LDL^T & Cholesky

This amounts to an LU decomposition of a positive-definite, symmetric matrix that's 2x as fast to compute!

Thm: A positive-definite symmetric matrix S can be uniquely decomposed as

$$S = LDL^T \qquad \text{and} \qquad S = \boxed{L_1 L_1^T} \longleftarrow \text{Cholesky}$$

where:

- D: diagonal w/ positive diagonal entries
- L: lower-unitriangular
- L_1: lower-triangular with positive diagonal entries.

Proof: [supplement]

NB: Any such L_1 has full column rank, so $S = L_1 L_1^T$ is necessarily positive-definite & symmetric (last time).

NB: Let $U = DL^T$.

(scales the rows of L^T by the diagonal entries of D)

Then U is upper-Δ with positive diagonal entries

$\Longrightarrow$ in REF, so $S = LU$ is the LU decomposition!

This tells us how to compute an LDL^T decomposition.

Procedure to compute $S=LDL^T$:

Let S be a symmetric matrix.

(1) Compute the LU decomposition $S=LU$.

→ If you have to do a row swap then stop: S is not positive-definite.

→ If the diagonal entries of U are not all positive then stop: S is not positive-definite.

(2) Let $D=$ the matrix of diagonal entries of U (set the off-diagonal entries $=0$). Then

$$S=LDL^T.$$

NB: An LDL^T decomposition can be computed in $\sim\frac{1}{3}n^3$ flops (as opposed to $2/3\,n^3$ for LU). This requires a slightly more clever algorithm. See the supplement – it's also faster by hand!

NB: This is still an LU decomposition – lets you solve $Sx=b$ quickly.

NB: $S=QDQ^T$ and $S=LDL^T$ are both "diagonalizations" in the sense of quadratic forms (later).

Eg: Find the LDL^T decomposition of

$$S = \begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -1 \\ -2 & -1 & 14 \end{pmatrix}$$

2-column method:	L	U
	$\begin{pmatrix} \; & \; & \; \\ \; & \; & \; \\ \; & \; & \; \end{pmatrix}$	$\begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -1 \\ -2 & -1 & 14 \end{pmatrix}$
$R_2 -= 2R_1$ $R_3 += R_1$	$\begin{pmatrix} 1 & \; & \; \\ 2 & \; & \; \\ -1 & \; & \; \end{pmatrix}$	$\begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 3 \\ 0 & 3 & 12 \end{pmatrix}$
$R_3 -= 3R_2$	$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 3 & 1 \end{pmatrix}$	$\begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 3 \\ 0 & 0 & 3 \end{pmatrix}$

So $S = LDL^T$ for

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 3 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Check:

$$DL^T = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 3 \\ 0 & 0 & 3 \end{pmatrix} = U \checkmark$$

Cholesky from LDL^T:

If S is positive-definite then $S = LDL^T$ where D is diagonal with positive diagonal entries.

If $D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$ set $\sqrt{D} = \begin{pmatrix} \sqrt{d_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{d_n} \end{pmatrix}$

Then $\sqrt{D} \cdot \sqrt{D} = D$ and $\sqrt{D}^T = \sqrt{D}$, so

$$LDL^T = L\sqrt{D}\sqrt{D}L^T = (L\sqrt{D})(L\sqrt{D})^T$$

So just set

$$\boxed{L_1 = L\sqrt{D} \rightsquigarrow S = L_1 L_1^T}$$

Strang:

"$S = A^TA$ is how a positive-definite symmetric matrix is put together.

$S = L_1 L_1^T$ is how you pull it apart"

Eg: $\begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -1 \\ -2 & -1 & 14 \end{pmatrix} = L_1 L_1^T$ for

$$L_1 = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 3 & 1 \end{pmatrix}\begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{1} & 0 \\ 0 & 0 & \sqrt{3} \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 2\sqrt{2} & 1 & 0 \\ -\sqrt{2} & 3 & \sqrt{3} \end{pmatrix}$$

Quadratic Optimization

This is an important application of the spectral theorem and positive-definiteness. Also, SVD+QO+ε-stats=PCA.

It is the simplest case of quadratic programming, which is a big subfield of optimization. (So is least squares.)

For an example application, see the Wikipedia page for support-vector machine, an important tool in machine learning that reduces to a quadratic optimization problem. (There are tons of other applications.)

Def: An optimization problem means finding extremal values (minimum & maximum) of a function $f(x_1, \dots, x_n)$ subject to some constraint on $(x_1, \dots, x_n)$.

In quadratic optimization, we consider quadratic functions.

Def: A quadratic form in n variables is a function $q(x_1, \dots, x_n) =$ sum of terms of the form $a_{ij} x_i x_j$

Eg: $q(x_1, x_2) = \frac{5}{2} x_1^2 + \frac{5}{2} x_2^2 - x_1 x_2$

Non-eg: $q(x_1, x_2) = x_1^2 + x_2^2 + x_1 + x_2$ is not a quadratic form: x_1, x_2 are linear terms.

NB: Thinking of $x=(x_1,\dots,x_n)$ as a vector,

$$q(cx)=q(cx_1,\dots,cx_n)=\sum' a_{ij}(cx_i)(cx_j)$$

$$=\sum' c^2 a_{ij}x_ix_j = c^2 q(x)$$

$$\boxed{q(cx)=c^2q(x)}$$

In quadratic optimization, the constraint on $x=(x_1,\dots,x_n)$ is usually $\|x\|=1$, ie $x_1^2+\cdots+x_n^2=1$.

Quadratic Optimization Problem:

Given a quadratic form $q(x)$, find the minimum & maximum values of $q(x)$ subject to $\|x\|=1$.

Eg: $q(x_1,x_2)=3x_1^2-2x_2^2$

Maximum:

$$q(x_1,x_2)=3x_1^2-2x_2^2\leq 3x_1^2+3x_2^2$$

$$=3(x_1^2+x_2^2)=3\|x\|^2=3$$

So the maximum value is 3; it is achieved at $(x_1,x_2)=\pm(1,0)$: $q(\pm1,0)=3$.

Minimum:

$$q(x_1,x_2) = 3x_1^2 - 2x_2^2 \geq -2x_1^2 - 2x_2^2$$
$$= -2(x_1^2+x_2^2) = -2\|x\|^2 = -2$$

So the minimum value is -2; it is achieved at $(x_1,x_2) = \pm(0,1)$: $q(0,\pm 1) = -2$.

This example is easy because $q(x_1,x_2) = 3x_1^2 - 2x_2^2$ involves only squares of the coordinates: there is no cross-term x_1x_2

Def: A quadratic form is diagonal if it has the form $q(x_1,\ldots,x_n) =$ sum of terms of the form $\lambda_i x_i^2$.

Terms of the form $a_{ij}x_ix_j$ $(i \neq j)$ are cross-terms.

Quadratic Optimization of Diagonal Forms:

Let $q(x) = \sum_i \lambda_i x_i^2$. Order the x_i so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then

- The maximum value of $q(x)$ is λ_1.
- The minimum value of $q(x)$ is λ_n.

(subject to $\|x\|=1$).

NB: the λ_i could be negative.

Strategy: To solve a quadratic optimization problem, we want to diagonalize it to get rid of the cross terms.

To do this, we use symmetric matrices!

Fact: Every quadratic form can be written
$$q(x) = x^T S x$$
for a symmetric matrix S.

Eg: $S = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{pmatrix}$

$$\leadsto x^T S x = (x_1 \ x_2 \ x_3) \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$= (x_1 \ x_2 \ x_3) \begin{pmatrix} x_1 + 2x_2 + 3x_3 \\ 2x_1 + 4x_2 + 5x_3 \\ 3x_1 + 5x_2 + 6x_3 \end{pmatrix}$$

$$= x_1^2 + 2x_1x_2 + 3x_1x_3$$
$$+ 2x_2x_1 + 4x_2^2 + 5x_2x_3$$
$$+ 3x_3x_1 + 5x_3x_2 + 6x_3^2$$

$$= x_1^2 + 4x_2^2 + 6x_3^2 + 4x_1x_2 + 6x_1x_3 + 10x_2x_3$$

NB: The $(1,2)$ and $(2,1)$ entries contribute to the x_1x_2 coefficient.

Given q, how to get S?

The x_i^2 coefficients go on the diagonal, and half of the x_ix_j coefficient goes in the (i,j) and (j,i) entries.

$$q(x_1,x_2,x_3) = a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2$$
$$+ a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3$$

$$\rightsquigarrow S = \begin{pmatrix} a_{11} & a_{12}/2 & a_{13}/2 \\ a_{12}/2 & a_{22} & a_{23}/2 \\ a_{13}/2 & a_{23}/2 & a_{33} \end{pmatrix}$$

NB: q is diagonal $\iff S$ is diagonal: the a_{ij} are the coefficients of the cross-terms.

$$x^T \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} x = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2$$

How does this help quadratic optimization?

Orthogonally diagonalize!

$$q(x) = x^T S x$$

Find a diagonal matrix D and orthogonal matrix Q such that $S = QDQ^T$

$$\rightsquigarrow q(x) = x^T Q D Q^T x$$

Let $\boxed{x = Qy}$: this is a change of variables

$$q(x) = q(Qy) = (Qy)^T QDQ^T (Qy)$$
$$= y^T \underbrace{Q^T Q}_{I_n} D \underbrace{Q^T Q}_{I_n} y = y^T D y$$

This is now diagonal!

NB: Q is orthogonal $\Rightarrow \|x\| = \|Qy\| = \|y\|$

So $\boxed{\|x\| = 1 \Leftrightarrow \|y\| = 1}$

Eg: Find the minimum & maximum of

$$q(x_1, x_2) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - 5x_1x_2$$ ← cross term ☹

subject to $\|x\| = 1$.

$$q(x) = x^T \begin{pmatrix} 1/2 & -5/2 \\ -5/2 & 1/2 \end{pmatrix} x \rightsquigarrow S = \frac{1}{2}\begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix}$$

Orthogonally diagonalize: $S = QDQ^T$ for

$$Q = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \quad D = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}$$

Set $x = Qy$:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} -y_1 + y_2 \\ y_1 + y_2 \end{pmatrix}$$

$$\begin{cases} x_1 = \frac{1}{\sqrt{2}}(-y_1 + y_2) \\ x_2 = \frac{1}{\sqrt{2}}(y_1 + y_2) \end{cases}$$ is a linear change of variables

Then $q(x) = y^T \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} y = 3y_1^2 - 2y_2^2$.

Check:

$$q(x) = q\left(\tfrac{1}{\sqrt{2}}(-y_1+y_2), \tfrac{1}{\sqrt{2}}(y_1+y_2)\right)$$

$$= \tfrac{1}{2}\cdot\tfrac{1}{2}(-y_1+y_2)^2 + \tfrac{1}{2}\cdot\tfrac{1}{2}(y_1+y_2)^2 - 5\cdot\tfrac{1}{2}(-y_1+y_2)(y_1+y_2)$$

$$= \tfrac{1}{4}y_1^2 + \tfrac{1}{4}y_2^2 - \cancel{\tfrac{1}{2}y_1y_2} + \tfrac{1}{4}y_1^2 + \tfrac{1}{4}y_2^2 + \cancel{\tfrac{1}{2}y_1y_2} + \tfrac{5}{2}y_1^2 - \tfrac{5}{2}y_2^2$$

$$= \left(\tfrac{1}{4}+\tfrac{1}{4}+\tfrac{5}{2}\right)y_1^2 + \left(\tfrac{1}{4}+\tfrac{1}{4}-\tfrac{5}{2}\right)y_2^2 = 3y_1^2 - 2y_2^2 \checkmark$$

The maximum value of q subject to $\|x\| = \|y\| = 1$ is 3, achieved at

$$y = (\pm 1, 0) \rightsquigarrow x = Qy = \pm\tfrac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\end{pmatrix}$$

The minimum value of q subject to $\|x\| = \|y\| = 1$ is -2, achieved at

$$y = (0, \pm 1) \rightsquigarrow x = Qy = \pm\tfrac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$$

NB: The minimum value is the smallest diagonal entry of $D \rightsquigarrow$ smallest eigenvalue.

$Q\begin{pmatrix}\pm 1\\0\end{pmatrix}$ is $\pm$ the first column of Q

$\rightsquigarrow$ is a unit eigenvector for that eigenvalue.

Likewise for the largest eigenvalue.

Quadratic Optimization:

To find the minimum/maximum of a quadratic form $q(x)$ subject to $\|x\|=1$:

(1) Write $q(x)=x^TSx$ for a symmetric matrix S

(2) Orthogonally diagonalize $S=QDQ^T$ for

$$Q=\underbrace{\begin{pmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{pmatrix}}_{\text{eigenvectors}} \qquad D=\underbrace{\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}}_{\text{eigenvalues}}$$

Order the eigenvalues so $\lambda_1 \geq \cdots \geq \lambda_n$

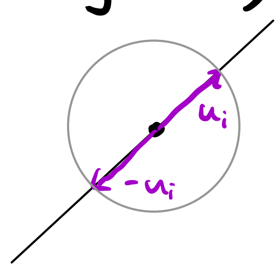
(3) The maximum value of $q(x)$ is the largest eigenvalue λ_1.

It is achieved for $x=$ any unit λ_1-eigenvector

The minimum value of $q(x)$ is the smallest eigenvalue λ_n.

It is achieved for $x=$ any unit λ_n-eigenvector.

NB: If $GM(\lambda_i)=1$ then the only unit λ_i-eigenvectors are $\pm u_i$. (only 2 unit vectors are on any line)



NB: $x=Qy$ diagonalizes q:

$$q(x)=\lambda_1 y_1^2+\cdots+\lambda_n y_n^2$$