Notes on Linear Algebra and Matrix Theory

Network Science

Massimo Franceschet

featuring

Enrico Bozzo

Summary

- Scalar product, Cauchy-Schwarz inequality, angle between two vectors, applications.
- Matrix product, transpose matrix, symmetric matrix, inverse matrix, applications.
- Square and overdetermined linear systems, applications.
- Eigenvectors, eigenvalues, implicit description, applications.
- Nonnegative matrices, Perron-Frobenius theory, graphs.
- Three important books.

Scalar product

The scalar product (a.k.a. dot product or inner product) of two real vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ is not a vector but a scalar defined as

$$x \cdot y = \sum_{k=1}^{n} x_k y_k.$$

From the definition it follows that the scalar product of two real vectors is commutative

$$x \cdot y = \sum_{k=1}^{n} x_k y_k = \sum_{k=1}^{n} y_k x_k = y \cdot x.$$

Euclidean norm and scalar product

By generalizing Pythagorean theorem, the Euclidean norm (or length) of a real vector x is defined as

$$\|x\| = \sqrt{\sum_{k=1}^{n} x_k^2}.$$

Notice that $\|x\| = 0$ if and only if $x_k = 0$ for $k = 1, \ldots, n$.
The Euclidean norm can be expressed by means of a scalar product

$$\|x\| = \sqrt{\sum_{k=1}^{n} x_k x_k} = \sqrt{x \cdot x}.$$

If only the sum of the squares of the entries of x matters then

$$\sum_{k=1}^{n} x_k^2 = \|x\|^2 = x \cdot x.$$

Cauchy-Schwarz inequality

The equality $x \cdot x = \|x\|^2$ can be, in some sense, generalized to two vectors, leading to the Cauchy-Schwarz inequality

$$|x \cdot y| \leq \|x\|\|y\|$$

where equality holds if and only if x and y lay on the same straight line.

Example If $x = (-1, 2)$ and $y = (1, -1)$ then $|x \cdot y| = |-1-2| = 3$ while $\|x\|\|y\| = \sqrt{5}\sqrt{2}$ and $3 < \sqrt{10}$.

If $x = (-1, 2)$ and $y = -2x = (2, -4)$ then $|x \cdot y| = |-2-8| = 10$ and $\|x\|\|y\| = \sqrt{5}\sqrt{20} = \sqrt{100} = 10$.

Angle between two vectors

If $x \neq 0 \neq y$ it is possible to rewrite Cauchy-Schwarz inequality in the form

$$-1 \leq \frac{x \cdot y}{\|x\|\|y\|} \leq 1.$$

The unique $0 \leq \theta \leq \pi$ such that

$$\cos(\theta) = \frac{x \cdot y}{\|x\|\|y\|}$$

is defined measure of the angle between x and y.

Example If $y = x$ then

$$\cos(\theta) = \frac{x \cdot x}{\|x\|\|x\|} = \frac{\|x\|^2}{\|x\|^2} = 1,$$

whence $\theta = 0$. On the contrary, if $y = -x$ then

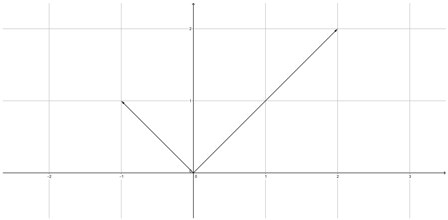
$$\cos(\theta) = \frac{x \cdot (-x)}{\|x\|\|-x\|} = \frac{-\|x\|^2}{\|x\|^2} = -1,$$

whence $\theta = \pi$.

Orthogonal vectors

If $x \cdot y = 0$ then x and y are defined orthogonal. In the case where $x \neq 0 \neq y$ this is equivalent to $\cos(\theta) = 0$, that is $\theta = \pi/2$. The null vector is orthogonal to every vector, even itself, and is the only vector with this property.

Example The two vectors $x = (2, 2)$ and $y = (-1, 1)$ are orthogonal.



Example: Statistics

Let $e = (1, \ldots, 1)$ a vector whose n entries are all equal to one. Notice that the mean of x can be expressed by means of a scalar product

$$m_x = \frac{1}{n}\sum_{k=1}^{n} x_k = \frac{1}{n} x \cdot e.$$

Analogously variance and standard deviation of x become

$$\sigma_x^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - m_x)^2 = \frac{1}{n}\|x - m_x e\|^2, \qquad \sigma_x = \frac{1}{\sqrt{n}}\|x - m_x e\|.$$

The covariance of x and y is

$$\sigma_{xy} = \frac{1}{n}\sum_{k=1}^{n}(x_k - m_x)(y_k - m_y) = \frac{1}{n}(x - m_x e) \cdot (y - m_y e)$$

Example: Statistics (continue)

By Cauchy-Schwarz inequality it follows that

$$-1 \leq \frac{(x - m_x e) \cdot (y - m_y e)}{\|x - m_x e\| \|y - m_y e\|} \leq 1.$$

Notice that

$$\frac{(x - m_x e) \cdot (y - m_y e)}{\|x - m_x e\| \|y - m_y e\|} = \frac{\frac{1}{n}(x - m_x e) \cdot (y - m_y e)}{\frac{1}{\sqrt{n}}\|x - m_x e\| \frac{1}{\sqrt{n}}\|y - m_y e\|} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}.$$

This ratio is the Pearson's correlation coefficient of x and y and is denoted by ρ_{xy}. In summary, by using Cauchy-Schwartz inequality, we proved that Pearson's correlation coefficient can only assume values between -1 and 1

$$-1 \leq \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \leq 1.$$

Example: Information Retrieval

Let $A = (a_{i,j})$ be matrix whose rows are associated with papers and columns with keywords in such a way that $a_{i,j} = 1$ if paper i contains the keyword j and $a_{i,j} = 0$ otherwise. Given a set of keywords we can represent it as a binary vector x and compute the angle between x and each of the rows of A in order to determine the papers more relevant for the given keywords.

As an example, suppose that the first two rows of A are $a = (1, 1, 1, 0, 0)$ and $b = (1, 1, 1, 1, 1)$. If $x = (1, 1, 0, 0, 0)$ then the angle θ between x and a is such that $\cos(\theta) = \sqrt{2/3}$. Analogously the angle η between x and b is such that $\cos(\eta) = \sqrt{3/5}$. The inequality $2/3 > 3/5$ implies $\theta < \eta$.

Matrix product

- Given an $n \times m$ matrix $A = (a_{i,j})$ and an $m \times p$ matrix $B = (b_{i,j})$ their product AB is an $n \times p$ matrix $C = (c_{i,j})$ such that
$$c_{i,j} = \sum_{k=1}^{m} a_{i,k} b_{k,j}.$$
Notice that $c_{i,j}$ is the scalar product of the i-th row of A and the j-th column of B, whose number of elements has to be the same. If $p \neq n$ then BA is not defined.
- If A is $n \times m$ and B is $m \times n$ then both AB and BA are defined: the former is $n \times n$ and the latter is $m \times m$ so their dimensions are different unless $n = m$.
- Even if A and B are both $n \times n$ it turns out that AB and BA are equal only for special choices of A and B, as for example if $A = B$.

Summarizing, matrix product is not commutative.

Transpose. Symmetry.

Given an $n \times m$ matrix A its transpose matrix A^T is the $m \times n$ matrix whose columns are the rows of A, or, equivalently, whose rows are the columns of A. Equivalently if $A = (a_{i,j})$ then $A^T = (a_{j,i})$. An $n \times n$ matrix A is symmetric if $A = A^T$. This means that $a_{i,j} = a_{j,i}$ so that A is specular along the main diagonal.

$$\begin{bmatrix} \ddots & & & & & \\ & \ddots & & & i,j & \\ & & \ddots & & & \\ & & & \ddots & & \\ & j,i & & & \ddots & \\ & & & & & \ddots \end{bmatrix}$$

Example: Bibliometrics

Let A be a matrix whose rows are associated with n papers and columns with m keywords. The two products AA^T and A^TA can both be computed and have interesting meanings.

- The product AA^T is $n \times n$ and its i, j entry is the number of keywords that appear both in paper i and in paper j.
- The product A^TA is $m \times m$ and its i, j entry is the number of papers that contain both the keyword i and the keyword j.

For example if

$$A \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

then

$$AA^T = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix} \quad \text{and} \quad A^TA = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}.$$

Notice that AA^T and A^TA are symmetric.

Square matrices. Inverses.

An $n \times n$ matrix A is a square matrix of order n. If A is square it is possible to compute its determinant $\det(A)$. If $\det(A) = 0$ then A is singular otherwise nonsingular. The inverse of a nonsingular matrix is a square matrix of the same order A^{-1} such that

$$AA^{-1} = A^{-1}A = I$$

where

$$I = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

is the identity matrix. Notice that a matrix and its inverse always commute, even if matrix product does not have this property. Warning: unfortunately square singular matrices and rectangular matrices lack of an inverse; for these matrices in some applications the role of the inverse is played from the generalized inverse (but that's another story).

Inversion formula for 2×2 matrices

It is easy to compute the determinant of a square matrix of order 2

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \Rightarrow \det(A) = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}.$$

If $\det(A) \neq 0$ the inverse of A is given by a simple formula

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} a_{2,2} & -a_{1,2} \\ -a_{2,1} & a_{1,1} \end{bmatrix}.$$

The formula for $n = 3$ is more complex and the inversion of a square matrix of order n, when n is big, is computationally quite heavy and preferably avoided.

Square linear systems and explicit representation of the solution

A linear system has the form

$$Ax = b$$

where A is a known matrix, b is a known vector, x is an unknown vector. If A is square and nonsingular then matrix inversion is fundamental, if not to compute, surely to represent the solution of the system in the explicit form

$$x = A^{-1}b.$$

Overdetermined linear systems

If A has more rows than columns the system is defined overdetermined. These systems are frequently impossible. Why should we bother of systems without solutions?

Example: ranking with Massey's method

In 1997, Kenneth Massey proposed a method for ranking college football teams. The main idea of Massey's method is enclosed in the simple equation

$$r_i - r_j = y_k$$

where r_i and r_j are the ratings of teams i and j and y_k is the absolute margin of victory for game k between teams i and j. Since the number of games is usually higher than the number of teams (for example in the Italian Serie A soccer league 20 teams play 380 games) the method leads to an overdetermined system. As a simple example consider just two teams that play three games

game	score team 1	score team 2
1	1	0
2	2	0
3	1	3

$$\Rightarrow \begin{cases} r_1 - r_2 = 1 \\ r_1 - r_2 = 2 \\ r_1 - r_2 = -2 \end{cases}$$

The first two equations make already the system impossible.

Example: straight line trough more than two points

Let us try to determine m and q in such a way that the straight line of equation $y = mx + q$ passes through $(0,1)$, $(1,0)$ and $(2,1)$. This leads to the overdetermined linear system

$$\begin{cases} q = 1 \\ m + q = 0 \\ 2m + q = 1 \end{cases} \qquad \Leftrightarrow \qquad \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} m \\ q \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

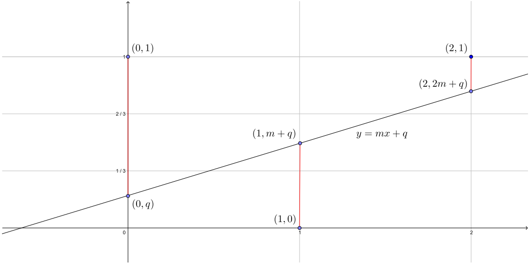
clearly impossible since the three given points do not lie on a straight line.

Least squares

Since an overdetermined system is frequently impossible it makes sense to look for a vector x, defined least square solution of the system, that minimizes $\|Ax - b\|^2$. For our geometric example we see that

$$\|Ax - b\|^2 = (q - 1)^2 + (m + q)^2 + (2m + q - 1)^2$$

is exactly the sum of the squares of the lengths of the vertical segments between the points and the straight line. Minimizing this sum of squares we hope to find a straight line that passes, as much as possible, "near" the tree points.



There are various techniques for solving this minimization problem. The simplest involves the solution of the linear system

$$A^T A x = A^T b,$$

known as system of the normal equations. For our geometric example

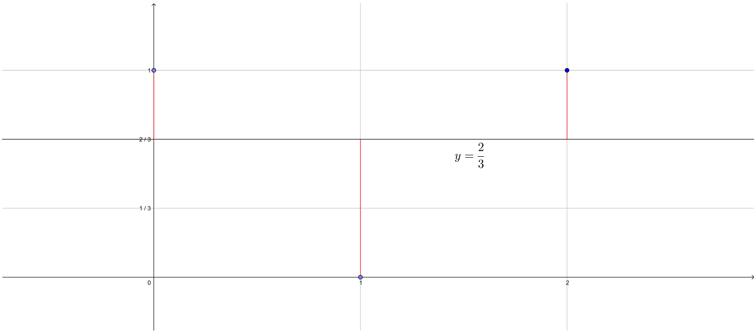
$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} m \\ q \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \quad \Rightarrow \begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} m \\ q \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}.$$

The matrix $A^T A$ is square, and in this example can be inverted, so that

$$\begin{bmatrix} m \\ q \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 3 & -3 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 2/3 \end{bmatrix}.$$

Exercise Find the least square solution of the system obtained in Massey's method example.

Least squares



We obtained the straight line nearest the three points, in the least squares sense. Notice that the line does not passes in the middle but is closer to the two upper points. Actually

$$2\left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2 = \frac{2}{3} < \frac{3}{4} = 3\left(\frac{1}{2}\right)^2.$$

(Right) Eigenvectors and eigenvalues

After this excursion in the world of rectangular matrices let's go back to square ones.
A vector $x \neq 0$ is a *right* eigenvector of a square matrix A if

$$Ax = \lambda x,$$

where λ is a scalar defined as eigenvalue associated with x.
The definition has a very intuitive geometrical meaning: If x is an eigenvector of A, then Ax has the same direction of x. The action of A on x is, in some sense, quite polite. In particular, in the case where $\lambda = 1$, x is left untouched by the action of A.

Centrality and power of a network

If we compare the two systems of equations $Ax = b$ and $Ax = \lambda x$ we recognize that the former is linear while the latter is nonlinear since λ and x, both unknowns, are multiplied together. Solving a nonlinear system is usually much more difficult, without any possibility to obtain an explicit expression for the unknowns. However, the system $Ax = \lambda x$ can be thought as an implicit description of x by means of itself. Concepts such as centrality and power of a network, are frequently approached by setting up equations that express some kind of implicit description: a node is important if it is linked to important nodes, a node is powerful if it is linked to nodes that are not powerful.

Characteristic polynomial

In theory, the eigenvalues of an $n \times n$ matrix A can be computed separately from the eigenvectors since they are the roots of the characteristic polynomial of A

$$p(\lambda) = \det(A - \lambda I).$$

The characteristic polynomial has degree n so that A has n eigenvalues, but some of them can be complex, even if A is real (however all the eigenvalues of a symmetric matrix are real), and some can be repeated.

Multiplicities

- The algebraic multiplicity μ_λ of an eigenvalue λ is the number of times it is repeated as root of the characteristic polynomial.
- The geometric multiplicity γ_λ of an eigenvalue λ is the number of linear independent eigenvectors associated with λ.
- It holds that $1 \leq \gamma_\lambda \leq \mu_\lambda \leq n$.
- If $\mu_\lambda = \gamma_\lambda$ then λ is defined semisimple.
- If $\mu_\lambda = \gamma_\lambda = 1$ then λ is defined simple.

(Left) Eigenvectors and eigenvalues

A vector $x \neq 0$ is a *left* eigenvector of a square matrix A if

$$xA = \lambda x,$$

where λ is a scalar defined as eigenvalue associated with x. Unless $A = A^T$ left eigenvectors are different from right eigenvectors. However, the set of the eigenvalues associated to right and left eigenvectors is exactly the same.

Spectrum. Spectral Radius.

The set of the distinct eigenvalues of A is defined as spectrum of A and denoted with $\sigma(A)$. Clearly $1 \leq |\sigma(A)| \leq n$. It is possible to show that $|\sigma(A)| = n$ if and only if all the eigenvalues are simple.

The spectral radius of A is the nonnegative number

$$\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|.$$

Example

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \Rightarrow p(\lambda) = \det\left(\begin{bmatrix} -\lambda & 1 \\ 1 & -\lambda \end{bmatrix}\right) = \lambda^2 - 1.$$

Hence

$$\sigma(A) = \{1, -1\} \qquad \text{and} \qquad \rho(A) = 1.$$

Notice that A is symmetric and all its eigenvalues are real.
Since $|\sigma(A)| = 2$ the two eigenvalues are simple.

Nonnegative matrices

A matrix is defined nonnegative, respectively, positive, provided all of its entries are nonnegative, respectively positive. These matrices have special spectral properties that depend solely on the pattern of the matrix and are independent from the magnitude of the positive entries.

Perron Theorem

Perron Theorem

If a matrix A is positive then:

1. the spectral radius $r = \rho(A) > 0$ and $r \in \sigma(A)$;
2. if $\lambda \in \sigma(A)$ and $\lambda \neq r$ then $|\lambda| < r$;
3. r is a simple eigenvalue of A;
4. the unique (up to a multiplicative constant) right eigenvector x such that $Ax = rx$ is positive.
5. the unique (up to a multiplicative constant) left eigenvector y such that $yA = ry$ is positive.

In many applications positive matrices are extremely rare. Graph theory is the right tool to obtain a result of much wider applicability.

Matrices and graphs

R. Brualdi and D. Cvetković in the preface of their book A COMBINATORIAL APPROACH TO MATRIX THEORY AND ITS APPLICATIONS state that

> *Generally, elementary (and some advanced) books on matrices ignore or only touch on the combinatorial or graph-theoretical connections with matrices. This is unfortunate in that these connections can be used to shed light on the subject, and to clarify and deepen one's understanding. In fact, a matrix and a (weighted) graph can each be regarded as different models of the same mathematical concept.*

Irreducibility is a good example of what Brualdi and Cvetković mean. It has two equivalent definitions: the graph theoretic one is more intuitive. However, the matrix theoretic definition suggests explicitly the concept of reduction in two smaller parts.

Irreducible matrices

Irreducible matrix-graph theoretic definition

A square matrix A is irreducible provided that the graph whose A is the adjacency matrix is strongly connected, that is, there is a directed path between any ordered couple of arbitrarily chosen vertexes. Otherwise it is reducible.

Irreducible matrix-matrix theoretic definition

A square matrix A is irreducible if does not exist a permutation matrix P such that

$$PAP^T = \begin{bmatrix} X & Y \\ O & Z \end{bmatrix}$$

where X and Z are square. Otherwise it is reducible.

Exercise

Which of the following matrices are irreducible?

$$(a) \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \qquad (b) \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$(c) \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad (d) \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Perron-Frobenius Theorem

Perron-Frobenius Theorem

If a nonnegative matrix A is irreducible then:

1. the spectral radius $r = \rho(A) > 0$ and $r \in \sigma(A)$;
2. r is a simple eigenvalue of A;
3. the unique (up to a multiplicative constant) right eigenvector x such that $Ax = rx$ is positive.
4. the unique (up to a multiplicative constant) left eigenvector y such that $yA = ry$ is positive.

Be careful!

The hypothesis of Perron-Frobenius Theorem is weaker but also the thesis is weaker since for an irreducible matrix A it is non excluded that $|\lambda| = \rho(A)$ for $\lambda \in \sigma(A)$ and $\lambda \neq \rho(A)$.

Example The nonnegative matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is easily recognized to be irreducible. Recall that $\sigma(A) = \{1, -1\}$ and $\rho(A) = 1$. Clearly $|-1| = 1$.

Power method in network science

For the convergence of power method, a numerical tool of very large use in network science, it is important that the spectral radius is strictly greater than all the moduli of the other eigenvalues.

Primitive matrices

It can be shown that nonnegative irreducible matrices having a positive power, defined primitive matrices, are such that their spectral radius is strictly greater than all the moduli of the other eigenvalues.

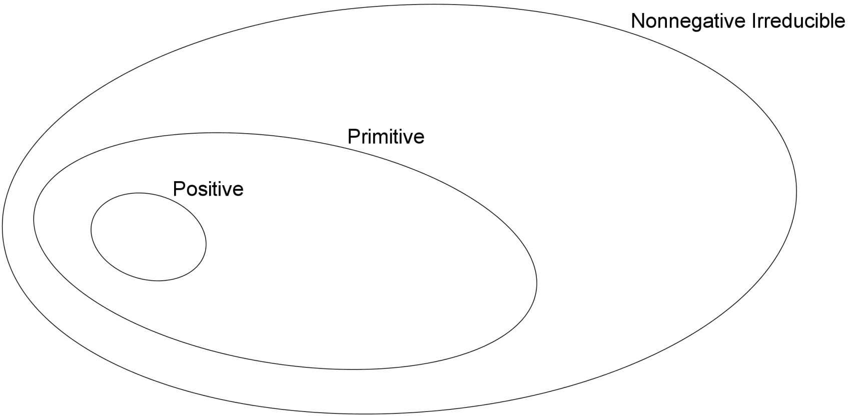
Example The matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is irreducible but not primitive (why?). The matrix $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ is primitive (why?).

Theorem

If a nonnegative irreducible matrix has a positive diagonal entry then it is primitive.

As an exercise, show with a counterexample that the converse does not hold.

Summary: matrix classes



Suggested reading

- R. Brualdi and D. Cvetković
  A combinatorial approach to matrix theory and its applications
  Chapman & Hall/CRC, 2009.
- C. Meyer
  Matrix Analysis and applied linear algebra
  SIAM, 2000.
  http://www.matrixanalysis.com/
- G. Strang
  Introduction to linear algebra.
  SIAM, Fifth Edition 2016.
  http://math.mit.edu/~gs/linearalgebra/