2. Positive semidefinite matrices

- definitions
- covariance matrix
- graph Laplacian
- pivoted Cholesky factorization

Positive semidefinite matrices

recall that an $n \times n$ symmetric matrix A is

- *positive semidefinite* if $x^T A x \geq 0$ for all x
- *positive definite* if $x^T A x > 0$ for all $x \neq 0$

the function $x^T A x$ is called a *quadratic form:*

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

Related terminology (for symmetric A)

- 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* it is not positive semidefinite or negative semidefinite

Nullspace of positive semidefinite matrix

for a positive semidefinite matrix,

$$Ax = 0 \qquad \Longleftrightarrow \qquad x^T A x = 0 \tag{1}$$

to show the "$\Leftarrow$" direction, assume x is nonzero and satisfies $x^T A x = 0$

- since A is positive semidefinite, the following function is nonnegative for all t:

$$f(t) = (x - tAx)^T A(x - tAx) = -2t\|Ax\|^2 + t^2 x^T A^3 x$$

- $f(t) \geq 0$ for all t is only possible if $Ax = 0$

note that (1) does not hold for indefinite symmetric matrices: the matrix

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

is nonsingular, so $Ax = 0$ only for $x = 0$; however $x^T A x = 0$ for $x = (1, 1)$

Positive semidefinite matrices in factored form

we will often encounter symmetric matrices in the product form

$$A = BB^T \tag{2}$$

- every matrix of this form is positive semidefinite:

$$x^T A x = x^T B B^T x = (B^T x)^T (B^T x) = \|B^T x\|^2 \geq 0$$

- on the next page, we show that

$$\operatorname{rank}(A) = \operatorname{rank}(B)$$

- later in the lecture we show that every p.s.d. matrix A has a factorization (2)

Rank and symmetric matrix product

we show that for any matrix B,

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

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

- factor $B = CD$ where C is $n \times r$, D is $r \times p$, $\text{rank}(C) = \text{rank}(D) = r$ (page 1.27):

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

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

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

- the matrices C and $\tilde{B} = CR^T$ are $n \times r$ and have rank r
- this implies that $\text{rank}(BB^T) = \text{rank}(\tilde{B}\tilde{B}^T) = r$ (see page 1.12)

Exercises

verify the following facts (A and B are symmetric $n \times n$ matrices)

1a. if A is p.s.d. and $\alpha \geq 0$, then αA is p.s.d

1b. if A is p.d. and $\alpha > 0$, then αA is p.d.

2a. if A and B are p.s.d., then $A + B$ is p.s.d.

2b. if A is p.s.d. and B is p.d., then $A + B$ is p.d.

3a. if A is p.s.d. and C is an $n \times m$ matrix then $C^T AC$ is p.s.d.

3b. if A is p.d. and C is $n \times m$ with linearly independent columns, then $C^T AC$ is p.d.

4. if A is p.d. then A^{-1} is p.d.

p.s.d. stands for positive semidefinite; p.d. stands for positive definite

Outline

- definitions
- **covariance matrix**
- graph Laplacian
- pivoted Cholesky factorization

Variance and covariance of random variables

let $x = (x_1, x_2, \ldots, x_n)$ be a random n-vector, with

$$\mu_i = \mathbf{E}\, x_i, \qquad \sigma_i = \sqrt{\mathbf{E}\,(x_i - \mu_i)^2}, \qquad \sigma_{ij} = \mathbf{E}\left((x_i - \mu_i)(x_j - \mu_j)\right) \quad \text{for } i \neq j$$

($\mathbf{E}$ denotes expectation)

- μ_i is the *mean* or *expected value* of x_i
- σ_i is the *standard deviation* and σ_i^2 is the *variance* of x_i
- σ_{ij}, for $i \neq j$, is the *covariance* of x_i and x_j
- $\rho_{ij} = \sigma_{ij}/(\sigma_i \sigma_j)$, for $i \neq j$, is the *correlation* between x_i and x_j
- variables x_i and x_j are *uncorrelated* if $\sigma_{ij} = 0$

Second moment matrix

the *second moment matrix* is the symmetric $n \times n$ matrix with i, j element $\mathbf{E}\,(x_i x_j)$:

$$S = \begin{bmatrix} \mathbf{E}(x_1^2) & \mathbf{E}\,(x_1 x_2) & \cdots & \mathbf{E}\,(x_1 x_n) \\ \mathbf{E}\,(x_2 x_1) & \mathbf{E}\,x_2^2 & \cdots & \mathbf{E}\,(x_2 x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{E}\,(x_n x_1) & \mathbf{E}\,(x_n x_2) & \cdots & \mathbf{E}\,x_n^2 \end{bmatrix} = \mathbf{E}\,(x x^T)$$

- on the right-hand side, expectation of a matrix applies element-wise
- the second moment matrix is positive semidefinite: for all a,

$$a^T S a = a^T\,\mathbf{E}\,(x x^T) a = \mathbf{E}\,(a^T x x^T a) = \mathbf{E}\,(a^T x)^2 \geq 0$$

$a^T S a$ is the expected value of the square of the scalar random variable $y = a^T x$

Covariance matrix

the *covariance matrix* (or *variance-covariance matrix*) is the symmetric $n \times n$ matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} = \mathbf{E}\left(\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ \vdots \\ x_n - \mu_n \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ \vdots \\ x_n - \mu_n \end{bmatrix}^T \right)$$

$$= \mathbf{E}\left((x - \mu)(x - \mu)^T\right)$$

- μ is the vector of means:

$$\mu = (\mu_1, \mu_2, \ldots, \mu_n) = (\mathbf{E}\, x_1, \mathbf{E}\, x_2, \ldots, \mathbf{E}\, x_n)$$

- the covariance matrix is positive semidefinite: for all a,

$$a^T \Sigma a = a^T\, \mathbf{E}\left((x - \mu)(x - \mu))^T\right)a = \mathbf{E}\left(a^T(x - \mu)\right)^2 \geq 0$$

  $a^T \mu$ is the mean and $a^T \Sigma a$ is the variance of the random variable $y = a^T x$

Correlation matrix

the *correlation matrix* has i, j element $\rho_{ij} = \sigma_{ij}/(\sigma_i \sigma_j)$ for $i \neq j$ and 1 for $i = j$:

$$C = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{bmatrix}$$

- $C = D\Sigma D$ where Σ is the covariance matrix and D is the diagonal matrix

$$D = \begin{bmatrix} \sigma_1^{-1} & 0 & \cdots & 0 \\ 0 & \sigma_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^{-1} \end{bmatrix}$$

  the expression $C = D\Sigma D$ shows that C is positive semidefinite

- C is the covariance matrix of the standardized variables $u_i = (x_i - \mu_i)/\sigma_i$

Linear combinations of random vectors

Scalar multiplication

- suppose x is a random vector with mean μ and covariance matrix Σ
- the random vector $y = \alpha x$ has mean and covariance matrix

$$\mathbf{E}\, y = \alpha\mu, \qquad \mathbf{E}\,((y - \mathbf{E}\, y)(y - \mathbf{E}\, y)^T) = \alpha^2\Sigma$$

Sum

- x, y are random n-vectors with means μ_x, μ_y, covariance matrices Σ_x, Σ_y
- the random vector $z = x + y$ has mean $\mathbf{E}\, z = \mu_x + \mu_y$
- if x and y are uncorrelated, *i.e.*,

$$\mathbf{E}\,(x_i - \mu_{x,i})(y_j - \mu_{y,j})) = 0, \quad i, j = 1, \ldots, n,$$

  then the covariance matrix of z is

$$\mathbf{E}\,((z - \mathbf{E}\, z)(z - \mathbf{E}\, z)^T) = \Sigma_x + \Sigma_y$$

Affine transformation

- suppose y is a random n-vector with mean μ_y and covariance matrix Σ_y
- define the random m-vector

  $$x = Ay + b$$

  where A is an $m \times n$ matrix, b is an m-vector
- the mean of x is

  $$\mathbf{E}\, x = \mathbf{E}\,(Ay + b) = A\mu_y + b$$

- the covariance matrix of x is

$$\begin{aligned} \Sigma &= \mathbf{E}\,((x - \mathbf{E}\, x)(x - \mathbf{E}\, x)^T) \\ &= \mathbf{E}\,((Ay - A\mu_y)(Ay - A\mu_y)^T) \\ &= A\,\mathbf{E}\,((y - \mu_y)(y - \mu_y)^T)A^T \\ &= A\Sigma_y A^T \end{aligned}$$

Example: factor model

suppose a random n-vector x has covariance matrix

$$\Sigma = AA^T + \sigma^2 I$$

x can be interpreted as being generated by a model $x = \mu + Ay + w$

- μ is the mean of x
- y is a random variable with mean zero and covariance matrix I
- w is random error or noise, uncorrelated with y, with $\mathbf{E}\, w = 0$, $\mathbf{E}\, ww^T = \sigma^2 I$

in statistics, this is known as a *factor model*

- components of y are common factors in x
- $x - \mu$ is a vector Ay in a subspace $\mathrm{range}(A)$ plus noise w

Estimate of mean

suppose the rows x_i^T in an $m \times n$ matrix X are observations of a random n-vector x

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_m^T \end{bmatrix}$$

- the sample estimate for the mean $\mathbf{E}\,x$ is

$$\bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i = \frac{1}{m} X^T \mathbf{1}$$

- subtracting $\bar{x}^T$ from each row gives the *centered* data matrix

$$X_\mathrm{c} = X - \mathbf{1}\bar{x}^T = (I - \frac{1}{m}\mathbf{1}\mathbf{1}^T)X$$

Estimate of covariance matrix

- the (Gram) matrix of X gives an an estimate of the second moment $\mathbf{E}\,(xx^T)$:

$$\frac{1}{m}X^T X = \frac{1}{m}\sum_{i=1}^{m} x_i x_i^T$$

- the Gram matrix of X_{c} gives an estimate of the covariance matrix:

$$\frac{1}{m}X_{\mathrm{c}}^T X_{\mathrm{c}} = \frac{1}{m}\sum_{i=1}^{m}(x_i - \bar{x})(x_i - \bar{x})^T$$

  this can also be expressed as

$$\begin{aligned}
\frac{1}{m}X_{\mathrm{c}}^T X_{\mathrm{c}} &= \frac{1}{m}X^T(I - \frac{1}{m}\mathbf{1}\mathbf{1}^T)^2 X \\
&= \frac{1}{m}X^T(I - \frac{1}{m}\mathbf{1}\mathbf{1}^T) X \\
&= \frac{1}{m}X^T X - \bar{x}\bar{x}^T
\end{aligned}$$

Outline

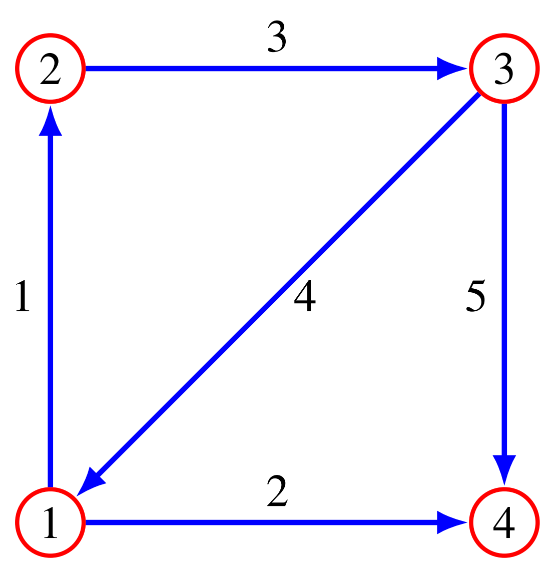
- definitions
- covariance matrix
- **graph Laplacian**
- pivoted Cholesky factorization

Incidence matrix

directed graph with m vertices, n edges, $m \times n$ incidence matrix A (page 1.6)

$$A_{ij} = \begin{cases} 1 & \text{edge } j \text{ ends at vertex } i \\ -1 & \text{edge } j \text{ starts at vertex } i \\ 0 & \text{otherwise} \end{cases}$$

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



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

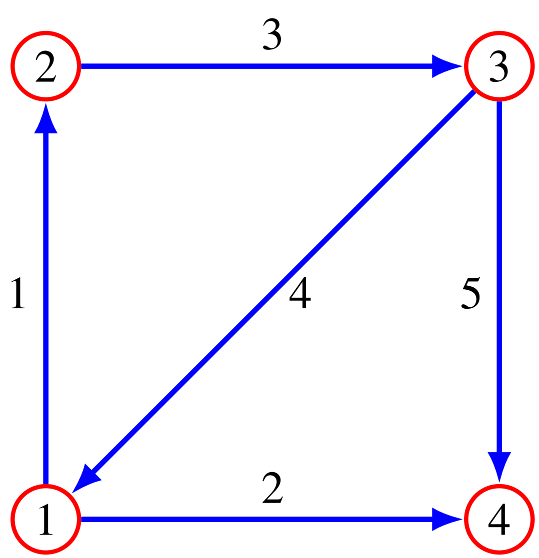
Graph Laplacian

the matrix $L = AA^T$ is called the graph *Laplacian*

- a symmetric $m \times m$ matrix with elements

$$L_{ij} = \begin{cases} \text{\#edges incident to vertex } i & \text{if } i = j \\ -1 & \text{if } i \neq j, \text{ and vertices } i \text{ and } j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

- does not depend on the orientation of the edges
- L is positive semidefinite with $\mathrm{rank}(L) = \mathrm{rank}(A)$

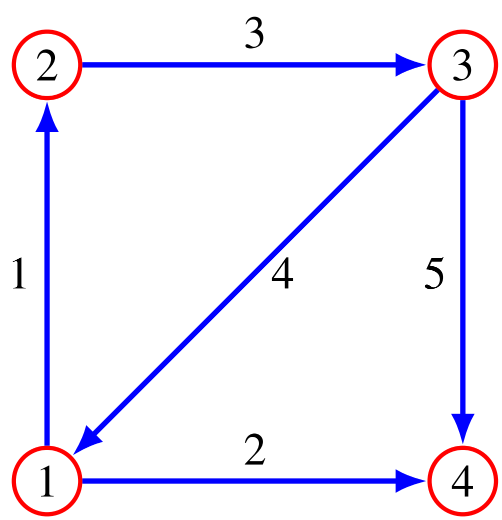


$$L = AA^T = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}$$

Weighted graph Laplacian

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

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



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

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

Outline

- definitions
- covariance matrix
- graph Laplacian
- **pivoted Cholesky factorization**

Pivoted Cholesky factorization

we show that 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

(Standard) Colesky factorization

the algorithm for Cholesky factorization $A = R^T R$ can be summarized as follows

- after k steps we have completed a partial factorization

$$A = \left[\begin{array}{ccc|c} R_{11} & \cdots & 0 & \\ \vdots & \ddots & \vdots & 0 \\ R_{1k} & \cdots & R_{kk} & \\ \hline R_{1,k+1} & \cdots & R_{k,k+1} & \\ \vdots & & \vdots & I \\ R_{1n} & \cdots & R_{kn} & \end{array}\right] \left[\begin{array}{cc} I & 0 \\ 0 & S_k \end{array}\right] \left[\begin{array}{ccc|ccc} R_{11} & \cdots & R_{1k} & R_{1,k+1} & \cdots & R_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & R_{kk} & R_{k,k+1} & \cdots & R_{kn} \\ \hline & 0 & & & I & \end{array}\right]$$

$$= \left[\begin{array}{cc} R_{1:k,1:k}^T & 0 \\ R_{1:k,(k+1):n}^T & I \end{array}\right] \left[\begin{array}{cc} I & 0 \\ 0 & S_k \end{array}\right] \left[\begin{array}{cc} R_{1:k,1:k} & R_{1:k,(k+1):n} \\ 0 & I \end{array}\right]$$

- row $k+1$ of R and the matrix S_{k+1} are found from the equality

$$S_k = \left[\begin{array}{cc} R_{k+1,k+1} & 0 \\ R_{k+1,(k+2):n}^T & I \end{array}\right] \left[\begin{array}{cc} 1 & 0 \\ 0 & S_{k+1} \end{array}\right] \left[\begin{array}{cc} R_{k+1,k+1} & R_{k+1,(k+2):n} \\ 0 & I \end{array}\right]$$

Update in standard Cholesky factorization

$$S_k = \begin{bmatrix} R_{k+1,k+1} & 0 \\ R^T_{k+1,(k+2):n} & I \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & S_{k+1} \end{bmatrix} \begin{bmatrix} R_{k+1,k+1} & R_{k+1,(k+2):n} \\ 0 & I \end{bmatrix}$$

- to simplify notation we partition S_k as $S_k = \begin{bmatrix} a & b^T \\ b & C \end{bmatrix}$
- row $k+1$ of R follows from

$$R_{k+1,k+1} = \sqrt{a}, \qquad R_{k+1,(k+2):n} = \frac{1}{R_{k+1,k+1}} b^T = \frac{1}{\sqrt{a}} b^T$$

- new matrix S_{k+1} is

$$S_{k+1} = C - R^T_{k+1,(k+2):n} R_{k+1,(k+2):n} = C - \frac{1}{a} b b^T$$

- the update fails when $(S_k)_{11} = a \leq 0$ (indicating that A is not positive definite)

Pivoted Cholesky factorization algorithm

the algorithm is readily extended to compute the pivoted Cholesky factorization

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

- after k steps we have computed a partial factorization

$$P_k A P_k^T = \begin{bmatrix} R_{1:k,1:k}^T & 0 \\ R_{1:k,(k+1):n}^T & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & S_k \end{bmatrix} \begin{bmatrix} R_{1:k,1:k} & R_{1:k,(k+1):n} \\ 0 & I \end{bmatrix}$$

- initially, $P_0 = I$ and $S_0 = A$
- if $S_k = 0$, the algorithm terminates with $r = k$
- before step $k + 1$ we reorder S_k to move largest diagonal element to position 1,1
- this reordering requires modifying P_k and reordering the columns of $R_{1:k,(k+1):n}$

Example

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix} = \begin{bmatrix} 6 & 3 & 10 & -1 \\ 3 & 18 & 15 & 0 \\ 10 & 15 & 25 & -5 \\ -1 & 0 & -5 & 6 \end{bmatrix}$$

Step 1

- apply symmetric reordering to move A_{33} to the 1,1 position
- find first row of R and S_1

$$\begin{bmatrix} A_{33} & A_{31} & A_{32} & A_{34} \\ A_{13} & A_{11} & A_{12} & A_{14} \\ A_{23} & A_{21} & A_{22} & A_{24} \\ A_{43} & A_{41} & A_{42} & A_{44} \end{bmatrix} = \begin{bmatrix} 25 & 10 & 15 & -5 \\ 10 & 6 & 3 & -1 \\ 15 & 3 & 18 & 0 \\ -5 & -1 & 0 & 6 \end{bmatrix}$$

$$= \left[\begin{array}{c|ccc} 5 & 0 & 0 & 0 \\ \hline 2 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & 2 & -3 & 1 \\ 0 & -3 & 9 & 3 \\ 0 & 1 & 3 & 5 \end{array}\right] \left[\begin{array}{c|ccc} 5 & 2 & 3 & -1 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

Example

Step 2

- move second diagonal element of S_1 to first position
- compute second row of R and S_2

$$
\begin{bmatrix} A_{33} & A_{32} & A_{31} & A_{34} \\ A_{23} & A_{22} & A_{21} & A_{24} \\ A_{13} & A_{12} & A_{11} & A_{14} \\ A_{43} & A_{42} & A_{41} & A_{44} \end{bmatrix}
$$

$$
= \left[\begin{array}{c|ccc} 5 & 0 & 0 & 0 \\ \hline 3 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & 9 & -3 & 3 \\ 0 & -3 & 2 & 1 \\ 0 & 3 & 1 & 5 \end{array}\right] \left[\begin{array}{c|ccc} 5 & 3 & 2 & -1 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]
$$

$$
= \left[\begin{array}{cc|cc} 5 & 0 & 0 & 0 \\ 3 & 3 & 0 & 0 \\ \hline 2 & -1 & 1 & 0 \\ -1 & 1 & 0 & 1 \end{array}\right] \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 4 \end{array}\right] \left[\begin{array}{cc|cc} 5 & 3 & 2 & -1 \\ 0 & 3 & -1 & 1 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]
$$

Example

Step 3

- move second diagonal element of S_2 to first position
- compute third row of R and S_3

$$
\begin{bmatrix} A_{33} & A_{32} & A_{34} & A_{31} \\ A_{23} & A_{22} & A_{24} & A_{21} \\ A_{43} & A_{42} & A_{44} & A_{41} \\ A_{13} & A_{12} & A_{14} & A_{11} \end{bmatrix}
$$

$$
= \left[\begin{array}{cc|cc} 5 & 0 & 0 & 0 \\ 3 & 3 & 0 & 0 \\ \hline -1 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right] \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 4 & 2 \\ 0 & 0 & 2 & 1 \end{array}\right] \left[\begin{array}{cc|cc} 5 & 3 & -1 & 2 \\ 0 & 3 & 1 & -1 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]
$$

$$
= \left[\begin{array}{ccc|c} 5 & 0 & 0 & 0 \\ 3 & 3 & 0 & 0 \\ -1 & 1 & 2 & 0 \\ \hline 2 & -1 & 1 & 1 \end{array}\right] \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right] \left[\begin{array}{ccc|c} 5 & 3 & -1 & 2 \\ 0 & 3 & 1 & -1 \\ 0 & 0 & 2 & 1 \\ \hline 0 & 0 & 0 & 1 \end{array}\right]
$$

Example

Result: since S_3 is zero, the algorithm terminates with the factorization

$$
\begin{bmatrix} A_{33} & A_{32} & A_{34} & A_{31} \\ A_{23} & A_{22} & A_{24} & A_{21} \\ A_{43} & A_{42} & A_{44} & A_{41} \\ A_{13} & A_{12} & A_{14} & A_{11} \end{bmatrix} = \begin{bmatrix} 25 & 15 & -5 & 10 \\ 15 & 18 & 0 & 3 \\ -5 & 0 & 6 & -1 \\ 10 & 3 & -1 & 6 \end{bmatrix}
$$

$$
= \begin{bmatrix} 5 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & 1 & 2 \\ 2 & -1 & 1 \end{bmatrix} \begin{bmatrix} 5 & 3 & -1 & 2 \\ 0 & 3 & 1 & -1 \\ 0 & 0 & 2 & 1 \end{bmatrix}
$$

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 2.5
- the pivoted Cholesky factorization proves the "only if" part
- other algorithms (symmetric eigenvalue decomposition) will be discussed later

Exercises

Exercise 1: explain why the pivoted Cholesky factorization returns a matrix R

$$R_{11} \geq R_{22} \geq \cdots \geq R_{rr}$$

Exercise 2: suppose A is a *symmetric* $n \times n$ matrix that satisfies

$$A^2 = A$$

1. show that A is positive semidefinite
2. since A is positive semidefinite, it can be factored as

   $$A = BB^T$$

   where B of size $n \times r$ and $r = \operatorname{rank}(A)$; show that B has orthonormal columns

hence, A is an orthogonal projection matrix (see 133A page 5.17)

Exercises

Exercise 3

we use $\circ$ to denote the component-wise product of matrices: if A, B are $n \times n$, then

$$(A \circ B)_{ij} = A_{ij} B_{ij}, \quad i, j = 1, \ldots, n$$

1. suppose D is $n \times r$ with columns d_k; show that

$$(DD^T) \circ B = \sum_{k=1}^{r} \mathbf{diag}(d_k) B \, \mathbf{diag}(d_k)$$

2. show that $A \circ B$ is positive semidefinite if A and B are positive semidefinite
3. show that $A \circ B$ is positive definite if A and B are positive definite

Exercises

Exercise 4

as an application of exercise 3, let

$$f(x) = c_0 + c_1 x + \cdots + c_d x^d$$

be a polynomial with nonnegative coefficients $c_0, \ldots, c_d$

suppose X is $n \times n$ and p.s.d., and define $Y = f(X)$ as the $n \times n$ matrix with

$$Y_{ij} = f(X_{ij}), \quad i, j = 1, \ldots, n$$

show that Y is positive semidefinite

an example is the polynomial kernel function $f(x) = (1 + x)^d$ (133A lecture 12)