Chap 6. Positive Definite Matrices

a real symmetric matrix A is <u>positive definite</u>

$\Leftrightarrow$ (I) $\mathbf{x}^T A \mathbf{x} > 0$ for all nonzero real vector $\mathbf{x} \neq \mathbf{0}$.

$\Leftrightarrow$ (II) all the eigenvalues of A satisfy $\lambda_i > 0$

$\Leftrightarrow$ (III) all the pivots (without row exchanges) $d_i > 0$

$\Leftrightarrow$ (IV) $\exists$ a matrix R with independent columns s.t. $A = R^T R$.

$(\mathbf{x}^T R^T R \mathbf{x} = (R\mathbf{x})^T (R\mathbf{x}) = \|R\mathbf{x}\|^2 > 0$ if $\mathbf{x} \neq \mathbf{0})$.

① Elimination : $A = LDL^T = (L\sqrt{D})(\sqrt{D}L^T) \Rightarrow R = \sqrt{D}L^T$

② Eigenvalues : $A = Q\Lambda Q^T = (Q\sqrt{\Lambda})(\sqrt{\Lambda}Q^T) \Rightarrow R = \sqrt{\Lambda}Q^T$

③ $R = Q\sqrt{\Lambda}Q^T$: the symmetric positive definite square root of A.

- The diagonalization $A = Q\Lambda Q^T$ leads to

$\mathbf{x}^T A \mathbf{x} = \mathbf{x}^T Q \Lambda Q^T \mathbf{x} = \mathbf{y}^T \Lambda \mathbf{y}$.

If A has rank r, there are r nonzero λ's and r perfect squares in $\mathbf{y}^T \Lambda \mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_r y_r^2$.

a real symmetric matrix A is <u>positive semidefinite</u>

$\Leftrightarrow$ (I) $\mathbf{x}^T A \mathbf{x} \geq 0$ for all $\mathbf{x}$

$\Leftrightarrow$ (II) all the eigenvalues of A satisfy $\lambda_i \geq 0$

$\Leftrightarrow$ (III) No pivots are negative

$\Leftrightarrow$ (IV) There is a matrix R, possibly with dependent columns, such that $A = R^T R$

Ellipsoids in n Dimensions

For a positive definite matrix, the equation $x^T A x = 1$ describes an ellipse in 2D and an ellipsoid in 3D.

Ellipsoid: $A = \begin{bmatrix} 4 & & \\ & 1 & \\ & & \frac{1}{9} \end{bmatrix}$, $x^T A x = 4x_1^2 + x_2^2 + \frac{1}{9}x_3^2 = 1$.

Ex 3:

$A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$ and $x^T A x = 5u^2 + 8uv + 5v^2 = 1$.

The ellipse is centered at $u = v = 0$, but the axes are not so clear. The axes no longer line up with the coordinate axes. We will show that the axes of the ellipse point toward the eigenvectors of A. Because $A = A^T$, those eigenvectors and axes are orthogonal. The major axis of the ellipse corresponds to the smallest eigenvalue of A.

$\lambda_1 = 1$, $x_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$; $\lambda_2 = 9$, $x_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

New squares: $5u^2 + 8uv + 5v^2 = \left(\frac{u}{\sqrt{2}} - \frac{v}{\sqrt{2}}\right)^2 + 9\left(\frac{u}{\sqrt{2}} + \frac{v}{\sqrt{2}}\right)^2 = 1$

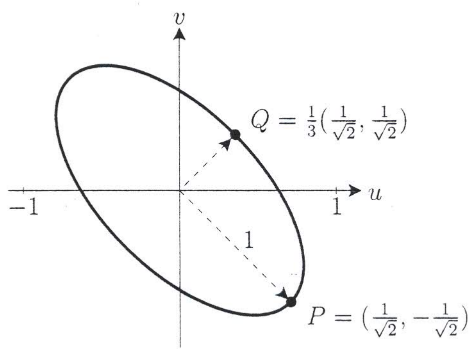

Figure 6.2 The ellipse $x^T A x = 5u^2 + 8uv + 5v^2 = 1$ and its principal axes.

$\boxed{6E}$ Suppose $A = Q\Lambda Q^T$ with $\lambda_i > 0$. Rotating $y = Q^T x$,

$x^T Q \Lambda Q^T x = 1$, $y^T \Lambda y = 1$, $\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 = 1$.

Its axes have lengths $1/\sqrt{\lambda_1}, \cdots, 1/\sqrt{\lambda_n}$ from the center. In the original x-space, they point along the eigenvectors of A.

6.3 Singular Value Decomposition

- The SVD is closely related to the $Q\Lambda Q^T$ factorization of a positive definite matrix. The eigenvalues are in the diagonal matrix Λ. The eigenvector matrix Q is orthogonal ($Q^TQ = I$) because eigenvectors of a symmetric matrix can be chosen to be orthonormal.
- For most matrices that is not true, and for rectangular matrices it is ridiculous (eigenvalues undefined). But now we allow the Q on the left and the Q^T on the right to be any two orthogonal matrices U and V^T — not necessarily transposes of each other. Then every matrix will split into $A = U\Sigma V^T$.
- The diagonal (but rectangular) matrix Σ has eigenvalues from A^TA, not from A! Those positive entries will be $\sigma_1, \cdots, \sigma_r$ (the singular values of A). They fill the first r places on the main diagonal of Σ — when A has rank r. The rest of Σ is zero.
- With rectangular matrices, the key is almost always to consider A^TA and AA^T.

Singular Value Decomposition:

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

$$A = U\Sigma V^T = (\text{orthogonal})(\text{diagonal})(\text{orthogonal})$$

The columns of U ($m \times m$) are eigenvectors of AA^T, and the columns of V ($n \times n$) are 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.

Remark 1: For positive definite matrices, Σ is Λ and $U\Sigma V^T$ is identical to $Q\Lambda Q^T$. For other symmetric matrices, any negative eigenvalues in Λ become positive in Σ.

Remark 2: U and V give orthonormal bases for all four fundamental subspaces:

$$\begin{bmatrix} \text{first } r \text{ columns of } U : \text{column space of } A \\ \text{last } m-r \text{ columns of } U : \text{left nullspace of } A \\ \text{first } r \text{ columns of } V : \text{row space of } A \\ \text{last } n-r \text{ columns of } V : \text{nullspace of } A \end{bmatrix}$$

Remark 3: The SVD chooses those bases in an extremely special way. They are more than just orthonormal. When A multiplies a column v_j of V, it produces σ_j times a column of U. That comes directly from $AV = U\Sigma$, looked at a column at a time.

Remark 4: Eigenvectors of AA^T and A^TA must go into the columns of U and V:

$$AA^T = (U\Sigma V^T)(V\Sigma^T U^T) = U\Sigma\Sigma^T U^T, \text{ and similarly}$$
$$A^TA = V\Sigma^T\Sigma V^T.$$

U must be the eigenvector matrix for AA^T. The eigenvalue matrix in the middle is $\Sigma\Sigma^T$ — which is $m\times m$ with $\sigma_1^2, \cdots, \sigma_r^2$ on the diagonal. From $A^TA = V\Sigma^T\Sigma V^T$, the V matrix must be the eigenvector matrix for A^TA. The diagonal matrix $\Sigma^T\Sigma$ $(n\times n)$ has the same $\sigma_1^2, \cdots, \sigma_r^2$.

Remark 5: Here is the reason that $Av_j = \sigma_j u_j$.

$$A^TAv_j = \sigma_j^2 v_j \Rightarrow AA^TAv_j = \sigma_j^2 (Av_j) \text{ eigenvector of } AA^T$$
$$\Downarrow$$
$$v_j^TA^TAv_j = \sigma_j^2 v_j^T v_j \Rightarrow \|Av_j\|^2 = \sigma_j^2$$

The unit eigenvector is $Av_j/\sigma_j = u_j$ $\therefore AV = U\Sigma$!

Ex 1: A has only one column: rank $r=1$, Σ has only $\sigma_1=3$.

$$\text{SVD}: \quad A=\begin{bmatrix}-1\\2\\2\end{bmatrix}=\begin{bmatrix}-\frac{1}{3}&\frac{2}{3}&\frac{2}{3}\\\frac{2}{3}&-\frac{1}{3}&\frac{2}{3}\\\frac{2}{3}&\frac{2}{3}&-\frac{1}{3}\end{bmatrix}\begin{bmatrix}3\\0\\0\end{bmatrix}\begin{bmatrix}1\end{bmatrix}=U_{3\times3}\Sigma_{3\times1}V^T_{1\times1}$$

A^TA is 1×1, whereas AA^T is 3×3. They both have eigenvalue 9 (whose square root is the 3 in Σ). The two zero eigenvalues of AA^T leave some freedom for the eigenvectors in the columns 2 and 3 of U. We kept that matrix orthogonal.

Ex 2:

Now A has rank 2 and $AA^T=\begin{bmatrix}2&-1\\-1&2\end{bmatrix}$ with $\lambda=3$ and 1.

$$\begin{bmatrix}-1&1&0\\0&-1&1\end{bmatrix}=U\Sigma V^T=\frac{1}{\sqrt{2}}\begin{bmatrix}-1&1\\1&1\end{bmatrix}\begin{bmatrix}\sqrt{3}&0&0\\0&1&0\end{bmatrix}\begin{bmatrix}1&-2&1\\-1&0&1\\1&1&1\end{bmatrix}\begin{matrix}/\sqrt{6}\\/\sqrt{2}\\/\sqrt{3}\end{matrix}$$

The columns of U are left singular vectors (unit eigenvectors of AA^T). The columns of V are right singular vectors (unit eigenvectors of A^TA).

Applications of the SVD

The SVD is terrific for numerically stable computations, because U and V are orthogonal matrices. They never change the length of a vector.

Σ could multiply by a large σ or divide by a small σ, and overflow the computer. But still Σ is as good as possible. The ratio $\sigma_{max}/\sigma_{min}$ is the condition number of an invertible $n\times n$ matrix. The availability of that information is another reason for the popularity of the SVD.

1. Image processing: We want to send a picture containing 1000×1000 color pixels. We may send 10^6 numbers. But it is better to find the essential information and send only that. Suppose we know the SVD. The key is in the singular values (in Σ). Some σ's are significant and others are extremely small. If we keep only 20 of them (throwing away 980), then we send only the corresponding 20 columns of U and V.

$$A = U\Sigma V^T = u_1\sigma_1 v_1^T + u_2\sigma_2 v_2^T + \cdots + u_r\sigma_r v_r^T$$

Any matrix is the sum of r matrices of rank 1. If only 20 terms are kept, we send 20 times 2000 numbers instead of a million (25 to 1 compression). The cost is in the computation of the SVD.

3. Polar decomposition: Every nonzero complex number z is a positive number r times a number $e^{i\theta}$: $z = r\,e^{i\theta}$.
If we think of z as a 1×1 matrix, r corresponds to a positive definite matrix and $e^{i\theta}$ corresponds to an orthogonal matrix.
The SVD extends this polar factorization to matrices of any size.

[Every real square matrix can be factored into $A = QS$, where Q is orthogonal and S is symmetric positive semidefinite. If A is invertible, then S is positive definite.]

pf)

$$A = U\Sigma V^T = (UV^T)(V\Sigma V^T) = QS$$

The factor $S = V\Sigma V^T$ is symmetric and semidefinite (because Σ is). The factor $Q = UV^T$ is an orthogonal matrix (because $Q^TQ = VU^TUV^T = I$).

Ex3: Polar decomposition

$$A = QS: \begin{bmatrix} 1 & -2 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}$$

Ex 4: Reverse polar decomposition

$$A = S'Q: \begin{bmatrix} 1 & -2 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Both S and S' are symmetric positive definite because A is invertible.

Application of $A = QS$:

a major use of the polar decomposition is in continuum mechanics (and recently in robotics and of cause in computer graphics!) In any deformation, it is important to separate stretching from rotation. The orthogonal matrix Q is a rotation, and possibly a reflection. The material feels no strain. The symmetric matrix S has eigenvalues $\sigma_1, \cdots, \sigma_r$, which are the stretching (or compression) factors. The diagonalization that displays those eigenvalues is the natural choice of axes – called principal axes. It is S that requires work on the material, and stores up elastic energy.

$S^2 = A^TA$ is symmetric positive definite when A is invertible. S is the symmetric positive definite square root of A^TA, and $Q = AS^{-1}$. In the reverse order $A = S'Q$, the matrix S' is the symmetric positive definite square root of AA^T.

4. Least squares: $A\mathbb{x}=\mathbb{b}$ has two possible difficulties.

① With dependent rows, $A\mathbb{x}=\mathbb{b}$ may have no solution.
That happens when $\mathbb{b}$ is outside the column space of A.
Instead of $A\mathbb{x}=\mathbb{b}$, we solve $A^TA\hat{\mathbb{x}}=A^T\mathbb{b}$.

② But if A has dependent columns, this $\hat{\mathbb{x}}$ will not be unique.
(any vector in the nullspace could be added to $\hat{\mathbb{x}}$.)
We have to choose a particular solution of $A^TA\hat{\mathbb{x}}=A^T\mathbb{b}$,
and we choose the shortest.

$$\boxed{\text{The optimal solution of } A\mathbb{x}=\mathbb{b} \text{ is the minimum length solution of } A^TA\hat{\mathbb{x}}=A^T\mathbb{b}.}$$

That minimum length solution will be called $\mathbb{x}^+$. It is our preferred choice as the best solution to $A\mathbb{x}=\mathbb{b}$ (which had no solution), and also to $A^TA\hat{\mathbb{x}}=A^T\mathbb{b}$ (which had too many).

Ex 5:

$A\hat{\mathbb{x}}=\mathbb{p}$ is $\begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \\ \hat{x}_4 \end{bmatrix}=\begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix}$

In the column space, the closest point to $\mathbb{b}=(b_1,b_2,b_3)^T$ is $\mathbb{p}=(b_1,b_2,0)^T$. Solving the first two equations,

$\hat{x}_1=b_1/\sigma_1$ and $\hat{x}_2=b_2/\sigma_2$

Now we face the second difficulty. To make $\hat{\mathbb{x}}$ as short as possible, we choose $\hat{x}_3=\hat{x}_4=0$. The minimum length solution is $\mathbb{x}^+$:

A^+ is pseudoinverse : $\mathbb{x}^+=\begin{bmatrix} b_1/\sigma_1 \\ b_2/\sigma_2 \\ 0 \\ 0 \end{bmatrix}=\begin{bmatrix} 1/\sigma_1 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$

$\mathbb{x}^+=A^+\mathbb{b}$ is the shortest

$$\Sigma = \begin{array}{c} \overbrace{\hphantom{\sigma_1 \quad \ddots \quad \sigma_r}}^{n} \\ \left[\begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \\ & & \end{matrix}\right] \end{array} \Big\} m \qquad \Sigma^+ = \begin{array}{c} \overbrace{\hphantom{1/\sigma_1 \quad \ddots \quad 1/\sigma_r}}^{m} \\ \left[\begin{matrix} 1/\sigma_1 & & \\ & \ddots & \\ & & 1/\sigma_r \\ & & \end{matrix}\right] \end{array} \Big\} n \qquad \Sigma^+ b = \begin{bmatrix} b_1/\sigma_1 \\ \vdots \\ b_r/\sigma_r \\ \\ \end{bmatrix}.$$

The shortest solution x^+ is always in the row space of A.
Any vector $\hat{x}$ can be split into a row space component x_r and a nullspace component: $\hat{x} = x_r + x_n$.

① The row space component also solves $A^TAx_r = A^Tb$. $(\because Ax_n = 0)$

② The components are orthogonal:
$\|\hat{x}\|^2 = \|x_r\|^2 + \|x_n\|^2$, so $\hat{x}$ is shortest when $x_n = 0$

③ all solutions of $A^TA\hat{x} = A^Tb$ have the same x_r.
That vector is x^+.

Every p in the column space comes from one and only one vector x_r in the row space. all we are doing is to choose that vector, $x^+ = x_r$, as the best solution to $Ax = b$.

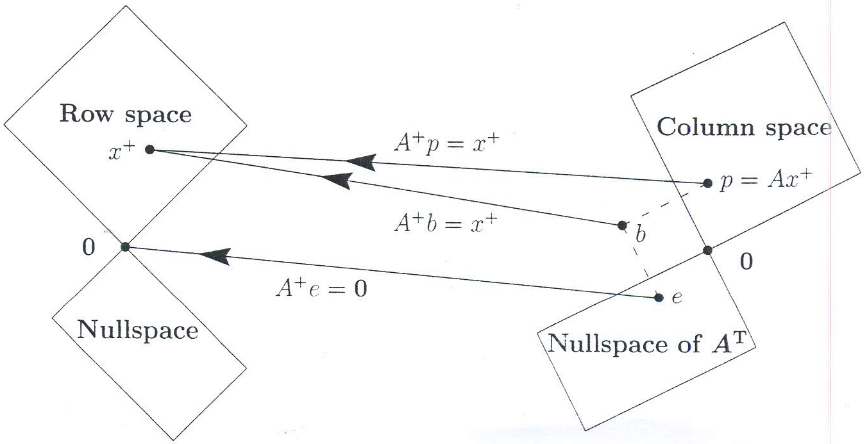

Figure 6.3 The pseudoinverse A^+ inverts A where it can on the column space.

Ex 6: $A\mathbf{x}=\mathbf{b}$ is $-x_1+2x_2+2x_3=18$, a whole plane of solutions.
The shortest solution is in the row space of $A=[-1,2,2]$, $\mathbf{x}^+=(-2,4,4)$ which satisfies the equation $A\mathbf{x}^+=18$.
There are longer solutions like $(-2,5,3)$, $(-2,7,1)$, or $(-6,3,3)$, but they all have nonzero components from the nullspace. The matrix that produces $\mathbf{x}^+$ from $\mathbf{b}=[18]$ is the pseudoinverse A^+.

$$A^+ = \begin{bmatrix} -1 & 2 & 2 \end{bmatrix}^+ = \begin{bmatrix} -\frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{9} \end{bmatrix} \quad \text{and} \quad A^+[18] = \begin{bmatrix} -2 \\ 4 \\ 4 \end{bmatrix}. \tag{6}$$

The row space of A is the column space of A^+. Here is a formula for A^+:

$$\textbf{If } \boldsymbol{A = U\Sigma V^{\mathrm{T}}} \textbf{ (the SVD), then its pseudoinverse is } \boldsymbol{A^+ = V\Sigma^+U^{\mathrm{T}}}. \tag{7}$$

Example 6 had $\sigma = 3$—the square root of the eigenvalue of $AA^{\mathrm{T}} = [9]$. Here it is again with Σ and Σ^+:

$$A = \begin{bmatrix} -1 & 2 & 2 \end{bmatrix} = U\Sigma V^{\mathrm{T}} = \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix}$$

$$V\Sigma^+U^{\mathrm{T}} = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{9} \end{bmatrix} = A^+.$$

The minimum length least-squares solution is $x^+ = A^+b = V\Sigma^+U^{\mathrm{T}}b$.

<proof> Multiplication by U^{T} leaves lengths unchanged
$\|A\mathbf{x}-\mathbf{b}\| = \|U\Sigma V^{\mathrm{T}}\mathbf{x}-\mathbf{b}\| = \|\Sigma V^{\mathrm{T}}\mathbf{x} - U^{\mathrm{T}}\mathbf{b}\|$
Let $\mathbf{y}=V^{\mathrm{T}}\mathbf{x}=V^{-1}\mathbf{x}$, with $\|V^{-1}\mathbf{x}\|=\|\mathbf{x}\|$. Then minimizing $\|A\mathbf{x}-\mathbf{b}\|$ is the same as minimizing $\|\Sigma\mathbf{y}-U^{\mathrm{T}}\mathbf{b}\|$. Now Σ is diagonal and the best $\mathbf{y}^+=\Sigma^+U^{\mathrm{T}}\mathbf{b}$, so the best $\mathbf{x}^+$ is shortest solution $\mathbf{x}^+=V\mathbf{y}^+=V\Sigma^+U^{\mathrm{T}}\mathbf{b}=A^+\mathbf{b}$.
$V\mathbf{y}^+$ is in the row space, and $A^{\mathrm{T}}A\mathbf{x}^+=A^{\mathrm{T}}\mathbf{b}$ from the SVD.