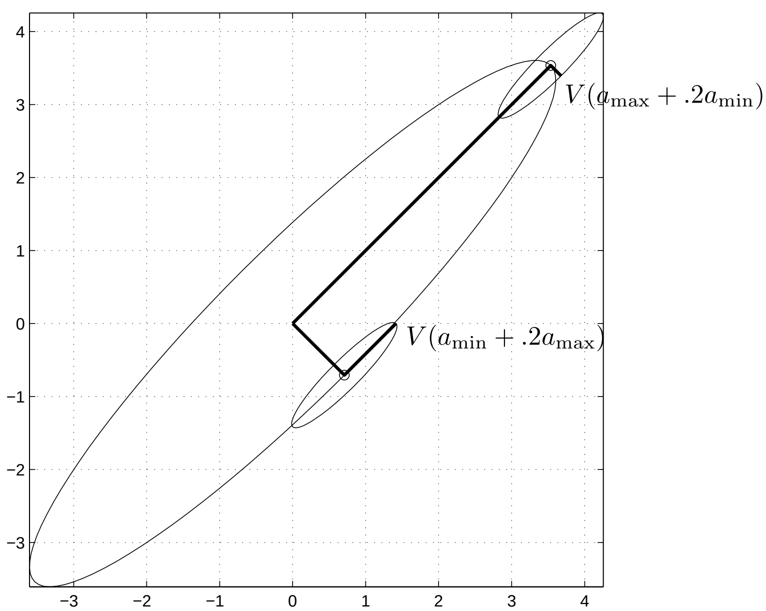
in $\mathbb{F}^n$. It will be some kind of ellipsoid, symmetric with respect to the origin. In particular, there will be a point $a_{\max}$ with $\|a_{\max}\| = 1$ for which $Va_{\max}$ will be as far from the origin as possible. There will also be a point $a_{\min}$ with $\|a_{\min}\| = 1$ for which $Va_{\min}$ will be as close to the origin as possible. In other words,

$$\kappa(V) = \|Va_{\max}\|/\|Va_{\min}\|,$$

saying that the condition number gives the ratio of the largest to the smallest diameter of the ellipsoid (7.4). The larger the condition number, the skinnier is the ellipsoid.

In particular, if $a = a_{\max}$ while $\varepsilon = a_{\min}$, then the relative error is 1 while the relative residual is $\|Va_{\min}\|/\|Va_{\max}\|$, and this is tiny to the extent that the ellipsoid is 'skinny'.

On the other hand, if $a = a_{\min}$ while $\varepsilon = a_{\max}$, then the relative error is still 1, but now the relative residual is $\|Va_{\max}\|/\|Va_{\min}\|$, and this is large to the extent that the ellipsoid is 'skinny'.



(7.5) Figure. Extreme effects of a 20% relative error on the relative residual, for $V = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$.

The worst-conditioned column maps V are those that fail to be 1-1 since, for them, $Va_{\min} = 0$, hence $\kappa(V) = \infty$.

On the other extreme, it follows directly from (7.3) that $\kappa(V) \ge 1$, and this lower bound is reached by any o.n. basis V since any o.n. basis is an isometry, by (6.23), i.e., $\|Va\| = \|a\|$ for all $a \in \mathbb{F}^n$. Thus o.n. bases are best-conditioned, and rightfully prized for that. It was for this reason that we took the trouble to prove that every finite-dimensional linear subspace of an inner product space has o.n. bases, and even discussed just how to construct such bases.

The map norm

As we now explain, the numbers $\|Va_{\max}\| = \max\{\|Va\| : \|a\| = 1\}$ and $1/\|Va_{\min}\| = 1/\min\{\|Va\| : \|a\| = 1\}$ both are examples of a map norm according to the following

(7.6) Definition: The **map norm**, $\|A\|$, of $A \in L(X,Y)$ is the smallest nonnegative number c for which

$$\|Ax\| \le c\|x\|, \quad \forall x \in X.$$

If X is trivial, then $\|A\| = 0$ for the sole $A \in L(X,Y)$. Otherwise

$$\|A\| = \sup_{x\ne0} \|Ax\|/\|x\| = \sup\{\|Ax\| : \|x\| = 1\}. \tag{7.7}$$

Here, the last equality follows from the absolute homogeneity of the norm and the homogeneity of A which combine to permit the conclusions that

$$\|Ax\|/\|x\| = \|A(x/\|x\|)\| \quad \text{and} \quad \|(x/\|x\|)\| = 1.$$

In these notes, we are only interested in *finite-dimensional* X and, for such X,

$$\|A\| = \max_{x\ne0} \|Ax\|/\|x\| = \max\{\|Ax\| : \|x\| = 1\}. \tag{7.8}$$

The reason for this is beyond the scope of these notes, but is now stated for the record: If X is finite-dimensional, then

$$F : x \mapsto \|Ax\|$$

is continuous and the *unit sphere*

$$\{x \in X : \|x\| = 1\}$$

is compact, hence F achieves its maximum value on that sphere. (For the same reason, F also achieves its minimum value on the unit sphere, and this justifies the existence of $a_{\max}$ and $a_{\min}$ in the preceding section.)

We conclude that *determination* of the map norm is a two-part process, as formalized in the following.

(7.9) Calculation of $\|A\|$: The number c equals the norm $\|A\|$ if and only if

(i) for all x, $\|Ax\| \le c\|x\|$; and

(ii) for some $x \ne 0$, $\|Ax\| \ge c\|x\|$.

The first says that $\|A\| \le c$, while second says that $\|A\| \ge c$, hence, together they say that $\|A\| = c$.

(7.10) Example: We compute $\|A\|$ in case $A \in \mathbb{F}^{m\times n}$ is of the simple form

$$A = [v][w]^c = vw^c$$

for some $v \in \mathbb{F}^m$ and some $w \in \mathbb{F}^n$. Since

$$Ax = (vw^c)x = v(w^cx),$$

we have

$$\|(vw^c)x\| = \|v\||w^cx| \le \|v\|\|w\|\|x\|,$$

the equality by the absolute homogeneity of the norm, and the inequality by (6.15)Cauchy's Inequality. This shows that $\|vw^c\| \le \|v\|\|w\|$. On the other hand, for the specific choice $x = w$, we get $(vw^c)w = v(w^cw) = v\|w\|^2$, hence $\|(vw^c)w\| = \|v\|\|w\|\|w\|$. Assuming that $w \ne 0$, this shows that $\|vw^c\| \ge \|v\|\|w\|$. However, this inequality is trivially true in case $w = 0$ since then $vw^c = 0$. So, altogether, we have that

$$\|vw^c\| = \|v\|\|w\|.$$

Note that we have, incidentally, proved that, for any $v \in \mathbb{F}^n$,

$$\|[v]\| = \|v\| = \|[v]^c\|. \tag{7.11}$$

□

As another example, note that, if also $B \in L(Y, Z)$ for some inner product space Z, then BA is defined and

$$\|(BA)x\| = \|B(Ax)\| \le \|B\| \, \|Ax\| \le \|B\| \, \|A\| \|x\|.$$

Therefore,

$$\|BA\| \le \|B\| \, \|A\|. \tag{7.12}$$

We are ready to discuss the condition (7.3) of a basis V in terms of map norms.

Directly from (7.8), $\max\{\|Va\| : \|a\| = 1\} = \|V\|$.

(7.13) Proposition: If $A \in L(X, Y)$ is invertible and $X \ne \{0\}$ is finite-dimensional, then

$$\|A^{-1}\| = 1/\min\{\|Ax\| : \|x\| = 1\}.$$

Proof: Since A is invertible, $y \in Y$ is nonzero if and only if $y = Ax$ for some nonzero $x \in X$. Hence,

$$\|A^{-1}\| = \max_{y \ne 0} \frac{\|A^{-1}y\|}{\|y\|} = \max_{x \ne 0} \frac{\|A^{-1}Ax\|}{\|Ax\|} = 1/\min_{x \ne 0} \frac{\|Ax\|}{\|x\|},$$

and this equals $1/\min\{\|Ax\| : \|x\| = 1\}$ by the absolute homogeneity of the norm and the homogeneity of A. □

In particular, $1/\|A^{-1}\|$ is the largest number c for which

$$c\|x\| \le \|Ax\|, \quad \forall x \in X.$$

We conclude that

$$\kappa(V) = \|V\| \|V^{-1}\|. \tag{7.14}$$

7.1 Complement (7.13)Proposition by discussing the situation when $X = \{0\}$.

7.2 Prove that $\kappa(V) \ge 1$ for any basis V with at least one column.

7.3 Determine $\kappa([\,])$.

Vector norms and their associated map norms

`MATLAB` provides the map norm of the matrix `A` by the statement `norm(A)` (or by the statement `norm(A,2)`, indicating that there are other map norms available).

The `norm` command gives the Euclidean norm when its argument is a ‘vector’. Specifically, `norm(v)` and `norm(v,2)` both give $\|v\| = \sqrt{v^c v}$. However, since in (present-day) `MATLAB`, everything is a matrix, there is room here for confusion since experimentation shows that `MATLAB` defines a ‘vector’ to be any 1-column matrix and any 1-row matrix. Fortunately, there is no problem with this, since, by (7.11), the norm of the *vector* v equals the norm of the *matrices* v and v^c.

□

The best explicit expression available for $\|A\|$ for an arbitrary $A \in \mathbb{F}^{m\times n}$ is the following:

$$\|A\| = \sqrt{\rho(A^{c}A)} = \sigma_1(A). \tag{7.15}$$

This formula cannot be evaluated in finitely many steps since the number $\rho(A^{c}A)$ is, by definition, the 'spectral radius' of $A^{c}A$, i.e., the smallest possible radius of a disk centered at the origin that contains all the eigenvalues of $A^{c}A$. The 2-norm of A also equals $\sigma_1(A)$ which is, by definition, the largest 'singular value' of A. In general, one can only compute approximations to this number.

For this reason (and others), other vector norms are in common use, among them the **max-norm**

$$\|x\|_\infty := \max_j |x_j|, \quad \forall x \in \mathbb{F}^n,$$

for which the associated map norm is easily computable. It is

$$\|A\|_\infty := \max_{x\neq 0} \|Ax\|_\infty/\|x\|_\infty = \max_i \sum_j |A(i,j)| = \max_i \|A(i,:)\|_1, \tag{7.16}$$

with

$$\|v\|_1 := \sum_j |v_j| \tag{7.17}$$

yet another vector norm, the socalled **1-norm**. The map norm associated with the 1-norm is also easily computable. It is

$$\|A\|_1 := \max_{x\neq 0} \|Ax\|_1/\|x\|_1 = \max_j \sum_i |A(i,j)| = \max_j \|A(:,j)\|_1 = \|A^{t}\|_\infty = \|A^{c}\|_\infty. \tag{7.18}$$

In this connection, the Euclidean norm is also known as the **2-norm**, since

$$\|x\| = \sqrt{x^{c}x} = \sqrt{\sum_j |x_j|^2} =: \|x\|_2.$$

Therefore, when it is important, one writes the corresponding map-norm with a subscript 2, too. For example, compare (7.18) with

$$\|A\| = \|A\|_2 = \|A^{c}\|_2 = \|A^{t}\|_2. \tag{7.19}$$

For the proof of these identities, recall from (6.15) that

$$\|x\|_2 = \max_{y\neq 0} |\langle x, y\rangle|/\|y\|_2, \qquad x \in \mathbb{F}^n. \tag{7.20}$$

Hence,

$$\|A\|_2 = \max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \max_{x\neq 0}\max_{y\neq 0} \frac{|\langle Ax, y\rangle|}{\|x\|_2\|y\|_2} = \max_{y\neq 0}\max_{x\neq 0} \frac{|\langle x, A^{c}y\rangle|}{\|x\|_2\|y\|_2} = \max_{y\neq 0} \frac{\|A^{c}y\|_2}{\|y\|_2} = \|A^{c}\|_2. \tag{7.21}$$

The equality $\|A^{t}\| = \|A^{c}\|$ holds in any of the map-norms discussed since they all depend only on the absolute values of the entries of the matrix A.

The `MATLAB` statement `norm(A,inf)` provides the norm $\|\mathtt{A}\|_\infty$ in case `A` is a 'matrix', i.e., not a 'vector'. If `A` happens to equal $[v]$ or $[v]^{t}$ for some vector v, then `norm(A,inf)` returns the max-norm of that vector, i.e., the number $\|v\|_\infty$. By (7.16), this is ok if `A` $= [v]$, but gives, in general, the wrong result if `A` $= v^{t}$. This is an additional reason for sticking with the rule of using only $(n, 1)$-matrices for representing n-vectors in `MATLAB`.

The 1-norm, $\|\mathtt{A}\|_1$, is supplied by the statement `norm(A,1)`.

□

All three (vector-)norms mentioned so far are, indeed, norms in the sense of the following definition.

(7.22) Definition: The map $\| \,\| : X \to \mathbb{R} : x \mapsto \|x\|$ is a **vector norm**, provided it is

(i) **positive definite**, i.e., $\forall\{x \in X\}\ \|x\| \ge 0$ with equality if and only if $x = 0$;

(ii) **absolutely homogeneous**, i.e., $\forall\{\alpha \in \mathbb{F}, x \in X\}\ \|\alpha x\| = |\alpha|\|x\|$;

(iii) **subadditive**, i.e., $\forall\{x, y \in X\}\ \|x + y\| \le \|x\| + \|y\|$.

This last inequality is called the **triangle inequality**, and the vector space X supplied with a vector norm is called a **normed vector space**.

The absolute value is a vector norm for the vector space $\mathbb{F} = \mathbb{F}^1$. From this, it is immediate that both the max-norm and the 1-norm are vector norms for $\mathbb{F}^n$. As to the norm $x \mapsto \sqrt{x^{\mathrm{c}}x}$ on an inner product space and, in particular, the Euclidean or 2-norm on $\mathbb{F}^n$, only the triangle inequality might still be in doubt, but it is an immediate consequence of (6.15)Cauchy's Inequality, which gives that

$$\langle x, y\rangle + \langle y, x\rangle = 2\,\mathrm{Re}\langle x, y\rangle \le 2|\langle x, y\rangle| \le 2\|x\|\|y\|,$$

and therefore:

$$\|x + y\|^2 = \|x\|^2 + \langle x, y\rangle + \langle y, x\rangle + \|y\|^2 \le (\|x\| + \|y\|)^2.$$

Also, for X finite-dimensional, and both X and Y normed vector spaces, with norms $\| \,\|_X$ and $\| \,\|_Y$ respectively, the vector space $L(X, Y)$ is a normed vector space with respect to the corresponding map norm

$$\|A\| := \|A\|_{X,Y} := \max_{x \ne 0} \frac{\|Ax\|_Y}{\|x\|_X}. \tag{7.23}$$

All statements about the map norm $\|A\|$ made in the preceding section hold for any of the map norms $\|A\|_{X,Y}$ since their proofs there use only the fact that $x \mapsto \sqrt{x^{\mathrm{c}}x}$ is a norm according to (7.22)Definition. In particular, we will feel free to consider

$$\kappa(A)_p := \|A\|_p\|A^{-1}\|_p, \quad p = 1, 2, \infty, \quad A \in \mathbb{F}^n.$$

Why all these different norms? Each norm associates with a vector just one number, and, as with bases, any particular situation may best be handled by a particular norm.

For example, in considering the condition of the power basis $V := [()^{j-1} : j = 1{:}k]$ for $\Pi_{<k}$, we might be more interested in measuring the size of the residual $p - V\widehat{a}$ in terms of the max-norm

$$\|f\|_{[c\,..\,d]} := \max\{|f(t)| : c \le t \le d\}$$

over the interval $[c\,..\,d]$ of interest, rather than in the averaging way supplied by the corresponding 2-norm

$$\left(\int_a^b |f(t)|^2\,\mathrm{d}t\right)^{1/2}.$$

In any case, any two norms on a finite-dimensional vector space are equivalent in the following sense.

(7.24) Proposition: For any two norms, $\| \,\|'$ and $\| \,\|''$, on a finite-dimensional vector space X, there exists a positive constant c so that

$$\|x\|'' \le c\|x\|', \quad \forall x \in X.$$

This is just the statement that the map norm

$$\|\,\mathrm{id}_X\| := \max_{x\neq 0} \|x\|''/\|x\|'$$

is finite.

For example, for any $x \in \mathbb{F}^n$,

$$\|x\|_1 \le \sqrt{n}\|x\|_2, \quad \text{and} \quad \|x\|_2 \le \sqrt{n}\|x\|_\infty, \quad \text{while} \quad \|x\|_1 \ge \|x\|_2 \ge \|x\|_\infty. \tag{7.25}$$

Finally, given that it is very easy to compute the max-norm $\|A\|_\infty$ of $A \in \mathbb{F}^{m\times n}$ and much harder to compute the 2-norm $\|A\| = \|A\|_2$, why does one bother at all with the 2-norm? One very important reason is the availability of a large variety of *isometries*, i.e., matrices A with

$$\|Ax\| = \|x\|, \quad \forall x.$$

Each of these provides an o.n. basis for its range, and, by (6.20)Proposition, each finite-dimensional linear subspace of an inner product space has o.n. bases.

In contrast, the only $A \in \mathbb{F}^{n\times n}$ that are isometries in the max-norm, i.e., satisfy

$$\|Ax\|_\infty = \|x\|_\infty, \quad \forall x \in \mathbb{F}^n,$$

are of the form

$$\mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_n)P,$$

with P a permutation matrix and each ε_j a scalar of absolute value 1.

For this reason, we continue to rely on the 2-norm. In fact, any norm without a subscript or other adornment is meant to be the 2-norm (or, more generally, the norm in the relevant inner product space).

7.4 Prove that, for any $\alpha \in \mathbb{F}$, the linear map $M_\alpha : X \to X : x \mapsto \alpha x$ on the normed vector space $X \neq \{0\}$ has map norm $|\alpha|$.

7.5 Prove that, for any diagonal matrix $D \in \mathbb{F}^{m\times n}$ and for $p = 1, 2, \infty$, $\|D\|_p = \max_j |D(j,j)|$.

8. Factorization and rank

The need for factoring linear maps

In order to compute with a linear map $A \in L(X,Y)$, we have to factor it through a coordinate space. This means that we have to write it as

$$A = V\Lambda^{\rm t}, \qquad \text{with } \ V \in L(\mathbb{F}^r, Y), \ \text{hence } \Lambda^{\rm t} \in L(X, \mathbb{F}^r)\ .$$

The following picture might be helpful:

$$\begin{array}{ccc} X & \stackrel{A}{\longrightarrow} & Y \\ \Lambda^{\rm t}\searrow & & \nearrow V \\ & \mathbb{F}^r & \end{array}$$

For example, recall how you apply the linear map D of differentiation to a polynomial $p \in \Pi_k$: First you get the polynomial coefficients of that polynomial, and then you write down Dp in terms of those coefficients.

To test my claim, carry out the following thought experiment: You know that there is exactly one polynomial p of degree $\le k$ that matches given ordinates at given $k+1$ distinct abscissae, i.e., that satisfies

$$p(\tau_i) = y_i, \qquad i = 0{:}k$$

for given data $(\tau_i, y_i), i = 0{:}k$. Now, try, e.g., to compute the first derivative of the polynomial p of degree ≤ 3 that satisfies $p(j) = (-1)^j$, $j = 1,2,3,4$. Can you do it without factoring the linear map $D : \Pi_3 \to \Pi_3$ through some coordinate space?

As another *example*, recall how we dealt with **coordinate map**s, i.e., the inverse of a basis. We saw that, even though a basis $V : \mathbb{F}^n \to \mathbb{F}^m$ for some linear subspace X of $\mathbb{F}^m$ is a concrete matrix, its inverse, V^{-1} is, offhand, just a formal expression. For actual work, we made use of any *matrix* $\Lambda^{\rm t} : \mathbb{F}^m \to \mathbb{F}^n$ that is 1-1 on X, thereby obtaining the *factorization*

$$V^{-1} = (\Lambda^{\rm t} V)^{-1}\Lambda^{\rm t}$$

in which $\Lambda^{\rm t} V$ is a square matrix, hence $(\Lambda^{\rm t} V)^{-1}$ is also a matrix.

The smaller one can make $\#V$ in a factorization $A = V\Lambda^{\rm t}$ of $A \in L(X,Y)$, the cheaper is the calculation of A.

Definition: The smallest r for which $A \in L(X,Y)$ can be factored as $A = V\Lambda^{\rm t}$ with $V \in L(\mathbb{F}^r, Y)$ (hence $\Lambda^{\rm t} \in L(X, \mathbb{F}^r)$) is called the **rank** of A. This is written

$$r = \operatorname{rank} A.$$

Any factorization $A = V\Lambda^{\rm t}$ with $\#V = \operatorname{rank} A$ is called **minimal**.

As an *example*,

$$A := \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} [\,1 \quad 2 \quad 3 \quad 4 \quad 5\,],$$