# Basic Ideas from Linear Algebra and Vector Norms

P. Sam Johnson

# Introduction

The analysis and derivation of algorithms in the matrix computation area requires a facility with certain aspects of linear algebra. Some of the basics such as Independence, Subspace, Basis, and Dimension are revieweds.

We next discuss a notion called, norm. Norms serve the same purpose on vector spaces that absolute value does on the real line: they furnish a measure of distance. More precisely, $\mathbb{R}^n$ together with a norm on $\mathbb{R}^n$ defines a metric space.

Therefore, we have the familiar notions of neighborhood, open sets, convergence, and continuity when working with vectors and vector-valued functions.

# Independence, Subspace, Basis, and Dimension

A set of vectors $\{a_1, \ldots, a_n\}$ in $\mathbb{R}^m$ is linearly independent if $\sum_{j=1}^{n} \alpha_j a_j = 0$ implies $\alpha(1:n) = 0$. Otherwise, a nontrivial combination of the $a_i$ is zero and $\{a_1, \ldots, a_n\}$ is said to be linearly dependent.

A subspace of $\mathbb{R}^m$ is a subset that is also a vector space. Given a collection of vectors $a_1, \ldots, a_n \in \mathbb{R}^m$, the set of all linear combinations of these vectors is a subspace referred to as the span of $\{a_1, \ldots, a_n\}$:

$$span\{a_1, \ldots, a_n\} = \left\{ \sum_{j=1}^{n} \beta_j a_j : \beta_j \in \mathbb{R} \right\}.$$

If $\{a_1, \ldots, a_n\}$ is independent and $b \in span\{a_1, \ldots, a_n\}$, then $b$ is a unique linear combination of the $a_j$.

# Independence, Subspace, Basis, and Dimension (Contd...)

If $S_1, \ldots, S_k$ are subspaces of $\mathbb{R}^m$, then their sum is the subspace defined by $S = \{a_1 + a_2 + \cdots + a_k : a_i \in S_i, i = 1 : k\}$. $S$ is said to be a direct sum if each $v \in S$ has a unique representation $v = a_1 + \cdots + a_k$ with $a_i \in S_i$. In this case we write $S = S_1 \oplus \cdots \oplus S_k$. The intersection of the $S_i$ is also a subspace, $S = S_1 \cap S_2 \cap \ldots \cap S_k$.

The subset $\{a_{i_1}, \ldots, a_{i_k}\}$ is a maximal linearly independent subset of $\{a_1, \ldots, a_n\}$ if it is linearly independent and is not properly contained in any linearly independent subset of $\{a_1, \ldots, a_n\}$. If $\{a_{i_1}, \ldots, a_{i_k}\}$ is maximal, then $span\{a_1, \ldots, a_n\} = span\{a_{i_1}, \ldots, a_{i_k}\}$ and $\{a_{i_1}, \ldots, a_{i_k}\}$ is a basis for $span\{a_1, \ldots, a_n\}$. If $S \subseteq \mathbb{R}^m$ is a subspace, then it is possible to find independent basic vectors $a_1, \ldots, a_k \in S$ such that $S = span\{a_1, \ldots, a_k\}$. All bases for a subspace $S$ have the same number of elements. This number is the dimension and is denoted by $\dim(S)$.

# Range, Null Space, and Rank

There are two important subspaces associated with an $m$-by-$n$ matrix $A$. The range of $A$ is defined by

$$ran(A) = \{y \in \mathbb{R}^m : y = Ax \text{ for some } x \in \mathbb{R}^n\},$$

and the null space of $A$ is defined by

$$null(A) = \{x \in \mathbb{R}^n : Ax = 0\}.$$

If $A = [a_1, \cdots, a_n]$ is a column partitioning, then

$$ran(A) = span\{a_1, \ldots, a_n\}$$

The rank of a matrix $A$ is defined by

$$rank(A) = \dim(ran(A)).$$

It can be shown that $rank(A) = rank(A^T)$. We say that $A \in \mathbb{R}^{m \times n}$ is rank deficient if $rank(A) < \min\{m, n\}$. If $A \in \mathbb{R}^{m \times n}$, then

$$\dim(null(A)) + rank(A) = n.$$

# Matrix Inverse

The $n$-by-$n$ identity matrix $I_n$ is defined by the column partitioning

$$I_n = [e_1, \ldots, e_n]$$

where $e_k$ is the $k$th "canonical" vector:

$$e_k = (\underbrace{0, \ldots, 0}_{k-1}, 1, \underbrace{0, \ldots, 0}_{n-k})^T.$$

The canonical vectors arise frequently in matrix analysis and if their dimension is ever ambiguous, we use superscripts, i.e., $e_k^{(n)} \in \mathbb{R}^n$.

If $A$ and $X$ are in $\mathbb{R}^{n \times n}$ and satisfy $AX = I$, then $X$ is the inverse of $A$ and is denoted by $A^{-1}$. If $A^{-1}$ exists, then $A$ is said to be nonsingular. Otherwise, we say $A$ is singular.

# Matrix Inverse (Contd...)

Several matrix inverse properties have an important role to play in matrix computations. The inverse of a product is the reverse product of the inverses:

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{1}$$

The transpose of the inverse is the inverse of the transpose:

$$(A^{-1})^T = (A^T)^{-1} \equiv A^{-T}. \tag{2}$$

The identity

$$B^{-1} = A^{-1} - B^{-1}(B - A)A^{-1} \tag{3}$$

shows how the inverse changes if the matrix changes.

# Matrix Inverse (Contd...)

The Sherman-Morrison-Woodbury formula gives a convenient expression for the inverse of $(A + UV^T)$ where $A \in \mathbb{R}^{n \times n}$ and $U$ and $V$ are $n$-by-$k$:

$$(A + UV^T)^{-1} = A^{-1} - A^{-1}U(I + V^T A^{-1} U)^{-1} V^T A^{-1}. \tag{4}$$

A rank $k$ correction to a matrix results in a rank $k$ correction of the inverse. In (4) we assume that both $A$ and $(I + V^T A^{-1} U)$ are nonsingular.

Any of these facts can be verified by just showing that the "proposed" inverse does the job. For example, here is how to confirm (3):

$$B(A^{-1} - B^{-1}(B - A)A^{-1}) = BA^{-1} - (B - A)A^{-1} = I.$$

# Determinant

If $A = (a) \in \mathbb{R}^{1\times 1}$, then its determinant is given by $\det(A) = a$. The determinant of $A \in \mathbb{R}^{n\times n}$ is defined in terms of order $n-1$ determinants:

$$\det(A) = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det(A_{1j}).$$

Here, $A_{1j}$ is an $(n-1)$-by-$(n-1)$ matrix obtained by deleting the first row and $j$th column of $A$. Useful properties of the determinant include

$$\begin{aligned}
\det(AB) &= \det(A)\det(B) && A, B \in \mathbb{R}^{n\times n}\\
\det(A^T) &= \det(A) && A \in \mathbb{R}^{n\times n}\\
\det(cA) &= c^n \det(A) && c \in \mathbb{R}, A \in \mathbb{R}^{n\times n}\\
\det(A) \neq 0 &\Leftrightarrow A \text{ is nonsingular} && A \in \mathbb{R}^{n\times n}
\end{aligned}$$

# Differentiation

Suppose $\alpha$ is a scalar and that $A(\alpha)$ is an $m$-by-$n$ matrix with entries $a_{ij}(\alpha)$. If $a_{ij}(\alpha)$ is a differentiable function of $\alpha$ for all $i$ and $j$, then by $\dot{A}(\alpha)$ we mean the matrix

$$\dot{A}(\alpha) = \frac{d}{d\alpha}A(\alpha) = \left(\frac{d}{d\alpha}a_{ij}(\alpha)\right) = (\dot{a}_{ij}(\alpha)).$$

The differentiation of a parameterized matrix turns out to be a handy way to examine the sensitivity of various matrix problems.

## Exercises 1.

1. *Show that if* $A \in \mathbb{R}^{m \times n}$ *has rank* $p$*, then there exists an* $X \in \mathbb{R}^{m \times n}$ *and a* $Y \in \mathbb{R}^{n \times p}$ *such that* $A = XY^T$*, where* $rank(X) = rank(Y) = p$*.*
2. *Suppose* $A(\alpha) \in \mathbb{R}^{m \times r}$ *and* $B(\alpha) \in \mathbb{R}^{r \times n}$ *are matrices whose entries are differentiable functions of the scalar a. Show*

$$\frac{d}{d\alpha}[A(\alpha)B(\alpha)] = \left[\frac{d}{d\alpha}A(\alpha)\right]B(\alpha) + A(\alpha)\left[\frac{d}{d\alpha}B(\alpha)\right].$$

3. *Suppose* $A(\alpha) \in \mathbb{R}^{n \times n}$ *has entries that are differentiable functions of the scalar* $\alpha$*. Assuming* $A(\alpha)$ *is always nonsingular, show*

$$\frac{d}{d\alpha}[A(\alpha)^{-1}] = -A(\alpha)^{-1}\left[\frac{d}{d\alpha}A(\alpha)\right]A(\alpha)^{-1}.$$

# Exercises (Contd...)

## Exercises 2.

4. *Suppose* $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ *and that* $\phi(x) = \frac{1}{2}x^T Ax - x^T b$*. Show that the gradient of* $\phi$ *is given by* $\nabla\phi(x) = \frac{1}{2}(A^T + A)x - b$*.*
5. *Assume that both* $A$ *and* $A + uv^T$ *are nonsingular where* $A \in \mathbb{R}^{n\times n}$ *and* $u, v \in \mathbb{R}$*. Show that if* $x$ *solves* $(A + uv^T)x = b$*, then it also solves a perturbed right hand side problem of the form* $Ax = b + \alpha x$*. Give an expression for* $\alpha$ *in terms of* $A, u,$ *and* $v$*.*

# Definition of Vector Norm

A vector norm on $\mathbb{R}^n$ is a function $f : \mathbb{R}^n \to \mathbb{R}$ that satisfies the following properties:

$$\begin{aligned} & f(x) \geq 0 && x \in \mathbb{R}^n, && (f(x) = 0 \text{ iff } x = 0) \\ & f(x+y) \leq f(x) + f(y) && x, y \in \mathbb{R}^n && \\ & f(\alpha x) = |\alpha| f(x) && \alpha \in \mathbb{R}, x \in \mathbb{R}^n && \end{aligned}$$

We denote such a function with a double bar notation: $f(x) = \|x\|$. Sub-scripts on the double bar are used to distinguish between various norms.

A useful class of vector norms are the $p$-norms defined by

$$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}} \qquad p \geq 1. \tag{5}$$

# Various Norms

Of these the $1, 2$, and $\infty$ norms are the most important:

$$\begin{aligned} \|x\|_1 &= |x_1| + \cdots + |x_n| \\ \|x\|_2 &= (|x_1|^2 + \cdots + |x_n|^2)^{\frac{1}{2}} = (x^T x)^{\frac{1}{2}} \\ \|x\|_\infty &= \max_{1 \le i \le n} |x_i| \end{aligned}$$

A unit vector with respect to the norm $\|\cdot\|$ is a vector $x$ that satisfies $\|x\| = 1$.

# Some Vector Norm Properties

A classic result concerning $p$-norms is the Holder inequality:

$$|x^T y| \leq \|x\|_p \|y\|_q \qquad \frac{1}{p} + \frac{1}{q} = 1. \tag{6}$$

A very important special case of this is the Cauchy-Schwartz inequality:

$$|x^T y| \leq \|x\|_2 \|y\|_2. \tag{7}$$

All norms on $\mathbb{R}^n$ are equivalent, i.e., if $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are norms on $\mathbb{R}^n$, then there exist positive constants, $c_1$ and $c_2$ such that

$$c_1 \|x\|_\alpha \leq \|x\|_\beta \leq c_2 \|x\|_\alpha \tag{8}$$

for all $x \in \mathbb{R}^n$.

# Some Vector Norm Properties

For example, if $x \in \mathbb{R}^n$, then

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2 \tag{9}$$

$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty \tag{10}$$

$$\|x\|_\infty \leq \|x\|_1 \leq n\|x\|_\infty. \tag{11}$$

# Absolute and Relative Error

Suppose $\hat{x} \in \mathbb{R}^n$ is an approximation to $x \in \mathbb{R}^n$. For a given vector norm $\|\cdot\|$ we say that

$$\varepsilon_{aba} = \|\hat{x} - x\|$$

is the absolute error in $\hat{x}$. If $x \neq 0$, then

$$\varepsilon_{rel} = \frac{\|\hat{x} - x\|}{\|x\|}$$

prescribes the relative error in $\hat{x}$. Relative error in the $\infty$-norm can be translated into a statement about the number of correct significant digits in $\hat{x}$. In particular, if

$$\frac{\|\hat{x} - x\|_\infty}{\|x\|_\infty} \approx 10^{-p},$$

then the largest component of $\hat{x}$ has approximately $p$ correct significant digits.

# Absolute and Relative Error (Contd...)

**Example 3.**

*If* $x = (1.234\ .05674)^T$ *and* $\hat{x} = (1.235\ .05128)^T$, *then*

$$\frac{\|\hat{x} - x\|_\infty}{\|x\|_\infty} \approx .0043 \approx 10^{-3}.$$

*Note than* $\hat{x}_1$ *has about three significant digits that are correct while only one significant digit in* $\hat{x}_2$ *is correct.*

We say that a sequence $\{x^{(k)}\}$ of $n$-vectors converges to $x$ if

$$\lim_{k \to \infty} \|x^{(k)} - x\| = 0.$$

Note that because of (8), convergence in the $\alpha$-norm implies convergence in the $\beta$-norm and vice versa.

## Exercises 4.

1. *Show that if* $x \in \mathbb{R}^n$, *then* $\lim_{p\to\infty} \|x\|_p = \|x\|_\infty$.
2. *Prove the Cauchy-Schwartz inequality* (7) *by considering the inequality* $0 \le (ax + by)^T(ax + by)$ *for suitable scalars* $a$ *and* $b$.
3. *Verify that* $\|\cdot\|_1, \|\cdot\|_2$, *and* $\|\cdot\|_\infty$ *are vector norms.*
4. *Verify* (9)-(11). *When is equality achieved in each result?*
5. *Show that in* $\mathbb{R}^n$, $x^{(i)} \to x$ *if and only if* $x_k^{(i)} \to x_k$ *for* $k = 1 : n$.
6. *Show that any vector norm on* $\mathbb{R}^n$ *is uniformly continuous by verifying the inequality* $|\,\|x\| - \|y\|\,| \le \|x - y\|$.
7. *Let* $\|\cdot\|$ *be a vector norm on* $\mathbb{R}^m$ *and assume* $A \in \mathbb{R}^{m\times n}$. *Show that if* $rank(A) = n$, *then* $\|x\|_A = \|Ax\|$ *is a vector norm on* $\mathbb{R}^n$.

## Exercises 5.

8. *Let $x$ and $y$ be in $\mathbb{R}^n$ and define $\psi : \mathbb{R} \to \mathbb{R}$ by $\psi(\alpha) = \|x - \alpha y\|_2$. Show that $\psi$ is minimized when $\alpha = x^T y / y^T y$.*

9. (a) *Verify that $\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}}$ is a vector norm on $\mathbb{C}^n$.*
   (b) *Show that if $x \in \mathbb{C}^m$ then $\|x\|_p \leq c\,(\|Re(x)\|_p + \|Im(x)\|_p)$.*
   (c) *Find a constant $c_n$ such that $c_n(\|Re(x)\|_2 + \|Im(x)\|_2) \leq \|x\|_2$ for all $x \in \mathbb{C}^m$.*

10. *Prove or disprove:*

$$v \in \mathbb{R}^n \Rightarrow \|v\|_1 \|v\|_\infty \leq \frac{1 + \sqrt{n}}{2} \|v\|_2.$$

# Reference Books


1. Gene H. Golub and Charles F. Van Loan, Matrix Computations, 3rd Edition, Hindustan book agency, 2007.
2. A.R. Gourlay and G.A. Watson, Computational methods for matrix eigen problems, John Wiley & Sons, New York, 1973.
3. W.W. Hager, Applied numerical algebra, Prentice-Hall, Englewood Cliffs, N.J, 1988.
4. D.S. Watkins, Fundamentals of matrix computations, John Wiley and sons, N.Y, 1991.
5. C.F. Van Loan, Introduction to scientific computing: A Matrix vector approach using Matlab, Prentice-Hall, Upper Saddle River, N.J, 1997.