MATH 5330: Computational Methods of Linear Algebra

Lecture Note 1: Preliminaries

Xianyi Zeng
Department of Mathematical Sciences, UTEP

1 Linear vector spaces

A **linear space** or **vector space** over $\mathbb{R}$ (or a general field $\mathbb{F}$) is a set $\mathcal{V}$ – elements of $\mathcal{V}$ are called *vectors* and the elements of $\mathbb{R}$ (or $\mathbb{F}$) are called *scalars* – such that:

1. There is an operation *addition* (denoted by $+$): $\mathcal{V}\times\mathcal{V}\to\mathcal{V}$, such that $\mathcal{V}$ is a *Abelian group* under $+$:
   (a) $\forall \boldsymbol{v}_1,\boldsymbol{v}_2\in\mathcal{V}\Rightarrow \boldsymbol{v}_1+\boldsymbol{v}_2\in\mathcal{V}$.
   (b) $\forall \boldsymbol{v}_1,\boldsymbol{v}_2\in\mathcal{V}\Rightarrow \boldsymbol{v}_1+\boldsymbol{v}_2=\boldsymbol{v}_2+\boldsymbol{v}_1$.
   (c) $\forall \boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3\in\mathcal{V}\Rightarrow (\boldsymbol{v}_1+\boldsymbol{v}_2)+\boldsymbol{v}_3=\boldsymbol{v}_1+(\boldsymbol{v}_2+\boldsymbol{v}_3)$.
   (d) There exists an identity element, denoted by $\mathbf{0}$, such that $\forall\boldsymbol{v}\in\mathcal{V}$ there is $\boldsymbol{v}+\mathbf{0}=\boldsymbol{v}$.
   (e) For any $\boldsymbol{v}\in\mathcal{V}$, there exist a unique element $\boldsymbol{v}_1\in\mathcal{V}$, such that $\boldsymbol{v}+\boldsymbol{v}_1=\mathbf{0}$; we denote $\boldsymbol{v}_1$ by $-\boldsymbol{v}$.
2. There is an operation *scalar multiplication* (denoted by $\times$, $*$, $\cdot$, or simply writing the two parts side by side): $\mathbb{R}\times\mathcal{V}\to\mathcal{V}$, such that:
   (a) $\forall\alpha\in\mathbb{R}$ and $\forall\boldsymbol{v}\in\mathcal{V}$: $\alpha\boldsymbol{v}\in\mathcal{V}$.
   (b) $\forall\alpha,\beta\in\mathbb{R}$ and $\forall\boldsymbol{v}\in\mathcal{V}$: $(\alpha\beta)\boldsymbol{v}=\alpha(\beta\boldsymbol{v})$.
   (c) $\forall\alpha\in\mathbb{R}$ and $\forall\boldsymbol{v}_1,\boldsymbol{v}_2\in\mathcal{V}$: $\alpha(\boldsymbol{v}_1+\boldsymbol{v}_2)=\alpha\boldsymbol{v}_1+\alpha\boldsymbol{v}_2$.
   (d) $\forall\alpha_1,\alpha_2\in\mathbb{R}$ and $\forall\boldsymbol{v}\in\mathcal{V}$: $(\alpha_1+\alpha_2)\boldsymbol{v}=\alpha_1\boldsymbol{v}+\alpha_2\boldsymbol{v}$.
   (e) $\forall\boldsymbol{v}\in\mathcal{V}\Rightarrow 1\cdot\boldsymbol{v}=\boldsymbol{v}, 0\cdot\boldsymbol{v}=\mathbf{0}$.

A finite subset $V=\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n\}$ of $\mathcal{V}$ is *linearly independent* if:

$$\sum_{i=1}^{n}\alpha_i\boldsymbol{v}_i=\mathbf{0},\quad\Rightarrow\quad \alpha_i=0, i=1,\cdots n\,. \tag{1.1}$$

We will frequently use the Einstein notation convention where repeated index denotes summation. For example, the left hand side of (1.1) is equivalently written as $\alpha_i\boldsymbol{v}_i$.

A linearly independent set V is called a **basis** if any $\boldsymbol{v}\in\mathcal{V}$ can be expressed as:

$$\boldsymbol{v}=\alpha_i\boldsymbol{v}_i \tag{1.2}$$

for some $\alpha_i\in\mathbb{R}$. This representation is unique, and we call $(\alpha_1,\cdots,\alpha_n)$ the *coordinates* of $\boldsymbol{v}$ under the basis V. The number n is called the *dimension* of $\mathcal{V}$ (denoted by $\dim(\mathcal{V})$) – it is independent of the choice of the basis $\boldsymbol{V}$, hence the dimension is a property of the vector space $\mathcal{V}$ itself. In general,

the dimension of a vector space $\mathcal{V}$ can be infinite; in this course we only consider finite dimensional spaces.

Let $\mathcal{W}$ be a non-empty subset of $\mathcal{V}$, such that $\mathcal{W}$ is a linear vector space under the same addition and scalar multiplication operators, then $\mathcal{W}$ is called a *linear subspace* or simply *subspace* of $\mathcal{V}$. A trivial subspace of any $\mathcal{V}$ is $\mathcal{V}$ itself. Let $\mathcal{W}$ has the dimension m, then $1 \le m \le n$ and we call $n-m$ the *codimension* of $\mathcal{W}$, denoted by $\mathrm{codim}(\mathcal{W})$. We have:

$$\mathrm{codim}(\mathcal{W}) = \dim(\mathcal{V}) - \dim(\mathcal{W})\,. \tag{1.3}$$

Any basis of $\mathcal{W}$ can be extended to a basis of $\mathcal{V}$.

In this course, we mostly deal with the Euclidean spaces[1] $\mathbb{R}^n$ and their subspaces, where the addition and scalar multiplication coincide with common practices. In particular, we call the vectors $\boldsymbol{e}_1 = (1,0,\cdots,0)$, $\boldsymbol{e}_2 = (0,1,\cdots,0)$, $\cdots$, and $\boldsymbol{e}_n = (0,\cdots,0,1)$ the *standard basis* of $\mathbb{R}^n$. Under the standard basis, an element $\boldsymbol{v} = (v_1,v_2,\cdots,v_n) \in \mathbb{R}^n$ has the coordinates $(v_1,\cdots,v_n)$. In this case, we usually identify the coordinate of a vector in $\mathbb{R}^n$ with the vector itself.

Remark 1. *In fact, for any n-dimensional vector space $\mathcal{V}$, once a basis $\{\boldsymbol{e}_1^v,\cdots,\boldsymbol{e}_n^v\}$ is fixed we can identify any element $\boldsymbol{v} = v_i\boldsymbol{e}_i^v$ of $\mathcal{V}$ with its coordinate $(v_1,\cdots,v_n)$ under this basis. This coordinate is clearly an element of $\mathbb{R}^n$; when no confusion will be caused, we will denote this coordinate vector in $\mathbb{R}^n$ also by $\boldsymbol{v}$.*

2 Matrices

Let $\mathcal{V}$ be an n-dimensional vector space and $\mathcal{W}$ be an m-dimensional vector space. A **linear map** or a **linear transformation** L: $\mathcal{V} \to \mathcal{W}$ is defined by:

1. $\forall \alpha \in \mathbb{R}$ and $\forall \boldsymbol{v} \in \mathcal{V}$: $L(\alpha\boldsymbol{v}) = \alpha L(\boldsymbol{v})$.
2. $\forall \boldsymbol{v}_1,\boldsymbol{v}_2 \in \mathcal{V} \Rightarrow L(\boldsymbol{v}_1+\boldsymbol{v}_2) = L(\boldsymbol{v}_1)+L(\boldsymbol{v}_2)$.

By definition, any linear map is identified by its operation on a basis of $\mathcal{V}$. Let $V = \{\boldsymbol{e}_1^v,\cdots,\boldsymbol{e}_n^v\}$ and $W = \{\boldsymbol{e}_1^w,\cdots,\boldsymbol{e}_m^w\}$ be two bases of $\mathcal{V}$ and $\mathcal{W}$, respectively, and we suppose:

$$L(\boldsymbol{e}_i^v) = a_{ij}\boldsymbol{e}_j^w, \quad \forall i = 1,\cdots,n\,, \tag{2.1}$$

where a_{ij} are real numbers. Then for any $\boldsymbol{v} = v_i\boldsymbol{e}_i^v \in \mathcal{V}$ we have:

$$L(\boldsymbol{v}) = v_i L(\boldsymbol{e}_i^v) = v_i a_{ij}\boldsymbol{e}_j^w\,. \tag{2.2}$$

Hence if we denote $\boldsymbol{w} = L(\boldsymbol{v}) = w_i\boldsymbol{e}_i^w$, we have:

$$w_j = a_{ij}v_i \quad \forall j = 1,\cdots,m\,. \tag{2.3}$$

Per Remark 1, the preceding relation can be written as:

$$\boldsymbol{w} = \boldsymbol{A}^t\boldsymbol{v}\,, \tag{2.4}$$

[1] This is to be made precise later.

where $\boldsymbol{w} \in \mathbb{R}^m$ and $\boldsymbol{v} \in \mathbb{R}^n$ are the coordinates identified with the two vectors, respectively; the matrix $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ is defined by $\boldsymbol{A} = [a_{ij}]_{n\times m}$, and ^{t} denotes the matrix transpose. In fact, we have a one-to-one correspondence of all the $n \times m$ real matrices and linear transformations between the two vector spaces (with their bases given).

It is not difficult to see that $\mathbb{R}^{n\times m}$, or the set of all linear transforms from $\mathcal{V}$ to $\mathcal{W}$ (denoted by $\mathcal{L}(\mathcal{V},\mathcal{W})$), is also a linear vector space. For example, let $\alpha_1,\alpha_2 \in \mathbb{R}$ and $L_1,L_2 \in \mathcal{L}(\mathcal{V},\mathcal{W})$ be arbitrary, then $\alpha_1 L_1 + \alpha_2 L_2$ also belongs to $\mathcal{L}(\mathcal{V},\mathcal{W})$ and it is defined as:

$$(\alpha_1 L_1 + \alpha_2 L_2)(\boldsymbol{v}) = \alpha_1 L_1(\boldsymbol{v}) + \alpha_2 L_2(\boldsymbol{v}), \quad \forall \boldsymbol{v} \in \mathcal{V}\,. \tag{2.5}$$

Its dimension is mn.

The *range* of the operator $L \in \mathcal{L}(\mathcal{V},\mathcal{W})$ is a subset of $\mathcal{W}$, which is consist of elements that can be expressed as $L(\boldsymbol{v})$ for some $\boldsymbol{v} \in \mathcal{V}$. The *kernel* of the operator is a subset of $\mathcal{V}$, which contains the elements of $\boldsymbol{v} \in \mathcal{V}$ such that $L(\boldsymbol{v}) = \boldsymbol{0} \in \mathcal{W}$. We denote these two sets by $\mathrm{Ra}(L) \subseteq \mathcal{W}$ and $\mathrm{Ker}(L) \subseteq \mathcal{V}$, respectively. Clearly $\mathrm{Ra}(L)$ is a subspace of $\mathcal{W}$ and it corresponds to the column space of A^t; $\mathrm{Ker}(L)$ is a subspace of $\mathcal{V}$ and it corresponds to the null space of A^T (denoted similarly as $\mathrm{Ker}(A^t)$). Furthermore, we have:

$$\dim \mathrm{Ra}(L) = \mathrm{rank} A^t = n - \dim \mathrm{Ker}(A^t) = \dim \mathcal{V} - \dim \mathrm{Ker}(L)\,. \tag{2.6}$$

We can associate the "matrix multiplication" with composition of linear transforms. Let $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{Z}$ be vector spaces with dimension n, m, and k, respectively; and we fix a basis for each of them. Suppose $L_1 \in \mathcal{L}(\mathcal{V},\mathcal{W})$ and $L_2 \in \mathcal{L}(\mathcal{W},\mathcal{Z})$, and under the chosen bases the two operators correspond to the matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times k}$; then it is not difficult to verify that $L_2 \circ L_1 \in \mathcal{L}(\mathcal{V},\mathcal{Z})$ and it correspond to the matrix $AB \in \mathbb{R}^{n\times k}$.

In the special case when $\mathcal{W} = \mathcal{Z} = \mathcal{V}$ and we choose the same basis for all of them, if $L \in \mathcal{L}(\mathcal{V},\mathcal{V})$ has the full range: $\mathrm{Ra}(L) = \mathcal{V}$, the corresponding matrix A is invertible. Furthermore, the inverse of L (denoted by L^{-1}) exists and it corresponds to A^{-1}.

Finally, we consider the structure of a linear map $L \in \mathcal{L}(\mathcal{V},\mathcal{V})$; especially we consider L under different bases. Suppose L corresponds to A under the basis $\{\boldsymbol{e}_i^v : 1 \le i \le n\}$ and to $\tilde{A}$ under another basis $\{\tilde{\boldsymbol{e}}_i^v : 1 \le i \le n\}$. Note that $L = I \circ L \circ I$, where I is the identity map from $\mathcal{V}$ to itself. Applying the preceding discussion to the map compositions $I \circ L \circ I : \mathcal{V} \to \mathcal{V} \to \mathcal{V} \to \mathcal{V}$, but equipping the first and the fourth $\mathcal{V}$ with the basis $\{\tilde{\boldsymbol{e}}_i^v\}$ whereas the middle two $\mathcal{V}$ with the basis $\{\boldsymbol{e}_i^v\}$, we obtain:

$$\tilde{A}^t = T^{-t} A^t T^t\,, \tag{2.7}$$

where T is the matrix corresponds to I in the last map of the chain, or equivalently $T = [t_{ij}]_{n\times n}$ is given by:

$$\boldsymbol{e}_i^v = t_{ij} \tilde{\boldsymbol{e}}_j^v, \quad \forall 1 \le i \le n\,. \tag{2.8}$$

As a consequence, if the matrix A is diagonalizable we can choose T such that its rows are the eigenvectors of A^t, then the resulting $\tilde{A}$ is diagonal, meaning that the action of the linear map L is scaling each coordinate in the basis $\{\tilde{\boldsymbol{e}}_i^v\}$ by a fixed quantity that equals a diagonal element of $\tilde{A}$.

3 Bilinear forms and special matrices

A **bilinear form** is a map B: $\mathcal{V} \times \mathcal{V} \to \mathbb{R}$, where for all $\alpha_1,\alpha_2 \in \mathbb{R}$ and all $\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3 \in \mathcal{V}$ we have:

1. $B(\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2, \boldsymbol{v}_3) = \alpha_1 B(\boldsymbol{v}_1, \boldsymbol{v}_3) + \alpha_2 B(\boldsymbol{v}_2, \boldsymbol{v}_3)$.

2. $B(\boldsymbol{v}_3, \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2) = \alpha_1 B(\boldsymbol{v}_3, \boldsymbol{v}_1) + \alpha_2 B(\boldsymbol{v}_3, \boldsymbol{v}_2)$.

If in addition $B(\boldsymbol{v}_1, \boldsymbol{v}_2) = B(\boldsymbol{v}_2, \boldsymbol{v}_1)$ for all $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathcal{V}$, we call B a **symmetric bilinear form**. In this case, one of the two preceding requirements is redundant.

Due to the linearity, a bilinear form is also determined by its action on vectors in a basis. Let $B(\boldsymbol{e}_i^v, \boldsymbol{e}_j^v) = g_{ij}$, then we have for $\boldsymbol{v} = v_i \boldsymbol{e}_i^v$ and $\boldsymbol{w} = w_i \boldsymbol{e}_i^v$:

$$B(\boldsymbol{v}, \boldsymbol{w}) = v_i w_j g_{ij} = \boldsymbol{v}^t G \boldsymbol{w} \,. \tag{3.1}$$

Hence any bilinear form is also related to a matrix $G = [g_{ij}]_{n \times n}$; and if B is symmetric, this matrix is also **symmetric**: $g_{ij} = g_{ji}, \forall 1 \leq i, j \leq n$, or $G = G^t$.

Of special interest is the symmetric bilinear form associated to **positive-definite matrices**:

$$\boldsymbol{v}^t \boldsymbol{G} \boldsymbol{v} \geq 0, \quad \forall \boldsymbol{v} \in \mathbb{R}^n \,, \tag{3.2}$$

and

$$\boldsymbol{v}^t \boldsymbol{G} \boldsymbol{v} = 0 \text{ if and only if } \boldsymbol{v} = \boldsymbol{0} \,. \tag{3.3}$$

In this case, the bilinear form is also called positive definite:

1. $B(\boldsymbol{v}, \boldsymbol{v}) \geq 0, \quad \forall \boldsymbol{v} \in \mathcal{V}$.

2. $B(\boldsymbol{v}, \boldsymbol{v}) = 0$ if and only if $\boldsymbol{v} = 0$.

Such symmetric positive-definite bilinear forms are called *inner product* on the space $\mathcal{V}$. For the latter, we denote the bilinear form by angled brackets:

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle_G = B(\boldsymbol{v}, \boldsymbol{w}), \quad \forall \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V} \,, \tag{3.4}$$

where $B(\cdot, \cdot)$ is defined by the symmetric positive-definite matrix G.

There are at least two things that we can talk about once an inner product is defined.

First, the above properties allow us to define a **norm** on elements of $\mathcal{V}$:

$$||\boldsymbol{v}||_B = ||\boldsymbol{v}||_G \stackrel{\text{def}}{=\!=} \sqrt{B(\boldsymbol{v}, \boldsymbol{v})} \,, \tag{3.5}$$

which is essentially a measure of the "distance" of the element $\boldsymbol{v}$ to the zero vector $\boldsymbol{0}$. Due to the linearity, we can also define the *distance* between any two elements of $\mathcal{V}$ by:

$$d(\boldsymbol{v}, \boldsymbol{w}) \stackrel{\text{def}}{=\!=} ||\boldsymbol{v} - \boldsymbol{w}||_G. \tag{3.6}$$

For a general positive-definite matrix G, we call this $||\cdot||_G$ the *Riemannian metric* of the vector space $\mathcal{V}$.

Now consider the case when $\mathcal{V} = \mathbb{R}^n$ and the standard basis is chosen. A special metric is given by setting $G = I$, the identity matrix of $\mathbb{R}^{n \times n}$, and we have:

$$||\boldsymbol{v}||_I = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} \,, \tag{3.7}$$

the familiar *Euclidean distance* on the n-dimensional real spaces. In this course, we will mostly consider Euclidean spaces and omit the I in the subscript; similarly the corresponding inner product is denoted by $\langle\cdot,\cdot\rangle$.

Second, we may extend the concept of "orthogonality" to general inner-product spaces. In particular, let $\langle\cdot,\cdot\rangle_G$ be the inner product, then $\boldsymbol{v}$ and $\boldsymbol{w}$ are orthogonal to each other under this inner product if:

$$\langle \boldsymbol{v},\boldsymbol{w}\rangle_G = 0, \quad \text{or equivalently} \quad \boldsymbol{v}^t G\boldsymbol{w} = 0\,. \tag{3.8}$$

In the case of Euclidean spaces $G = I$, and the inner product coincide with the dot product and (3.8) reduces to the common orthogonality condition.

4 Vector and matrix norms

We've seen that matrices have tight relation to linear spaces and linear transforms between these spaces. We can interpret most topics of this course in this geometrical context. For example, solving the **linear system** $A\boldsymbol{x} = \boldsymbol{b}$ can be interpreted as finding a vector $\boldsymbol{x}$ in a suitably defined vector space $\mathcal{V}$ that is mapped to $\boldsymbol{b}$ in another vector space $\mathcal{W}$ by a linear map that corresponds to A^t. Also, the process of **orthogonal reduction** is the same as the following: Given a symmetric bilinear form B that corresponds to the positive definite matrix G under one particular basis, constructing another basis such that B corresponds to the identity matrix I under this new basis. The eigenvalue problem or the generalized eigenvalue problem have more in-depth interpretation in this geometrical setting, and we will get back to this subject later in this course.

Meanwhile, we focus on how to measure certain quantities in the problem of solving a linear system $A\boldsymbol{x} = \boldsymbol{b}$ for a square matrix $A \in \mathbb{R}^{n\times n}$ (or a linear transform that maps a vector space into itself). We first consider the *well-posedness* of this problem:

1. For any $\boldsymbol{b} \in \mathbb{R}^n$ there is a vector $\boldsymbol{x} \in \mathcal{R}^n$ such that $A\boldsymbol{x} = \boldsymbol{b}$.

2. This solution is unique.

Clearly, a sufficient and necessary condition for the problem to be well posed is that A is nonsingular. If A is singular, $A\boldsymbol{x} = \boldsymbol{b}$ only possesses a solution when $\boldsymbol{b} \in \mathrm{Ra}(A)$; and in the latter case the solution is not unique. We shall talk about the technique to solve a linear system with singular matrix A much later in the course when we get to the point of singular value decomposition and generalized inverse; at the moment let us restrict ourself to nonsingular matrices. The general form, however, proves to be useful in several applications including the conjugate gradient method.

Let A be nonsingular, the first thing we need to realize in seeking numerical solutions to $A\boldsymbol{x} = \boldsymbol{b}$ is that the data is often **not** exact. This can either due to measure imprecision or to roundoff errors; and the errors may reside either in $\boldsymbol{b}$ or in A. Thus the actual problem that we solve is:

$$(A + \delta A)(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{b} + \delta\boldsymbol{b}\,, \tag{4.1}$$

where δA and $\delta\boldsymbol{b}$ are small matrix and vector that represent the errors in data, and $\delta\boldsymbol{x}$ is a vector that represents the error in the numerical solution. A fundamental question that relates to almost any numerical method is whether this method is *stable* and how we can estimate the errors in the solution given the errors in data.

A useful tool for this purpose in the context of linear systems is the vector and matrix norms. A norm, as mentioned before, is a measure of the difference between any vector and the zero vector; and the ones induced by bilinear forms are special cases. Formally, we define a **norm** on a vector space $\mathcal{V}$ as a non-negative function $||\cdot|| : \mathcal{V} \to \overline{\mathbb{R}^+}$:

1. $||\boldsymbol{v}|| \geq 0, \quad \forall \boldsymbol{v} \in \mathcal{V}$ and $||\boldsymbol{v}|| = 0$ if and only if $\boldsymbol{v} = 0$.
2. For all $\alpha \in \mathbb{R}$ and $\boldsymbol{v} \in \mathcal{V}$, $||\alpha \boldsymbol{v}|| = |\alpha| ||\boldsymbol{v}||$.
3. The triangle inequality: $\forall \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$, $||\boldsymbol{v} + \boldsymbol{w}|| \leq ||\boldsymbol{v}|| + ||\boldsymbol{w}||$.

Other than the norms that are induced by bilinear forms, other commonly used norms on $\mathcal{R}^n$ are the L^p-norms (or simply p-norms), where $1 \leq p \leq \infty$. In particular, for $1 \leq p < \infty$, we define:

$$||\boldsymbol{v}||_p = (|v_1|^p + \cdots + |v_n|^p)^{1/p}, \tag{4.2}$$

and for $p = \infty$ we define:

$$||\boldsymbol{v}||_\infty = \max_{1 \leq i \leq n} |v_i|. \tag{4.3}$$

Mathematically these norms do not differ from each other very much, in the sense that any two norms of $\mathcal{R}^n$ are equivalent to each other; that is, two norms $||\cdot||_1$ and $||\cdot||_2$[2] are equivalent if and only if there exist positive constants $0 < c_1 < c_2$ such that for all $\boldsymbol{v} \in \mathcal{V}$:

$$c_1 ||\boldsymbol{v}||_1 \leq ||\boldsymbol{v}||_2 \leq c_2 ||\boldsymbol{v}||_1. \tag{4.4}$$

Different norms, however, have significant different implications in numerical analysis. For example, minimization in L^2 norm leads to the least squares problem that are much easier to solve than minimization in other norms; whereas minimization in L^1 norm has the special property that small entries tend to be driven to zero, hence it has important application in data compression. Another commonly used norm is the L^∞ norm, for example trying to reduce $||\delta \boldsymbol{x}||_\infty$ essentially means trying to minimize the errors in each component of $\boldsymbol{x}$ uniformly.

If we regard any matrix A as a linear map between $\mathbb{R}^n$ and itself (in what follows, we exchange the role of A and A^t and simply denote this map by $\boldsymbol{x} \mapsto A\boldsymbol{x}$), any norm on $\mathbb{R}^n$ induces a norm on $\mathbb{R}^{n \times n}$ as follows:

$$||A|| \stackrel{\text{def}}{=\!=} \max_{\boldsymbol{x}: ||\boldsymbol{x}|| \neq 0} \frac{||A\boldsymbol{x}||}{||\boldsymbol{x}||}. \tag{4.5}$$

If the vector norm is denoted by $||\boldsymbol{x}||_p$, we also denote the induced matrix-norm by $||A||_p$.

Remark 2. *We have omitted the important questions of showing (4.2), (4.3), or (4.5) indeed defines a norm, especially the triangle inequalities.*

We have the following results of several induced matrix norms: Let $A \in \mathbb{R}^{n \times n}$, then:

$$||A||_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}| \tag{4.6}$$

$$||A||_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}| \tag{4.7}$$

$$||A||_2 = \sigma_{\max}, \tag{4.8}$$

[2]Here we mean any two generic norms, not to be confused with the L^p-norm defined before.

where $\sigma_{\max}$ is the largest singular value of A, or equivalently $\sigma_{\max}^2$ is the largest eigenvalue of A^tA.

Proof. We will show (4.6) and (4.8), the other one is left as exercise. For (4.6), according to Exercise 2 we may assume in the following $||x||_1 = 1$ or equivalently $\sum_{i=1}^n |x_i| = 1$. Then:

$$\begin{aligned}
||A||_1 &= \max_{||\boldsymbol{x}||_1=1} \sum_{i=1}^n |a_{ij}x_j| \le \max_{||\boldsymbol{x}||_1=1} \sum_{i=1}^n \sum_{j=1}^n |a_{ij}||x_j| \\
&= \max_{||\boldsymbol{x}||=1} \sum_{j=1}^n (\sum_{i=1}^n |a_{ij}|)|x_j| \le \max_{||\boldsymbol{x}||=1} \sum_{j=1}^n (\max_{1\le k\le n} \sum_{i=1}^n |a_{ik}|)|x_j| \\
&= \max_{||\boldsymbol{x}||=1} \max_{1\le k\le n} \sum_{i=1}^n |a_{ik}| \sum_{j=1}^n |x_j| = \max_{||\boldsymbol{x}||=1} \max_{1\le k\le n} \sum_{i=1}^n |a_{ik}| = \max_{1\le k\le n} \sum_{i=1}^n |a_{ij}| \,.
\end{aligned}$$

In addition, let j_0 be the index such that $\sum_{i=1}^n |a_{ij_0}| = \max_{1\le j\le n} \sum_{i=1}^n |a_{ij}|$, then we may select $\boldsymbol{x} = \boldsymbol{e}_{j_0}$. In this case $||\boldsymbol{x}||_1 = 1$, and:

$$A\boldsymbol{x} = A\boldsymbol{e}_{j_0} = (a_{1j_0}, a_{2j_0}, \cdots, a_{nj_0})^t \Rightarrow ||A\boldsymbol{x}||_1 = \sum_{i=1}^n |a_{ij_0}| = \max_{1\le j\le n} \sum_{i=1}^n |a_{ij}| \,,$$

which completes the proof of (4.6).

As for (4.8), we need to use a few results from linear algebra. In particular it is well known that any symmetric positive-semidefinite matrix can be diagonalized by a set of orthonormal vectors. Applying this result to A^tA we obtain $A^tA = Q^{-1}DQ = Q^tDQ$, where Q is orthogonal and $D = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_n^2)$ is diagonal. Let $||\boldsymbol{x}||_2 = 1$ be arbitrary, then using the fact that orthogonal transformation preserves the L^2-norm, we have:

$$||A\boldsymbol{x}||_2^2 = \boldsymbol{x}^tA^tA\boldsymbol{x} = \boldsymbol{x}^tQ^tDQ\boldsymbol{x} = \boldsymbol{y}^tD\boldsymbol{y} = \sum_{i=1}^n \sigma_i^2 y_i^2 \,,$$

here $\boldsymbol{y} = Q\boldsymbol{x}$ and thusly $||\boldsymbol{y}||_2 = 1$. It follows that:

$$||A\boldsymbol{x}||_2^2 \le \max_{1\le j\le n} \sigma_j^2 \sum_{i=1} y_i^2 = \sigma_{\max}^2 \,.$$

Since $\boldsymbol{x}$ is arbitrary, we obtain $||A||_2 \le \sigma_{\max}$.

For the equality, we may suppose $\sigma_{j_0}^2 = \max_{1\le j\le n} \sigma_j^2$ and choose $\boldsymbol{x}$ as the j_0'th column of Q, which completes the proof. □

Finally, we derive an *a priori* estimate of the error for a simpler version of (4.1), where $\delta A = 0$. In particular, we are interested in the relative errors $||\delta\boldsymbol{x}||/||\boldsymbol{x}||$ and $||\delta\boldsymbol{b}||/||\boldsymbol{b}||$, mainly due to the mechanism of floating point numbers. Suppose $A(\boldsymbol{x}+\delta\boldsymbol{x}) = \boldsymbol{b}+\delta\boldsymbol{b}$ and $A\boldsymbol{x} = \boldsymbol{b}$, on the one hand we derive:

$$\delta\boldsymbol{x} = A^{-1}\delta\boldsymbol{b} \quad \Rightarrow \quad ||\delta\boldsymbol{x}|| \le \left|\left|A^{-1}\right|\right| ||\delta\boldsymbol{b}|| \,.$$

On the other hand:

$$A\boldsymbol{x} = \boldsymbol{b} \quad \Rightarrow \quad ||\boldsymbol{b}|| \le ||A||\,||\boldsymbol{x}|| \,.$$

Combining the two inequalities and ignoring the trivial case $\boldsymbol{b}=\mathbf{0}$, we obtain:

$$\frac{||\delta \boldsymbol{x}||}{||\boldsymbol{x}||} \leq ||A|| \left|\left|A^{-1}\right|\right| \frac{||\delta \boldsymbol{b}||}{||b||} \,. \tag{4.9}$$

The number $||A|| \left|\left|A^{-1}\right|\right|$ is called the **condition number** of the matrix A, denoted by $\kappa(A)$. The estimate (4.9) states that the relative error in the solution cannot grow from the relative error in the data by a factor of $\kappa(A)$.

Exercises

Exercise 1. *Prove that the L^1-norm and the L^∞-norm on $\mathbb{R}^n$ are equivalent. That is, find the two constants c_1 and c_2 such that (4.4) holds for these two norms. You may see that these two constants depend on the dimension n, which provides another reason why despite their mathematical equivalence, the two norms are treated extremely differently for practical applications when n is very large.*

Exercise 2. *Show that (4.5) is equivalent to:*

$$||A|| = \max_{\boldsymbol{x}:||\boldsymbol{x}||=1} ||A\boldsymbol{x}|| \,. \tag{4.10}$$

Exercise 3. *Let A be diagonalizable with real eigenvalues. The spectral radius of a matrix A, denoted by $\rho(A)$, is the largest absolute value of eigenvalues of A:*

$$\rho(A) = \max\{|\lambda| : \text{ there exists nonzero } \boldsymbol{v} \in \mathbb{R}^n \text{ such that } A\boldsymbol{v} = \lambda \boldsymbol{v}\} \,. \tag{4.11}$$

Show that any induced matrix norm (4.5) satisfies:

$$||A|| \geq \rho(A) \,. \tag{4.12}$$

Exercise 4. *Prove (4.7).*

Exercise 5. *We would like to derive an estimate of the growth in relative errors for the equation (4.1). Particularly, we suppose A is non-singular and δA is small enough such that $A+\delta A$ is also nonsingular and the following inequality is true:*

$$\frac{||\delta A||}{||A||} < \frac{1}{\kappa(A)} \,.$$

Our estimate proceeds as below. Taking the difference between (4.1) and $A\boldsymbol{x}=\boldsymbol{b}$ gives:

$$(A+\delta A)\delta \boldsymbol{x} = \delta \boldsymbol{b} - \delta A \boldsymbol{x} \,, \tag{4.13}$$

and we begin with estimating $\left|\left|(A+\delta A)^{-1}\right|\right|$. For this purpose, show that:

$$\left|\left|(A+\delta A)^{-1}\right|\right| \leq \frac{\kappa(A)}{||A||} \cdot \frac{1}{1-\kappa(A)(||\delta A||/||A||)} \,.$$

Next, using this estimate and (4.13) to derive:

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