Rob 501 Handouts
Advanced Properties of Matrices
J.W. Grizzle

Advanced Properties of Matrices

Sources:

- Symmetric Matrices
  http://ocw.mit.edu/courses/mathematics/18-06sc-linear-algebra-fall-2011/positive-defi
  symmetric-matrices-and-positive-definiteness/MIT18_06SCF11_Ses3.1sum.pdf
- Positive Definite Matrices.
  http://wwwf.imperial.ac.uk/~das01/MyWeb/M3S3/Handouts/Positive.pdf
- Singular Value Decompsition
  http://www.math.tau.ac.il/~turkel/notes/SVD.pdf
- Schur Complement
  http://www.cis.upenn.edu/~jean/schur-comp.pdf
- QR Decomposition
  https://inst.eecs.berkeley.edu/~ee127a/book/login/l_mats_qr.html

Symmetric matrices and positive definiteness

Symmetric matrices are good – their eigenvalues are real and each has a complete set of orthonormal eigenvectors. Positive definite matrices are even better.

Symmetric matrices

A *symmetric matrix* is one for which $A = A^T$. If a matrix has some special property (e.g. it's a Markov matrix), its eigenvalues and eigenvectors are likely to have special properties as well. For a symmetric matrix with real number entries, the eigenvalues are real numbers and it's possible to choose a complete set of eigenvectors that are perpendicular (or even orthonormal).

If A has n independent eigenvectors we can write $A = S\Lambda S^{-1}$. If A is symmetric we can write $A = Q\Lambda Q^{-1} = Q\Lambda Q^T$, where Q is an orthogonal matrix. Mathematicians call this the *spectral theorem* and think of the eigenvalues as the "spectrum" of the matrix. In mechanics it's called the *principal axis theorem*.

In addition, any matrix of the form $Q\Lambda Q^T$ will be symmetric.

Real eigenvalues

Why are the eigenvalues of a symmetric matrix real? Suppose A is symmetric and $A\mathbf{x} = \lambda\mathbf{x}$. Then we can conjugate to get $\overline{A}\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$. If the entries of A are real, this becomes $A\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$. (This proves that complex eigenvalues of real valued matrices come in conjugate pairs.)

Now transpose to get $\overline{\mathbf{x}}^T A^T = \overline{\mathbf{x}}^T\overline{\lambda}$. Because A is symmetric we now have $\overline{\mathbf{x}}^T A = \overline{\mathbf{x}}^T\overline{\lambda}$. Multiplying both sides of this equation on the right by $\mathbf{x}$ gives:

$$\overline{\mathbf{x}}^T A\mathbf{x} = \overline{\mathbf{x}}^T\overline{\lambda}\mathbf{x}.$$

On the other hand, we can multiply $A\mathbf{x} = \lambda\mathbf{x}$ on the left by $\overline{\mathbf{x}}^T$ to get:

$$\overline{\mathbf{x}}^T A\mathbf{x} = \overline{\mathbf{x}}^T\lambda\mathbf{x}.$$

Comparing the two equations we see that $\overline{\mathbf{x}}^T\overline{\lambda}\mathbf{x} = \overline{\mathbf{x}}^T\lambda\mathbf{x}$ and, unless $\overline{\mathbf{x}}^T\mathbf{x}$ is zero, we can conclude $\lambda = \overline{\lambda}$ is real.

How do we know $\overline{\mathbf{x}}^T\mathbf{x} \neq 0$?

$$\overline{\mathbf{x}}^T\mathbf{x} = \begin{bmatrix} \overline{x}_1 & \overline{x}_2 & \cdots & \overline{x}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2.$$

If $\mathbf{x} \neq \mathbf{0}$ then $\overline{\mathbf{x}}^T\mathbf{x} \neq 0$.

With complex vectors, as with complex numbers, multiplying by the conjugate is often helpful.

Symmetric matrices with real entries have $A = A^T$, real eigenvalues, and perpendicular eigenvectors. If A has complex entries, then it will have real eigenvalues and perpendicular eigenvectors if and only if $A = \overline{A}^T$. (The proof of this follows the same pattern.)

Projection onto eigenvectors

If $A = A^T$, we can write:

$$\begin{aligned} A &= Q\Lambda Q^T \\ &= \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \\ &= \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^T + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T \end{aligned}$$

The matrix $\mathbf{q}_k \mathbf{q}_k^T$ is the projection matrix onto $\mathbf{q}_k$, so every symmetric matrix is a combination of perpendicular projection matrices.

Information about eigenvalues

If we know that eigenvalues are real, we can ask whether they are positive or negative. (Remember that the signs of the eigenvalues are important in solving systems of differential equations.)

For very large matrices A, it's impractical to compute eigenvalues by solving $|A - \lambda I| = 0$. However, it's not hard to compute the pivots, and the signs of the pivots of a symmetric matrix are the same as the signs of the eigenvalues:

$$\text{number of positive pivots} = \text{number of positive eigenvalues.}$$

Because the eigenvalues of $A + bI$ are just b more than the eigenvalues of A, we can use this fact to find which eigenvalues of a symmetric matrix are greater or less than any real number b. This tells us a lot about the eigenvalues of A even if we can't compute them directly.

Positive definite matrices

A *positive definite matrix* is a symmetric matrix A for which all eigenvalues are positive. A good way to tell if a matrix is positive definite is to check that all its pivots are positive.

Positive Definite Matrices. We are only covering Property 2. Enjoy the others!

M3S3/S4 STATISTICAL THEORY II

POSITIVE DEFINITE MATRICES

Definition: Positive Definite Matrix
A square, $p \times p$ symmetric matrix A is *positive definite* if, for all $x \in \mathbb{R}^p$,

$$x^\mathsf{T} A x > 0$$

Properties: Suppose that A

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}$$

is a positive definite matrix.

1. The $r \times r$ $(1 \leq r \leq p)$ submatrix A_r,

$$A_r = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rr} \end{bmatrix}$$

   is also positive definite.

2. The p eigenvalues of A, $\lambda_1, \ldots, \lambda_p$ are **positive**. Conversely, if all the eigenvalues of a matrix B are positive, then B is positive definite.

3. There exists a unique decomposition of A

$$A = LL^\mathsf{T} \tag{1}$$

   where L is a lower triangular matrix

$$L = [l_{ij}] = \begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pp} \end{bmatrix}.$$

   Equation (1) gives the *Cholesky Decomposition* of A.

4. There exists a unique decomposition of A

$$A = SS \tag{2}$$

   where S can be denoted $A^{1/2}$. S is the *matrix square root* of A.

5. There exists a unique decomposition of A

$$A = VDV^\mathsf{T} \tag{3}$$

where

$$D = \mathrm{diag}(\lambda_1, \ldots, \lambda_p) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}$$

is the diagonal matrix composed of the eigenvalues of A, and V is an *orthogonal matrix*

$$V^{\mathsf{T}} V = \mathbf{1}$$

Equation (3) gives the *Singular Value Decomposition* of A.

6. As $A = VDV^{\mathsf{T}}$,

$$|A| = |VDV^{\mathsf{T}}| = |V||D||V^{\mathsf{T}}| = |V|^2|D| = |D| > 0$$

as

$$|V| = 1 \qquad \text{and} \qquad |D| = \prod_{i=1}^{p} \lambda_i > 0$$

by 2 and 5.

7. By 6., as $|A| > 0$, A is *non-singular*, that is, the *inverse* of A, A^{-1} exists such that

$$AA^{-1} = A^{-1}A = \mathbf{1}.$$

In fact

$$A^{-1} = (VDV^{\mathsf{T}})^{-1} = VD^{-1}V^{\mathsf{T}}$$

as

$$V^{-1} = V^{\mathsf{T}}.$$

8. A^{-1} is positive definite.

9. For $x \in \mathbb{R}^p$,

$$\min_{1 \le i \le p} \lambda_i \le \frac{x^{\mathsf{T}} A x}{x^{\mathsf{T}} x} \le \max_{1 \le i \le p} \lambda_i$$

10. If A and B are positive definite, then

    (i) $|A + B| \le |A| + |B|$.
    (ii) If $A - B$ is positive definite, $|A| > |B|$.
    (iii) $B^{-1} - A^{-1}$ is positive definite.

Singular Value Decomposition (SVD)

Singular Value Decomposition

Notes on Linear Algebra

Chia-Ping Chen

Department of Computer Science and Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan ROC

Introduction

- The singular value decomposition, SVD, is just as amazing as the LU and QR decompositions.
- It is closely related to the diagonal form $A = Q\Lambda Q^T$ of a symmetric matrix. What happens if the matrix is not symmetric?
- It turns out that we can factorize A by $Q_1 \Sigma Q_2^T$, where Q_1, Q_2 are orthogonal and Σ is nonnegative and diagonal-like. The diagonal entries of Σ are called the singular values.

SVD Theorem

- Any $m \times n$ real matrix A can be factored into

$$A = Q_1 \Sigma Q_2^T = \text{(orthogonal)(diagonal)(orthogonal)}.$$

- The matrices are constructed as follows: The columns of Q_1 ($m \times m$) are the eigenvectors of AA^T, and the columns of Q_2 ($n \times n$) are the eigenvectors of A^TA. The r singular values on the diagonal of Σ ($m \times n$) are the square roots of the nonzero eigenvalues of both AA^T and A^TA.

Proof of SVD Theorem

The matrix $A^T A$ is real symmetric so it has a complete set of orthonormal eigenvectors: $A^T A x_j = \lambda_j x_j$, and

$$x_i^T A^T A x_j = \lambda_j x_i^T x_j = \lambda_j \delta_{ij}.$$

For positive λ_j's (say $j = 1, \ldots, r$), we define $\sigma_j = \sqrt{\lambda_j}$ and $q_j = \frac{Ax_j}{\sigma_j}$. Then $q_i^T q_j = \delta_{ij}$. Extend the q_i's to a basis for R^m. Put x's in Q_2 and q's in Q_1, then

$$(Q_1^T A Q_2)_{ij} = q_i^T A x_j = \begin{cases} 0 & \text{if } j > r, \\ \sigma_j q_i^T q_j = \sigma_j \delta_{ij} & \text{if } j \leq r. \end{cases}$$

That is, $Q_1^T A Q_2 = \Sigma$. So $A = Q_1 \Sigma Q_2^T$.

Remarks

- For positive definite matrices, SVD is identical to $Q\Lambda Q^T$. For indefinite matrices, any negative eigenvalues in Λ become positive in Σ.
- The columns of Q_1, Q_2 give orthonormal bases for the fundamental subspaces of A. (Recall that the nullspace of A^TA is the same as A).
- $AQ_2 = Q_1\Sigma$, meaning that A multiplied by a column of Q_2 produces a multiple of column of Q_1.
- $AA^T = Q_1\Sigma\Sigma^TQ_1^T$ and $A^TA = Q_2\Sigma^T\Sigma Q_2^T$, which mean that Q_1 must be the eigenvector matrix of AA^T and Q_2 must be the eigenvector matrix of A^TA.

Applications of SVD

- Through SVD, we can expand a matrix to be a sum of rank-one matrices

$$A = Q_1 \Sigma Q_2^T = u_1 \sigma_1 v_1^T + \cdots + u_r \sigma_r v_r^T.$$

- Suppose we have a 1000×1000 matrix, for a total of 10^6 entries. Suppose we use the above expansion and keep only the 50 most most significant terms. This would require $50(1 + 1000 + 1000)$ numbers, a save of space of almost 90%.
- This is used in image processing and information retrieval (e.g. Google).

SVD for Image

A picture is a matrix of gray levels. This matrix can be approximated by a small number of terms in SVD.

Pseudoinverse

- Suppose $A = Q_1 \Sigma Q_2^T$ is the SVD of an $m \times n$ matrix A. The pseudoinverse of A is defined by

$$A^+ = Q_2 \Sigma^+ Q_1^T,$$

  where Σ^+ is $n \times m$ with diagonals $\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_r}$.
- The pseudoinverse of A^+ is A, or $A^{++} = A$.
- The minimum-length least-square solution to $Ax = b$ is $x^+ = A^+ b$. This is a generalization of the least-square problem when the columns of A are not required to be independent.

Proof of Minimum Length

Multiplication by Q_1^T leaves the length unchanged, so

$$|Ax-b| = |Q_1 \Sigma Q_2^T x - b| = |\Sigma Q_2^T x - Q_1^T b| = |\Sigma y - Q_1^T b|,$$

where $y = Q_2^T x = Q_2^{-1} x$. Since Σ is a diagonal matrix, we know the minimum-length least-square solution is $y^+ = \Sigma^+ Q_1^T b$. Since $|y| = |x|$, the minimum-length least-square solution for x is

$$x^+ = Q_2 y^+ = Q_2 \Sigma Q_1^T b = A^+ b.$$

Schur Complement

The Schur Complement and Symmetric Positive Semidefinite (and Definite) Matrices

Jean Gallier

December 10, 2010

1 Schur Complements

In this note, we provide some details and proofs of some results from Appendix A.5 (especially Section A.5.5) of *Convex Optimization* by Boyd and Vandenberghe [1].

Let M be an $n \times n$ matrix written a as 2×2 block matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where A is a $p \times p$ matrix and D is a $q \times q$ matrix, with $n = p + q$ (so, B is a $p \times q$ matrix and C is a $q \times p$ matrix). We can try to solve the linear system

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix},$$

that is

$$\begin{aligned} Ax + By &= c \\ Cx + Dy &= d, \end{aligned}$$

by mimicking Gaussian elimination, that is, assuming that D is invertible, we first solve for y getting

$$y = D^{-1}(d - Cx)$$

and after substituting this expression for y in the first equation, we get

$$Ax + B(D^{-1}(d - Cx)) = c,$$

that is,

$$(A - BD^{-1}C)x = c - BD^{-1}d.$$

If the matrix $A - BD^{-1}C$ is invertible, then we obtain the solution to our system

$$\begin{aligned} x &= (A - BD^{-1}C)^{-1}(c - BD^{-1}d) \\ y &= D^{-1}(d - C(A - BD^{-1}C)^{-1}(c - BD^{-1}d)). \end{aligned}$$

The matrix, $A - BD^{-1}C$, is called the *Schur Complement* of D in M. If A is invertible, then by eliminating x first using the first equation we find that the Schur complement of A in M is $D - CA^{-1}B$ (this corresponds to the Schur complement defined in Boyd and Vandenberghe [1] when $C = B^\top$).

The above equations written as

$$\begin{aligned} x &= (A - BD^{-1}C)^{-1}c - (A - BD^{-1}C)^{-1}BD^{-1}d \\ y &= -D^{-1}C(A - BD^{-1}C)^{-1}c + (D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1})d \end{aligned}$$

yield a formula for the inverse of M in terms of the Schur complement of D in M, namely

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{pmatrix}.$$

A moment of reflexion reveals that

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & 0 \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} \end{pmatrix} \begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix},$$

and then

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} I & 0 \\ -D^{-1}C & I \end{pmatrix} \begin{pmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{pmatrix} \begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix}.$$

It follows immediately that

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & 0 \\ D^{-1}C & I \end{pmatrix}.$$

The above expression can be checked directly and has the advantage of only requiring the invertibility of D.

Remark: If A is invertible, then we can use the Schur complement, $D - CA^{-1}B$, of A to obtain the following factorization of M:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix}.$$

If $D - CA^{-1}B$ is invertible, we can invert all three matrices above and we get another formula for the inverse of M in terms of $(D - CA^{-1}B)$, namely,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}.$$

If A, D and both Schur complements $A - BD^{-1}C$ and $D - CA^{-1}B$ are all invertible, by comparing the two expressions for M^{-1}, we get the (non-obvious) formula

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}.$$

Using this formula, we obtain another expression for the inverse of M involving the Schur complements of A and D (see Horn and Johnson [5]):

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}.$$

If we set $D = I$ and change B to $-B$ we get

$$(A + BC)^{-1} = A^{-1} - A^{-1}B(I - CA^{-1}B)^{-1}CA^{-1},$$

a formula known as the *matrix inversion lemma* (see Boyd and Vandenberghe [1], Appendix C.4, especially C.4.3).

2 A Characterization of Symmetric Positive Definite Matrices Using Schur Complements

Now, if we assume that M is symmetric, so that A, D are symmetric and $C = B^\top$, then we see that M is expressed as

$$M = \begin{pmatrix} A & B \\ B^\top & D \end{pmatrix} = \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BD^{-1}B^\top & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix}^\top,$$

which shows that M is similar to a block-diagonal matrix (obviously, the Schur complement, $A - BD^{-1}B^\top$, is symmetric). As a consequence, we have the following version of "Schur's trick" to check whether $M \succ 0$ for a symmetric matrix, M, where we use the usual notation, $M \succ 0$ to say that M is positive definite and the notation $M \succeq 0$ to say that M is positive semidefinite.

Proposition 2.1 *For any symmetric matrix, M, of the form*

$$M = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix},$$

if C is invertible then the following properties hold:

(1) $M \succ 0$ iff $C \succ 0$ and $A - BC^{-1}B^\top \succ 0$.

(2) If $C \succ 0$, then $M \succeq 0$ iff $A - BC^{-1}B^\top \succeq 0$.

Proof. (1) Observe that

$$\begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix}^{-1} = \begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix}$$

and we know that for any symmetric matrix, T, and any invertible matrix, N, the matrix T is positive definite ($T \succ 0$) iff $NTN^\top$ (which is obviously symmetric) is positive definite ($NTN^\top \succ 0$). But, a block diagonal matrix is positive definite iff each diagonal block is positive definite, which concludes the proof.

(2) This is because for any symmetric matrix, T, and any invertible matrix, N, we have $T \succeq 0$ iff $NTN^\top \succeq 0$. □

Another version of Proposition 2.1 using the Schur complement of A instead of the Schur complement of C also holds. The proof uses the factorization of M using the Schur complement of A (see Section 1).

Proposition 2.2 *For any symmetric matrix, M, of the form*

$$M = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix},$$

if A is invertible then the following properties hold:

(1) $M \succ 0$ iff $A \succ 0$ and $C - B^\top A^{-1} B \succ 0$.

(2) If $A \succ 0$, then $M \succeq 0$ iff $C - B^\top A^{-1} B \succeq 0$.

When C is singular (or A is singular), it is still possible to characterize when a symmetric matrix, M, as above is positive semidefinite but this requires using a version of the Schur complement involving the pseudo-inverse of C, namely $A - BC^\dagger B^\top$ (or the Schur complement, $C - B^\top A^\dagger B$, of A). But first, we need to figure out when a quadratic function of the form $\frac{1}{2}x^\top Px + x^\top b$ has a minimum and what this optimum value is, where P is a symmetric matrix. This corresponds to the (generally nonconvex) quadratic optimization problem

$$\text{minimize} \quad f(x) = \frac{1}{2}x^\top Px + x^\top b,$$

which has no solution unless P and b satisfy certain conditions.

3 Pseudo-Inverses

We will need pseudo-inverses so let's review this notion quickly as well as the notion of SVD which provides a convenient way to compute pseudo-inverses. We only consider the case of square matrices since this is all we need. For comprehensive treatments of SVD and pseudo-inverses see Gallier [3] (Chapters 12, 13), Strang [7], Demmel [2], Trefethen and Bau [8], Golub and Van Loan [4] and Horn and Johnson [5, 6].

Recall that every square $n \times n$ matrix, M, has a *singular value decomposition*, for short, *SVD*, namely, we can write

$$M = U\Sigma V^\top,$$

where U and V are orthogonal matrices and Σ is a diagonal matrix of the form

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0),$$

where $\sigma_1 \geq \cdots \geq \sigma_r > 0$ and r is the rank of M. The σ_i's are called the *singular values* of M and they are the positive square roots of the nonzero eigenvalues of $MM^\top$ and $M^\top M$. Furthermore, the columns of V are eigenvectors of $M^\top M$ and the columns of U are eigenvectors of $MM^\top$. Observe that U and V are not unique.

If $M = U\Sigma V^\top$ is some SVD of M, we define the *pseudo-inverse*, $M^\dagger$, of M by

$$M^\dagger = V\Sigma^\dagger U^\top,$$

where

$$\Sigma^\dagger = \mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0).$$

Clearly, when M has rank $r = n$, that is, when M is invertible, $M^\dagger = M^{-1}$, so $M^\dagger$ is a "generalized inverse" of M. Even though the definition of $M^\dagger$ seems to depend on U and V, actually, $M^\dagger$ is uniquely defined in terms of M (the same $M^\dagger$ is obtained for all possible SVD decompositions of M). It is easy to check that

$$\begin{aligned} MM^\dagger M &= M \\ M^\dagger MM^\dagger &= M^\dagger \end{aligned}$$

and both $MM^\dagger$ and $M^\dagger M$ are symmetric matrices. In fact,

$$MM^\dagger = U\Sigma V^\top V\Sigma^\dagger U^\top = U\Sigma\Sigma^\dagger U^\top = U \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} U^\top$$

and

$$M^\dagger M = V\Sigma^\dagger U^\top U\Sigma V^\top = V\Sigma^\dagger\Sigma V^\top = V \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} V^\top.$$

We immediately get

$$\begin{aligned} (MM^\dagger)^2 &= MM^\dagger \\ (M^\dagger M)^2 &= M^\dagger M, \end{aligned}$$

so both $MM^\dagger$ and $M^\dagger M$ are orthogonal projections (since they are both symmetric). We claim that $MM^\dagger$ is the orthogonal projection onto the range of M and $M^\dagger M$ is the orthogonal projection onto $\mathrm{Ker}(M)^\perp$, the orthogonal complement of $\mathrm{Ker}(M)$.

Obviously, $\mathrm{range}(MM^\dagger) \subseteq \mathrm{range}(M)$ and for any $y = Mx \in \mathrm{range}(M)$, as $MM^\dagger M = M$, we have

$$MM^\dagger y = MM^\dagger Mx = Mx = y,$$

so the image of $MM^\dagger$ is indeed the range of M. It is also clear that $\mathrm{Ker}(M) \subseteq \mathrm{Ker}(M^\dagger M)$ and since $MM^\dagger M = M$, we also have $\mathrm{Ker}(M^\dagger M) \subseteq \mathrm{Ker}(M)$ and so,

$$\mathrm{Ker}(M^\dagger M) = \mathrm{Ker}(M).$$

Since $M^\dagger M$ is Hermitian, $\mathrm{range}(M^\dagger M) = \mathrm{Ker}(M^\dagger M)^\perp = \mathrm{Ker}(M)^\perp$, as claimed.

It will also be useful to see that $\mathrm{range}(M) = \mathrm{range}(MM^\dagger)$ consists of all vector $y \in \mathbb{R}^n$ such that

$$U^\top y = \begin{pmatrix} z \\ 0 \end{pmatrix},$$

with $z \in \mathbb{R}^r$.

Indeed, if $y = Mx$, then

$$U^\top y = U^\top Mx = U^\top U\Sigma V^\top x = \Sigma V^\top x = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} V^\top x = \begin{pmatrix} z \\ 0 \end{pmatrix},$$

where Σ_r is the $r \times r$ diagonal matrix $\mathrm{diag}(\sigma_1, \ldots, \sigma_r)$. Conversely, if $U^\top y = \binom{z}{0}$, then $y = U\binom{z}{0}$ and

$$\begin{aligned} MM^\dagger y &= U \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} U^\top y \\ &= U \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} U^\top U \begin{pmatrix} z \\ 0 \end{pmatrix} \\ &= U \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} \begin{pmatrix} z \\ 0 \end{pmatrix} \\ &= U \begin{pmatrix} z \\ 0 \end{pmatrix} = y, \end{aligned}$$

which shows that y belongs to the range of M.

Similarly, we claim that $\mathrm{range}(M^\dagger M) = \mathrm{Ker}(M)^\perp$ consists of all vector $y \in \mathbb{R}^n$ such that

$$V^\top y = \begin{pmatrix} z \\ 0 \end{pmatrix},$$

with $z \in \mathbb{R}^r$.

If $y = M^\dagger Mu$, then

$$y = M^\dagger Mu = V \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} V^\top u = V \begin{pmatrix} z \\ 0 \end{pmatrix},$$

for some $z \in \mathbb{R}^r$. Conversely, if $V^\top y = \binom{z}{0}$, then $y = V\binom{z}{0}$ and so,

$$\begin{aligned} M^\dagger M V \begin{pmatrix} z \\ 0 \end{pmatrix} &= V \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} V^\top V \begin{pmatrix} z \\ 0 \end{pmatrix} \\ &= V \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} \begin{pmatrix} z \\ 0 \end{pmatrix} \\ &= V \begin{pmatrix} z \\ 0 \end{pmatrix} = y, \end{aligned}$$

which shows that $y \in \mathrm{range}(M^\dagger M)$.

If M is a symmetric matrix, then in general, there is no SVD, $U\Sigma V^\top$, of M with $U = V$. However, if $M \succeq 0$, then the eigenvalues of M are nonnegative and so the nonzero eigenvalues of M are equal to the singular values of M and SVD's of M are of the form

$$M = U\Sigma U^\top.$$

Analogous results hold for complex matrices but in this case, U and V are unitary matrices and $MM^\dagger$ and $M^\dagger M$ are Hermitian orthogonal projections.

If M is a normal matrix which, means that $MM^\top = M^\top M$, then there is an intimate relationship between SVD's of M and block diagonalizations of M. As a consequence, the pseudo-inverse of a normal matrix, M, can be obtained directly from a block diagonalization of M.

If M is a (real) normal matrix, then it can be block diagonalized with respect to an orthogonal matrix, U, as

$$M = U\Lambda U^\top,$$

where Λ is the (real) block diagonal matrix,

$$\Lambda = \mathrm{diag}(B_1, \ldots, B_n),$$

consisting either of 2×2 blocks of the form

$$B_j = \begin{pmatrix} \lambda_j & -\mu_j \\ \mu_j & \lambda_j \end{pmatrix}$$

with $\mu_j \neq 0$, or of one-dimensional blocks, $B_k = (\lambda_k)$. Assume that $B_1, \ldots, B_p$ are 2×2 blocks and that $\lambda_{2p+1}, \ldots, \lambda_n$ are the scalar entries. We know that the numbers $\lambda_j \pm i\mu_j$, and the λ_{2p+k} are the eigenvalues of A. Let $\rho_{2j-1} = \rho_{2j} = \sqrt{\lambda_j^2 + \mu_j^2}$ for $j = 1, \ldots, p$, $\rho_{2p+j} = \lambda_j$ for $j = 1, \ldots, n-2p$, and assume that the blocks are ordered so that $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_n$. Then, it is easy to see that

$$UU^\top = U^\top U = U\Lambda U^\top U\Lambda^\top U^\top = U\Lambda\Lambda^\top U^\top,$$

with

$$\Lambda\Lambda^\top = \mathrm{diag}(\rho_1^2, \ldots, \rho_n^2)$$

so, the singular values, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$, of A, which are the nonnegative square roots of the eigenvalues of $AA^\top$, are such that

$$\sigma_j = \rho_j, \quad 1 \leq j \leq n.$$

We can define the diagonal matrices

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$$

where $r = \mathrm{rank}(A)$, $\sigma_1 \geq \cdots \geq \sigma_r > 0$, and

$$\Theta = \mathrm{diag}(\sigma_1^{-1} B_1, \ldots, \sigma_{2p}^{-1} B_p, 1, \ldots, 1),$$

so that Θ is an orthogonal matrix and

$$\Lambda = \Theta\Sigma = (B_1, \ldots, B_p, \lambda_{2p+1}, \ldots, \lambda_r, 0, \ldots, 0).$$

But then, we can write

$$A = U\Lambda U^\top = U\Theta\Sigma U^\top$$

and we if let $V = U\Theta$, as U is orthogonal and Θ is also orthogonal, V is also orthogonal and $A = V\Sigma U^\top$ *is an SVD for* A. Now, we get

$$A^+ = U\Sigma^+ V^\top = U\Sigma^+\Theta^\top U^\top.$$

However, since Θ is an orthogonal matrix, $\Theta^\top = \Theta^{-1}$ and a simple calculation shows that

$$\Sigma^+\Theta^\top = \Sigma^+\Theta^{-1} = \Lambda^+,$$

which yields the formula

$$A^+ = U\Lambda^+ U^\top.$$

Also observe that if we write

$$\Lambda_r = (B_1, \ldots, B_p, \lambda_{2p+1}, \ldots, \lambda_r),$$

then Λ_r is invertible and

$$\Lambda^+ = \begin{pmatrix} \Lambda_r^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

Therefore, the pseudo-inverse of a normal matrix can be computed directly from any block diagonalization of A, as claimed.

Next, we will use pseudo-inverses to generalize the result of Section 2 to symmetric matrices $M = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}$ where C (or A) is singular.

4 A Characterization of Symmetric Positive Semidefinite Matrices Using Schur Complements

We begin with the following simple fact:

Proposition 4.1 *If P is an invertible symmetric matrix, then the function*

$$f(x) = \frac{1}{2}x^\top Px + x^\top b$$

has a minimum value iff $P \succeq 0$, in which case this optimal value is obtained for a unique value of x, namely $x^* = -P^{-1}b$, and with

$$f(P^{-1}b) = -\frac{1}{2}b^\top P^{-1}b.$$

Proof. Observe that

$$\frac{1}{2}(x + P^{-1}b)^\top P(x + P^{-1}b) = \frac{1}{2}x^\top Px + x^\top b + \frac{1}{2}b^\top P^{-1}b.$$

Thus,

$$f(x) = \frac{1}{2}x^\top Px + x^\top b = \frac{1}{2}(x + P^{-1}b)^\top P(x + P^{-1}b) - \frac{1}{2}b^\top P^{-1}b.$$

If P has some negative eigenvalue, say $-\lambda$ (with $\lambda > 0$), if we pick any eigenvector, u, of P associated with λ, then for any $\alpha \in \mathbb{R}$ with $\alpha \neq 0$, if we let $x = \alpha u - P^{-1}b$, then as $Pu = -\lambda u$ we get

$$\begin{aligned} f(x) &= \frac{1}{2}(x + P^{-1}b)^\top P(x + P^{-1}b) - \frac{1}{2}b^\top P^{-1}b \\ &= \frac{1}{2}\alpha u^\top P\alpha u - \frac{1}{2}b^\top P^{-1}b \\ &= -\frac{1}{2}\alpha^2\lambda \left\|u\right\|_2^2 - \frac{1}{2}b^\top P^{-1}b, \end{aligned}$$

and as α can be made as large as we want and $\lambda > 0$, we see that f has no minimum. Consequently, in order for f to have a minimum, we must have $P \succeq 0$. In this case, as $(x + P^{-1}b)^\top P(x + P^{-1}b) \geq 0$, it is clear that the minimum value of f is achieved when $x + P^{-1}b = 0$, that is, $x = -P^{-1}b$. $\square$

Let us now consider the case of an arbitrary symmetric matrix, P.

Proposition 4.2 *If P is a symmetric matrix, then the function*

$$f(x) = \frac{1}{2}x^\top Px + x^\top b$$

has a minimum value iff $P \succeq 0$ *and* $(I - PP^{\dagger})b = 0$, *in which case this minimum value is*

$$p^* = -\frac{1}{2} b^{\top} P^{\dagger} b.$$

Furthermore, if $P = U^{\top} \Sigma U$ *is an SVD of* P, *then the optimal value is achieved by all* $x \in \mathbb{R}^n$ *of the form*

$$x = -P^{\dagger} b + U^{\top} \begin{pmatrix} 0 \\ z \end{pmatrix},$$

for any $z \in \mathbb{R}^{n-r}$, *where* r *is the rank of* P.

Proof. The case where P is invertible is taken care of by Proposition 4.1 so, we may assume that P is singular. If P has rank $r < n$, then we can diagonalize P as

$$P = U^{\top} \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} U,$$

where U is an orthogonal matrix and where Σ_r is an $r \times r$ diagonal invertible matrix. Then, we have

$$\begin{aligned} f(x) &= \frac{1}{2} x^{\top} U^{\top} \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} U x + x^{\top} U^{\top} U b \\ &= \frac{1}{2} (Ux)^{\top} \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} U x + (Ux)^{\top} U b. \end{aligned}$$

If we write $Ux = \binom{y}{z}$ and $Ub = \binom{c}{d}$, with $y, c \in \mathbb{R}^r$ and $z, d \in \mathbb{R}^{n-r}$, we get

$$\begin{aligned} f(x) &= \frac{1}{2} (Ux)^{\top} \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} U x + (Ux)^{\top} U b \\ &= \frac{1}{2} (y^{\top}, z^{\top}) \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} + (y^{\top}, z^{\top}) \begin{pmatrix} c \\ d \end{pmatrix} \\ &= \frac{1}{2} y^{\top} \Sigma_r y + y^{\top} c + z^{\top} d. \end{aligned}$$

For $y = 0$, we get

$$f(x) = z^{\top} d,$$

so if $d \neq 0$, the function f has no minimum. Therefore, if f has a minimum, then $d = 0$. However, $d = 0$ means that $Ub = \binom{c}{0}$ and we know from Section 3 that b is in the range of P (here, U is $U^{\top}$) which is equivalent to $(I - PP^{\dagger})b = 0$. If $d = 0$, then

$$f(x) = \frac{1}{2} y^{\top} \Sigma_r y + y^{\top} c$$

and as Σ_r is invertible, by Proposition 4.1, the function f has a minimum iff $\Sigma_r \succeq 0$, which is equivalent to $P \succeq 0$.

Therefore, we proved that if f has a minimum, then $(I - PP^\dagger)b = 0$ and $P \succeq 0$. Conversely, if $(I - PP^\dagger)b = 0$ and $P \succeq 0$, what we just did proves that f does have a minimum.

When the above conditions hold, the minimum is achieved if $y = -\Sigma_r^{-1}c$, $z = 0$ and $d = 0$, that is for x^* given by $Ux^* = \begin{pmatrix} -\Sigma_r^{-1}c \\ 0 \end{pmatrix}$ and $Ub = \begin{pmatrix} c \\ 0 \end{pmatrix}$, from which we deduce that

$$x^* = -U^\top \begin{pmatrix} \Sigma_r^{-1}c \\ 0 \end{pmatrix} = -U^\top \begin{pmatrix} \Sigma_r^{-1}c & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} c \\ 0 \end{pmatrix} = -U^\top \begin{pmatrix} \Sigma_r^{-1}c & 0 \\ 0 & 0 \end{pmatrix} Ub = -P^\dagger b$$

and the minimum value of f is

$$f(x^*) = -\frac{1}{2} b^\top P^\dagger b.$$

For any $x \in \mathbb{R}^n$ of the form

$$x = -P^\dagger b + U^\top \begin{pmatrix} 0 \\ z \end{pmatrix}$$

for any $z \in \mathbb{R}^{n-r}$, our previous calculations show that $f(x) = -\frac{1}{2} b^\top P^\dagger b$. $\square$

We now return to our original problem, characterizing when a symmetric matrix, $M = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}$, is positive semidefinite. Thus, we want to know when the function

$$f(x, y) = (x^\top, y^\top) \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^\top A x + 2x^\top B y + y^\top C y$$

has a minimum with respect to both x and y. Holding y constant, Proposition 4.2 implies that $f(x, y)$ has a minimum iff $A \succeq 0$ and $(I - AA^\dagger)By = 0$ and then, the minimum value is

$$f(x^*, y) = -y^\top B^\top A^\dagger B y + y^\top C y = y^\top (C - B^\top A^\dagger B) y.$$

Since we want $f(x, y)$ to be uniformly bounded from below for all x, y, we must have $(I - AA^\dagger)B = 0$. Now, $f(x^*, y)$ has a minimum iff $C - B^\top A^\dagger B \succeq 0$. Therefore, we established that $f(x, y)$ has a minimum over all x, y iff

$$A \succeq 0, \quad (I - AA^\dagger)B = 0, \quad C - B^\top A^\dagger B \succeq 0.$$

A similar reasoning applies if we first minimize with respect to y and then with respect to x but this time, the Schur complement, $A - BC^\dagger B^\top$, of C is involved. Putting all these facts together we get our main result:

Theorem 4.3 *Given any symmetric matrix,* $M = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}$*, the following conditions are equivalent:*

(1) $M \succeq 0$ *(*M *is positive semidefinite).*

(2) $A \succeq 0, \quad (I - AA^{\dagger})B = 0, \quad C - B^{\top}A^{\dagger}B \succeq 0.$

(2) $C \succeq 0, \quad (I - CC^{\dagger})B^{\top} = 0, \quad A - BC^{\dagger}B^{\top} \succeq 0.$

If $M \succeq 0$ as in Theorem 4.3, then it is easy to check that we have the following factorizations (using the fact that $A^{\dagger}AA^{\dagger} = A^{\dagger}$ and $C^{\dagger}CC^{\dagger} = C^{\dagger}$):

$$\begin{pmatrix} A & B \\ B^{\top} & C \end{pmatrix} = \begin{pmatrix} I & BC^{\dagger} \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BC^{\dagger}B^{\top} & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} I & 0 \\ C^{\dagger}B^{\top} & I \end{pmatrix}$$

and

$$\begin{pmatrix} A & B \\ B^{\top} & C \end{pmatrix} = \begin{pmatrix} I & 0 \\ B^{\top}A^{\dagger} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & C - B^{\top}A^{\dagger}B \end{pmatrix} \begin{pmatrix} I & A^{\dagger}B \\ 0 & I \end{pmatrix}.$$

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, first edition, 2004.

[2] James W. Demmel. *Applied Numerical Linear Algebra*. SIAM Publications, first edition, 1997.

[3] Jean H. Gallier. *Geometric Methods and Applications, For Computer Science and Engineering*. TAM, Vol. 38. Springer, first edition, 2000.

[4] H. Golub, Gene and F. Van Loan, Charles. *Matrix Computations*. The Johns Hopkins University Press, third edition, 1996.

[5] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, first edition, 1990.

[6] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, first edition, 1994.

[7] Gilbert Strang. *Linear Algebra and its Applications*. Saunders HBJ, third edition, 1988.

[8] L.N. Trefethen and D. Bau III. *Numerical Linear Algebra*. SIAM Publications, first edition, 1997.

QR Decomposition

The QR decomposition of a matrix

Page = 33

Basic idea

The basic goal of the QR decomposition is to *factor* a matrix as a product of two matrices (traditionally called Q, R, hence the name of this factorization). Each matrix has a simple structure which can be further exploited in dealing with, say, linear equations.

The QR decomposition is nothing else than the Gram-Schmidt procedure applied to the columns of the matrix, and with the result expressed in matrix form. Consider a $m \times n$ matrix $A = (a_1, \ldots, a_n)$, with each $a_i \in \mathbf{R}^m$ a column of A.

Case when A is full column rank

Assume first that the a_i's (the columns of A) are linearly independent. Each step of the G-S procedure can be written as

$$a_i = (a_i^T q_1) q_1 + \ldots + (a_i^T q_{i-1}) q_{i-1} + \|\tilde{q}_i\|_2 q_i, \quad i = 1, \ldots, n.$$

We write this as

$$a_i = r_{i1} q_1 + \ldots + r_{i,i-1} q_{i-1} + r_{ii} q_i, \quad i = 1, \ldots, n,$$

where $r_{ij} = (a_i^T q_j)$ $(1 \le j \le i-1)$ and $r_{ii} = \|\tilde{q}_{ii}\|_2$.

Since the q_i's are unit-length and normalized, the matrix $Q = (q_1, \ldots, q_n)$ satisfies $Q^T Q = I_n$. The QR decomposition of a $m \times n$ matrix A thus allows to write the matrix in *factored* form:

$$A = QR, \quad Q = \begin{pmatrix} q_1 & \ldots & q_n \end{pmatrix}, \quad R = \begin{pmatrix} r_{11} & r_{12} & \ldots & r_{1n} \\ 0 & r_{22} & & r_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & & 0 & r_{nn} \end{pmatrix}$$

where Q is a $m \times n$ matrix with $Q^T Q = I_n$, and R is $n \times n$,upper-triangular.

Matlab syntax

```
>> [Q,R] = qr(A,0); % A is a mxn matrix, Q is mxn orthogonal, R is nxn upper triangular
```

Example: QR decomposition of a 4x6 matrix.

Page = 34

Case when the columns are not independent

When the columns of A are not independent, at some step of the G-S procedure we encounter a zero vector $\tilde{q}_j$, which means a_j is a linear combination of $a_{j-1}, \ldots, a_1$. The modified Gram-Schmidt procedure then simply skips to the next vector and continues.

In matrix form, we obtain $A = QR$, with $Q \in \mathbf{R}^{m \times r}$, $r = \mathbf{Rank}(A)$, and R has an upper staircase form, for example:

$$R = \left(\begin{array}{cccccc} * & * & * & * & * & * \\ 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * \end{array} \right).$$

(This is simply an upper triangular matrix with some rows deleted. It is still upper triangular.)

We can permute the columns of R to bring forward the first non-zero elements in each row:

$$R = \left(\begin{array}{cc} R_1 & R_2 \end{array} \right) P^T, \quad \left(\begin{array}{c|c} R_1 & R_2 \end{array} \right) := \left(\begin{array}{ccc|ccc} * & * & * & * & * & * \\ 0 & * & 0 & * & * & * \\ 0 & 0 & * & 0 & 0 & * \end{array} \right),$$

where P is a permutation matrix (that is, its columns are the unit vectors in some order), whose effect is to permute columns. (Since P is orthogonal, $P^{-1} = P^T$.) Now, R_1 is square, upper triangular, and *invertible*, since none of its diagonal elements is zero.

The QR decomposition can be written

$$AP = Q \left(\begin{array}{cc} R_1 & R_2 \end{array} \right),$$

where

1. $Q \in \mathbf{R}^{m \times r}$, $Q^T Q = I_r$;
2. r is the *rank* of A;
3. R_1 is $r \times r$ upper triangular, invertible matrix;
4. R_2 is a $r \times (n - r)$ matrix;
5. P is a $m \times m$ permutation matrix.

Matlab syntax

```
>> [Q,R,inds] = qr(A,0);  % here inds is a permutation vector such that A(:,inds) = Q*R
```

Full QR decomposition

The *full QR decomposition* allows to write $A = QR$ where $Q \in \mathbf{R}^{m \times m}$ is *square* and orthogonal ($Q^T Q = QQ^T = I_m$). In other words, the columns of Q are an orthonormal basis for the whole output space $\mathbf{R}^m$, not just for the range of A.

We obtain the full decomposition by appending an $m \times m$ identity matrix to the columns of A: $A \rightarrow [A, I_m]$. The QR decomposition of the augmented matrix allows to write

$$AP = QR = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R_1 & R_2 \\ 0 & 0 \end{pmatrix},$$

Page = 35

where the columns of the $m \times m$ matrix $Q = [Q_1, Q_2]$ are orthogonal, and R_1 is upper triangular and invertible. (As before, P is a permutation matrix.) In the G-S procedure, the columns of Q_1 are obtained from those of A, while the columns of Q_2 come from the extra columns added to A.

The full QR decomposition reveals the rank of A: we simply look at the elements on the diagonal of R that are not zero, that is, the size of R_1.

Matlab syntax

```
>> [Q,R] = qr(A); % A is a mxn matrix, Q is mxm orthogonal, R is mxn upper triangular
```

Example: QR decomposition of a 4x6 matrix.