8. Positive semidefinite matrices

- positive semidefinite matrices
- examples
- Cholesky factorization
- pivoted Cholesky factorization

Positive (semi)definite matrix

a *symmetric* matrix $A \in \mathbb{R}^{n \times n}$ is

- *positive semidefinite* if $x^T A x \geq 0$ for all x
- *positive definite* if $x^T A x > 0$ for all $x \neq 0$ (a subset of the p.s.d. matrices)

note: if $A \in \mathbb{R}^{n \times n}$ is symmetric, then the function $x^T A x$ is called a *quadratic form*

$$x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j = \sum_{i=1}^{n} A_{ii} x_i^2 + 2 \sum_{i>j} A_{ij} x_i x_j$$

Other terminology (for symmetric $A \in \mathbb{R}^{n \times n}$)

- A is *negative semidefinite* if $-A$ is positive semidefinite: $x^T A x \leq 0$ for all x
- A is *negative definite* if $-A$ is positive definite: $x^T A x < 0$ for all $x \neq 0$
- A is *indefinite* if it is neither positive semidefinite nor negative semidefinite

Example

$$A = \begin{bmatrix} 9 & 6 \\ 6 & a \end{bmatrix}$$

$$x^T A x = 9x_1^2 + 12x_1x_2 + ax_2^2 = (3x_1 + 2x_2)^2 + (a-4)x_2^2$$

- A is positive definite for $a > 4$

$$x^T A x > 0 \text{ for all nonzero } x$$

- A is positive semidefinite but not positive definite for $a = 4$

$$x^T A x \geq 0 \quad \text{for all} \quad x, \quad x^T A x = 0 \quad \text{for} \quad x = (2, -3)$$

- A is not positive semidefinite for $a < 4$

$$x^T A x < 0 \quad \text{for} \quad x = (2, -3)$$

Properties

- every positive definite matrix A has positive diagonal elements

$$A_{ii} = e_i^T A e_i > 0$$

- every positive semidefinite matrix A has nonnegative diagonal elements

$$A_{ii} = e_i^T A e_i \geq 0$$

- every positive definite matrix A is nonsingular

$$Ax = 0 \quad \Longrightarrow \quad x^T A x = 0 \quad \Longrightarrow \quad x = 0$$

last step follows from positive definiteness

Singular positive semidefinite matrices

if A is positive semidefinite, but not positive definite, then it is singular

to see this, assume A is positive semidefinite and $x^TAx = 0$ for some nonzero x

- since A is positive semidefinite the following function is nonnegative:

$$\begin{aligned} f(t) &= (x - tAx)^T A(x - tAx) \\ &= x^TAx - 2tx^TA^2x + t^2x^TA^3x = -2t\|Ax\|^2 + t^2x^TA^3x \end{aligned}$$

- $f(t) \geq 0$ for all t is only possible if $Ax = 0$, so A is singular since x is nonzero

Nullspace of positive semidefinite matrix: for a positive semidefinite matrix,

$$Ax = 0 \quad \Longleftrightarrow \quad x^TAx = 0$$

- the "$\Leftarrow$" direction follows from above argument
- this does not hold for indefinite symmetric matrices; for example

$$A = \operatorname{diag}(-1, 1)$$

  is nonsingular, so $Ax = 0$ only for $x = 0$; but $x^TAx = 0$ for $x = (1, 1)$

Schur complement

partition $n \times n$ symmetric matrix A as

$$A = \begin{bmatrix} A_{11} & A_{2:n,1}^T \\ A_{2:n,1} & A_{2:n,2:n} \end{bmatrix}$$

- the *Schur complement* of A_{11} is defined as the $(n-1) \times (n-1)$ matrix

$$S = A_{2:n,2:n} - \frac{1}{A_{11}} A_{2:n,1} A_{2:n,1}^T$$

- if A is positive definite, then S is positive definite

  to see this, take any $x \neq 0$ and define $y = -(A_{2:n,1}^T x)/A_{11}$, then

$$x^T S x = \begin{bmatrix} y \\ x \end{bmatrix}^T \begin{bmatrix} A_{11} & A_{2:n,1}^T \\ A_{2:n,1} & A_{2:n,2:n} \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} > 0$$

  because A is positive definite

Gram matrix

the *Gram matrix* of a matrix B is

$$A = B^T B$$

- every Gram matrix is positive semidefinite

$$x^T A x = x^T B^T B x = \|Bx\|^2 \geq 0 \quad \text{for all} \quad x$$

- a Gram matrix is positive definite if

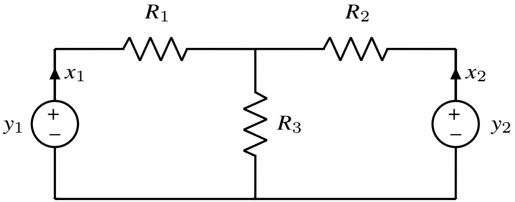
$$x^T A x = x^T B^T B x = \|Bx\|^2 > 0 \quad \text{for all} \quad x \neq 0$$

  i.e., B has linearly independent columns

Outline

- positive semidefinite matrices
- **examples**
- Cholesky factorization
- pivoted Cholesky factorization

Example: resistor circuit



mesh (loop) equations:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} R_1 + R_3 & R_3 \\ R_3 & R_2 + R_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

the matrix

$$A = \begin{bmatrix} R_1 + R_3 & R_3 \\ R_3 & R_2 + R_3 \end{bmatrix}$$

is positive definite if R_1, R_2, R_3 are positive to see this, observe

$$
\begin{aligned}
x^T A x &= (R_1 + R_3)\, x_1^2 + 2R_3 x_1 x_2 + (R_2 + R_3)\, x_2^2 \\
&= R_1 x_1^2 + R_2 x_2^2 + R_3\,(x_1 + x_2)^2 \\
&\geq 0
\end{aligned}
$$

and $x^T A x = 0$ only if $x_1 = x_2 = 0$

Physics interpretation

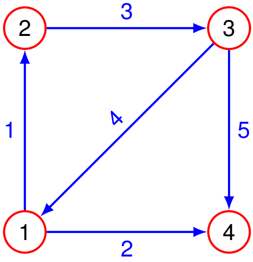
- $x^T A x = y^T x$ is the power delivered by sources, dissipated by resistors
- power dissipated by the resistors is positive unless both currents are zero

Graph Laplacian

recall the incidence matrix of a directed graph with m nodes and n edges:

$$A_{ij} = \begin{cases} 1 & \text{if edge } j \text{ point to node } i \\ -1 & \text{if edge } j \text{ point from node } i \\ 0 & \text{otherwise} \end{cases}$$

assume there are no self-loops and at most one edge between any two nodes



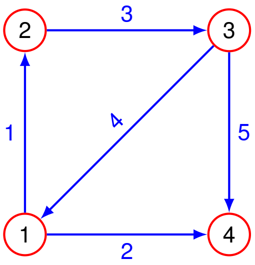
$$A = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Graph Laplacian

the positive semidefinite matrix $L = AA^T$ is called the *Laplacian* of the graph:

$$L_{ij} = \begin{cases} \text{degree of node} & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and nodes } i \text{ and } j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

the *degree* of a node is the number of edges incident to it



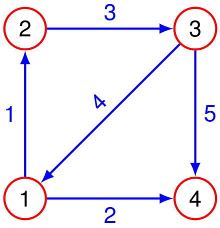
$$L = AA^T = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}$$

Laplacian quadratic form: $v^T L v = \|A^T v\|^2 = \sum_{\text{edges } i \to j} (v_j - v_i)^2$ is Dirichlet energy

Weighted graph Laplacian

- we associate a nonnegative weight w_k with edge k
- the weighted graph Laplacian is the matrix $L = A\operatorname{diag}(w)A^T$

$$L_{ij} = \begin{cases} \sum_{k \in \mathcal{N}_i} w_k & \text{if } i = j \quad \text{(where } \mathcal{N}_i \text{ are the edges incident to node } i\text{)} \\ -w_k & \text{if } i \neq j \text{ and edge } k \text{ is between nodes } i \text{ and } j \\ 0 & \text{otherwise} \end{cases}$$



$$L = \begin{bmatrix} w_1 + w_2 + w_4 & -w_1 & -w_4 & -w_2 \\ -w_1 & w_1 + w_3 & -w_3 & 0 \\ -w_4 & -w_3 & w_3 + w_4 + w_5 & -w_5 \\ -w_2 & 0 & -w_5 & w_2 + w_5 \end{bmatrix}$$

this is the conductance matrix of a resistive circuit (w_k is conductance in branch k)

Outline

- positive semidefinite matrices
- examples
- **Cholesky factorization**
- pivoted Cholesky factorization

LU factorization for positive definite matrices

a symmetric positive definite matrix $A \in \mathbb{R}^{n\times n}$ has an LU factorization:

$$A = LU$$

- L is unit lower triangular, U is upper triangular
- since A is positive definite, diagonal elements $U_{ii} > 0$ are positive
- expressing

$$U = \underbrace{\begin{bmatrix} U_{11} & & & & \\ & U_{22} & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & U_{nn} \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} 1 & \frac{U_{12}}{U_{11}} & \cdots & \cdots & \frac{U_{1n}}{U_{11}} \\ & 1 & \frac{U_{23}}{U_{22}} & \cdots & \frac{U_{2n}}{U_{22}} \\ & & \ddots & & \vdots \\ & & & \ddots & \vdots \\ & & & & 1 \end{bmatrix}}_{\tilde{U}}$$

the matrix A can be factored as $A = LD\tilde{U}$, which is unique

Symmetrizing the LU factorization

- since A is symmetric, we have

$$L D \tilde{U} = A = A^T = \tilde{U}^T D L^T$$

- since this factorization is unique, we have $L = \tilde{U}^T$ or

$$A = L D L^T$$

- if we write $D = D^{1/2} D^{1/2}$ with

$$D^{1/2} = \text{diag}(\sqrt{U_{11}}, \ldots, \sqrt{U_{nn}})$$

  we can express the LU as factorization

$$A = R^T R$$

  with $R^T = L D^{1/2}$ a lower triangular matrix

- this is called the *Cholesky factorization*

Cholesky factorization

every positive definite matrix $A \in \mathbb{R}^{n \times n}$ can be factored as

$$A = R^T R$$

- $R \in \mathbb{R}^{n \times n}$ is upper triangular with positive diagonal elements
- R is called the *Cholesky factor* of A
- can be interpreted as “square root” of a positive definite matrix

Recursive viewpoint of Cholesky factorization

$$\begin{bmatrix} A_{11} & A_{1,2:n} \\ A_{2:n,1} & A_{2:n,2:n} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 \\ R_{1,2:n}^T & R_{2:n,2:n}^T \end{bmatrix} \begin{bmatrix} R_{11} & R_{1,2:n} \\ 0 & R_{2:n,2:n} \end{bmatrix}$$

$$= \begin{bmatrix} R_{11}^2 & R_{11}R_{1,2:n} \\ R_{11}R_{1,2:n}^T & R_{1,2:n}^T R_{1,2:n} + R_{2:n,2:n}^T R_{2:n,2:n} \end{bmatrix}$$

---

given a symmetric positive definite matrix $A \in \mathbb{R}^{n \times n}$

1. compute first row of R:

$$R_{11} = \sqrt{A_{11}}, \quad R_{1,2:n} = \frac{1}{R_{11}} A_{1,2:n}$$

2. compute 2,2 block $R_{2:n,2:n}$ from

$$A_{2:n,2:n} - R_{1,2:n}^T R_{1,2:n} = R_{2:n,2:n}^T R_{2:n,2:n}$$

   this is a Cholesky factorization of order $n - 1$

Example

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 \\ R_{12} & R_{22} & 0 \\ R_{13} & R_{23} & R_{33} \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix}$$

- first row of R

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 3 & R_{22} & 0 \\ -1 & R_{23} & R_{33} \end{bmatrix} \begin{bmatrix} 5 & 3 & -1 \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix}$$

- second row of R

$$\begin{bmatrix} 18 & 0 \\ 0 & 11 \end{bmatrix} - \begin{bmatrix} 3 \\ -1 \end{bmatrix} \begin{bmatrix} 3 & -1 \end{bmatrix} = \begin{bmatrix} R_{22} & 0 \\ R_{23} & R_{33} \end{bmatrix} \begin{bmatrix} R_{22} & R_{23} \\ 0 & R_{33} \end{bmatrix}$$

$$\begin{bmatrix} 9 & 3 \\ 3 & 10 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 1 & R_{33} \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & R_{33} \end{bmatrix}$$

- third column of R : $10 - 1 = R_{33}^2$, so, $R_{33} = 3$

Example

we conclude

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 \\ R_{12} & R_{22} & 0 \\ R_{13} & R_{23} & R_{33} \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & 1 & 3 \end{bmatrix} \begin{bmatrix} 5 & 3 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}$$

Iterative computation of Cholesky factor

given a symmetric positive definite $A \in \mathbb{R}^{n\times n}$

set $S = A$ and initialize $R \in \mathbb{R}^{n\times n}$ as the zero matrix

for $k = 1, \ldots, n$

1. compute the diagonal entry

$$R_{kk} = \sqrt{S_{kk}}$$

2. compute the rest of the kth row

$$R_{k,k+1:n} = S_{k,k+1:n}/R_{kk}$$

3. update the trailing submatrix

$$S_{k+1:n,k+1:n} = S_{k+1:n,k+1:n} - R^T_{k,k+1:n} R_{k,k+1:n}$$

- algorithm gives the factorization $A = R^T R$
- complexity is $(1/3)n^3$ flops
- half the memory space and half the flops of the general LU factorization

Solving equations with positive definite A

---

given: $Ax = b$ with positive definite $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$

1. factor A as $A = R^T R$
2. solve $R^T R x = b$ in two steps
   (a) forward substitution: solve $R^T y = b$
   (b) back substitution: solve $Rx = y$

Complexity: $(1/3)n^3 + 2n^2 \approx (1/3)n^3$ flops

Computing Cholesky factorization of Gram matrix

- suppose A is an $m \times n$ matrix with linearly independent columns
- the Gram matrix $C = A^T A$ is positive definite

two methods for computing the Cholesky factor of C, given A

1. compute $C = A^T A$, then Cholesky factorization of C

$$C = R^T R$$

2. compute QR factorization $A = QR$; since

$$C = A^T A = R^T Q^T Q R = R^T R$$

the matrix R is the Cholesky factor of C

Example

$$A = \begin{bmatrix} 3 & -6 \\ 4 & -8 \\ 0 & 1 \end{bmatrix}, \quad C = A^T A = \begin{bmatrix} 25 & -50 \\ -50 & 101 \end{bmatrix}$$

1. Cholesky factorization:

$$C = \begin{bmatrix} 5 & 0 \\ -10 & 1 \end{bmatrix} \begin{bmatrix} 5 & -10 \\ 0 & 1 \end{bmatrix}$$

2. QR factorization

$$A = \begin{bmatrix} 3 & -6 \\ 4 & -8 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3/5 & 0 \\ 4/5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & -10 \\ 0 & 1 \end{bmatrix}$$

Rank and symmetric matrix product

for any matrix B,

$$\mathrm{rank}(BB^T) = \mathrm{rank}(B)$$

Proof: suppose B is $n \times p$ and $\mathrm{rank}(B) = r$

- factor $B = CD$, C is $n \times r$, D is $r \times p$, $\mathrm{rank}(C) = \mathrm{rank}(D) = r$ (page 4.18):

$$BB^T = C(DD^T)C^T$$

- the matrix DD^T is positive definite because D has full row rank
- let R be the $r \times r$ Cholesky factor of $DD^T = R^TR$ and define $\tilde{B} = CR^T$:

$$BB^T = CR^TRC^T = \tilde{B}\tilde{B}^T$$

- the matrices C and $\tilde{B} = CR^T$ are $n \times r$ and have rank r
- this implies that $\mathrm{rank}(BB^T) = \mathrm{rank}(\tilde{B}\tilde{B}^T) = r$ (see page 4.40)

Outline

- positive semidefinite matrices
- examples
- Cholesky factorization
- **pivoted Cholesky factorization**

Pivoted Cholesky factorization

the following factorization exists for every positive semidefinite A:

$$A = P^T R^T R P$$

- P is a permutation matrix
- R is $r \times n$, leading $r \times r$ submatrix is upper triangular with positive diagonal:

$$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1r} & R_{1,r+1} & \cdots & R_{1n} \\ 0 & R_{22} & \cdots & R_{2r} & R_{2,r+1} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & R_{rr} & R_{r,r+1} & \cdots & R_{rn} \end{bmatrix}$$

- can be chosen to satisfy $R_{11} \geq R_{22} \geq \cdots \geq R_{rr} > 0$
- r is the rank of A

Interpretation: we permute the rows and columns of A and factor

$$P^T A P = R^T R$$

Factorization theorem for positive semidefinite matrices

a positive semidefinite $n \times n$ matrix A has rank r if and only it can be factored as

$$A = BB^T$$

where B is $n \times r$ with linearly independent columns

- “if” statement follows from page 8.23
- the pivoted Cholesky factorization proves the “only if” part
- other algorithms (symmetric eigendecomposition) can also show “only if” part

Sparse Cholesky factorization

if A is sparse and positive definite, it is usually factored as

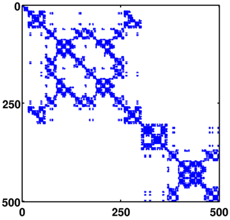
$$A = PR^TRP^T$$

P a permutation matrix; R upper triangular with positive diagonal elements

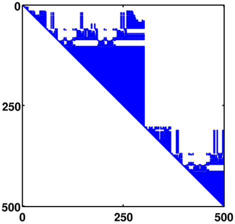
- if A is very sparse, R is often (but not always) sparse
- choice of permutation greatly affects the sparsity R
- there exist several heuristic methods for choosing a good permutation
- if R is sparse, the cost of the factorization is much less than $(1/3)n^3$

Example

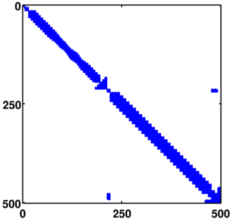
sparsity pattern of A



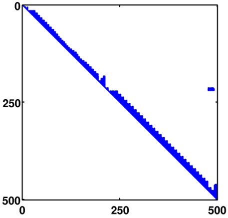
Cholesky factor of A



pattern of P^TAP



Cholesky factor of P^TAP



References and further readings

- L. Vandenberghe, *EE133A Lecture Notes,* University of California, Los Angeles.
- L. Vandenberghe. *EE133B Lecture Notes*, University of California, Los Angeles.