Positive Definite Matrices 2

Dhruv Kohli

Department of Mathematics
Indian Institute of Technology, Guwahati

Outline

Singular Vector Decomposition

Singular Vector Decomposition

Any m by n matrix can be factored into:
$A = U_{m\times m}\Sigma_{m\times n}V_{n\times n}^T = \text{(orthogonal)(Diagonal)(orthogonal)}$

1. $$AA^T = U\Sigma\Sigma^T U^T, \; A^T A = V\Sigma^T\Sigma V^T$$
   AA^T and $A^T A$ are positive definite if rows and columns of A are linearly independent (Why?), respectively, otherwise they are positive semidefinite (Why?).
2. Eigenvalues of AA^T and $A^T A$ are necessarily nonnegative (Why?) and the square root of the positive eigenvalues go into Σ as $\sigma_1, \sigma_2, \ldots, \sigma_r$ where r is the rank(A). Note that the number of nonzero (in our case, positive) eigenvalues equals rank(A). (Why?)

Singular Vector Decomposition contd.

3. Orthonormal eigenvectors of AA^T and A^TA can be obtained (Why?), which fill the columns of U and V respectively.
4. Proof of $Av_j = \sigma_j u_j$ and therefore, $AV = U\Sigma$:

$$A^TAv_j = \sigma_j^2 v_j \implies (AA^T)Av_j = \sigma_j^2 Av_j$$

Av_j is an eigenvector (ku_j) of AA^T

$$\|Av_j\|^2 = v_j^T A^T A v_j = \sigma_j^2 v_j^T v_j \Rightarrow Av_j = ku_j = \sigma_j u_j$$

5. U and V give orthonormal bases for all four fundamental subspaces (Why?):

$$\begin{aligned} r \text{ columns of } U&\text{: column space of } A \\ m-r \text{ columns of } U&\text{: left null space of } A \\ r \text{ columns of } V&\text{: row space of } A \\ n-r \text{ columns of } V&\text{: null space of } A \end{aligned}$$

Singular Vector Decomposition contd.

6. $A = U\Sigma V^T = \sum_{i=1}^{r} u_i \sigma_i v_i^T$, sum of $r = \text{rank}(A)$ matrices of rank 1.
7. For positive definite matrices, $\Sigma = \Lambda$ and $U = V = Q$ (Why?).
8. The columns of U are left singular vectors (unit eigenvectors of AA^T) and the columns of V are right singular vectors (unit eigenvectors of A^TA).
9. Try to workout the complex analog of above equations. Note that Σ will still be a real matrix (Why?).

Applications of SVD
Image compression

- Let A be an image of dimension 1000×1000. To transfer A one must send 1000×1000 numbers.
- $A = U\Sigma V^T$. The key is the singular values in Σ. Some σ's are significant while others are extremely small. Pick $q(= 20)$ σ's to approximate the image using,

$$A \approx U_{1:q}\Sigma_{1:q}V_{1:q} = \sum_{i=1}^{q} u_i\sigma_i v_i^T$$

- The new A can be sent using $q \times 2000 = 20 \times 2000$.

Applications of SVD

Effective rank, Polar Decomposition

- An efficient way to compute rank of a matrix is to compute the number of singular values in its SVD which are above a tolerance.
- 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.

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

- $S^2 = V\Sigma^2 V^T = A^T A$ which is positive definite if A is invertible (Why?). Then S is symmetric positive definite square root of $A^T A$ and $Q = AS^{-1}$.
- *A could be rectangular as long as $A^T A$ is positive definite* (A must have independent columns).
- In the reverse order, $A = S'Q$ where $S' = U\Sigma U^T$ is the symmetric positive definite square root of AA^T.

Applications of SVD
Least Squares

- Recall: $Ax = b$ has two possible difficulties: Rows of A are dependent in which case there might not be any solution (Why?). Then, instead of $Ax = b$ we solve $A^T A\hat{x} = A^T b$. But if Columns of A are dependent then the solution will not be unique. So, we have to choose a particular solution of $A^T A\hat{x} = A^T b$ and we choose the shortest one (x^+).
- **The shortest solution $\mathbf{x}^+$ is always in the row space of A.** Since, $\hat{x} = x_r + x_n,\ x_r \perp x_n,\ \|\hat{x}\|^2 = \|x_r\|^2 + \|x_n\|^2$ which is minimized when $x_n = 0$.
- Alls solutions of $A^T A\hat{x} = A^T b$ have same x_r. **That vector is $\mathbf{x}^+$.**

Applications of SVD

Least Squares contd.

Result 1

If $A = U\Sigma V^T$, then its pseudoinverse is $V\Sigma^+U^T$ where Σ^+ has all positive entries inverted. The minimum length least squares solution is $x^+ = A^+b = V\Sigma^+U^T$.

Proof-Hints:

First note that the minimum length least square solution of $Dy = c$ where D is diagonal is $y = D^+c$.

$$U\Sigma V^T x = b \Rightarrow \Sigma V^T x = U^T b$$

$$\Rightarrow \Sigma y = U^T b \text{ where } y = V^T x \text{ and } \|y\| = \|x\|$$

$$\Rightarrow y = \Sigma^+ U^T b \Rightarrow V^T x = \Sigma^+ U^T b$$

$$\Rightarrow x = V\Sigma^+ U^T b = A^+ b$$

Applications of SVD

Least Squares contd.

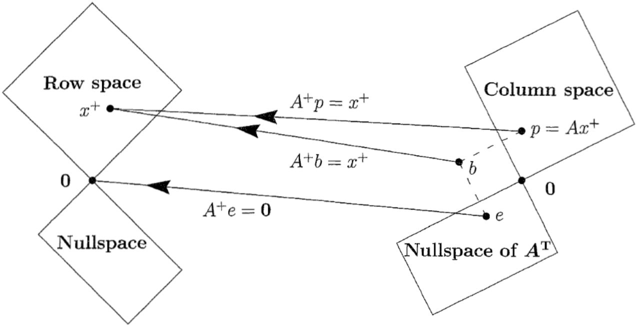

Figure: The pseudoinverse A^+ inverts A where it can on the column space.

Minimum Principles

- We want to find the minimum principle that is equivalent to $Ax = b$ and the minimization equivalent of $Ax = \lambda x$.

Result 2

If A is symmetric positive definite, then $P(x) = \frac{1}{2}x^TAx - x^Tb$ reaches its minimum at $Ax = b$ and $P_{\text{min}} = -\frac{1}{2}b^TA^{-1}b$.

Proof - Hints:

$$\begin{aligned} P(y) - P(x) &= \frac{1}{2}y^TAy - y^Tb - \frac{1}{2}x^TAx + x^Tb \\ &= \frac{1}{2}y^TAy - y^TAx + \frac{1}{2}x^TAx \\ &= \frac{1}{2}(y-x)^TA(y-x) > 0 \end{aligned}$$

In applications, $\frac{1}{2}x^TAx$ is the internal energy and $-x^Tb$ is the external work. The system automatically goes to $x = A^{-1}b$ where total energy $P(x)$ is minimum.

Minimum Principles contd.

Constrained Minimization

To minimize $P(x) = \frac{1}{2}x^T Ax - x^T b$ under constraint $Cx = d$, we need more unknowns (equal to number of equations in constraint) which are called Lagrange multipliers. Then, we minimize $L(x, y) = P(x) + y^T(Cx - d)$.

$$\frac{\partial L}{\partial x} = 0 \Rightarrow Ax + C^T y = b \text{ and } \frac{\partial L}{\partial y} = 0 \Rightarrow Cx = d$$

The minimum occurs if A is symmetric positive definite.

Minimum Principles contd.

- Those "dual unknowns" y tell how much the constrained minimum $P_{C/\min}$ exceeds the unconstrained $P_{\min}$. The sentivity of minimum is given by:

$$
\begin{aligned}
P_{C/\min} &= \frac{1}{2}(x^T b - y^T d) - x^T b = -\frac{1}{2}(x^T b - y^T d) \\
&= -\frac{1}{2}(b^T A^{-1} b - y^T C A^{-1} b - y^T d) \\
&= P_{\min} + \frac{1}{2} y^T (C A^{-1} b - d) \geq P_{\min}
\end{aligned}
$$

- Least square equations (normal equations) $A^T A \hat{x} = A^T b$ can also be obtained by minimization of $E = \|Ax - b\|^2$ which on expansion is $x^T A^T A x - 2x^T A^T b + b^T b$.

Minimum Principles contd.

The Rayleigh Quotient

Result 3 - Rayleigh Principle

For a given symmetric matrix A, the minimum value of Rayleigh quotient $R(x) = \frac{x^TAx}{x^Tx}$ is the smallest eigenvalue λ_1 which is achieved at the corresponding eigenvector x_1 and largest value is λ_n at x_n.

Proof - Hints: Use $A = Q\Lambda Q^T$ **or**

- Restrict $x^TAx = 1$. Then we need a point on this ellipsoid farthest from the origin - vector x of maximum length, which must be the longest axis x_1 (corresponds to λ_1).
- **The diagonal entries of any symmetric matrix are between $\boldsymbol{\lambda_1}$ and $\boldsymbol{\lambda_n}$** (Why?).

Minimum Principles contd.
Intertwining of the Eigenvalues

- This needs some more understanding.

The Finite Element Method

- Consider BVP $-u'' = f(x),\ u(0) = u(1) = 0$.
- The problem is infinite dimensional (the vector b is replaced by function f, and the matrix A becomes $-\frac{d^2}{dx^2}$).
- The energy whose minimum is required is given by replacing inner products with integral,

$$\begin{aligned} P(v) &= \frac{1}{2}v^T Av - v^T f \\ &= \frac{1}{2}\int_0^1 (v(x))(-v''(x))dx - \int_0^1 v(x)f(x)dx \end{aligned}$$

- $P(v)$ is minimized over all functions $v(x)$ that satisfy $v(1) = v(0) = 0$. Function giving minimum will be $u(x)$.

The Finite Element Method contd.

Rayleigh-Ritz Principle

- Using integration by parts,

$$P(v) = \frac{1}{2}\int_0^1 (v'(x))^2 dx - \int_0^1 v(x)f(x)dx$$

- The Rayleigh-Ritz principle produces an n-dimesnional problem by choosing only n trial functions $V_1(x), ., V_n(x)$. From all combinations, $V = \sum_{i=1}^{n} y_i V_i(x)$, we look for one (call it U) which minimizes $P(V)$.

$$P(V) = \frac{1}{2}\int_0^1 \left(\sum_{i=1}^{n} V_i'(x) y_i\right)^2 dx - \int_0^1 \left(\sum_{i=1}^{n} y_i V_i(x)\right) dx$$

$$= \frac{1}{2} y^T A y - y^T b, \ A_{ij} = \int V_i' V_j' dx, b_k = \int f V_k dx$$

The Finite Element Method contd.

Rayleigh-Ritz Method

Rayleigh-Ritz method has three steps:

1. Choose trial functions $V_1, V_2, \ldots, V_n$.
   - V_i's should be extremely simple to proceed further.
   - Some combination of V_i's should actually be close to $u(x)$ otherwise useless to proceed.
2. Compute the coeffecients A_{ij} and b_j.
3. Solve $Ay = b$ to find $U(x) = \sum_{i=1}^{n} y_i V_i(x)$

- **Key idea that makes finite elements successful - Use of piecewise polynomials as trial functions**.
- Example of piecewise linear finite element - V_j is "hat function" which has height 1 at node $x_j = jh$ and zero at all other nodes (h is interval length).

- Consider the problem of finding eigenfunction $u(x)$ such that $-u'' = \lambda u,\ u(0) = u(1) = 0$.
- Rayleigh quotient, $R(v) = \frac{\int_0^1 v(x)(-v''(x))dx}{\int_0^1 v(x)^2 dx} = \frac{\int_0^1 v'(x)^2}{\int_0^1 v(x)^2 dx}$.
- Using trial functions,
$$R(V) = \frac{\int_0^1 \left(\sum_{i=1}^n y_i V_i'(x)\right)^2 dx}{\int_0^1 \left(\sum_{i=1}^n y_i V_i(x)\right)^2 dx} = \frac{y^T A y}{y^T M y}$$
- Minimization of $R(V)$ is equivalent to solving generalized eigenvalue problem $Ay = \lambda My$ for the smallest eigenvalue Λ_1. Using corresponding eigenvector y_1 we approximate the eigenfunction $U = \sum_i y_{1_i} V_i$.
- For $\lambda = \pi^2$, the function $\sin \pi x$ minimizes $\frac{v^T A v}{v^T v}$.

Bibliography

Gilbert Strang. *Linear algebra and its applications*. Belmont, CA: Thomson, Brooks/Cole, 2006.