1: intro

Sunday, September 6, 2020 2:18 PM

Chapter 1.7

We're going to start by reviewing one of the fundamental problems of linear algebra.

Let $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$, and assume A is invertible.
Solve for $Ax = b$.

(Note, also related to the optimization problem minimize $J(v)$ subject to $Av = b$, $v \in \mathbb{R}^n$)
(or the optimization problem minimize $\|Ax - b\|_2^2$ for $A \in \mathbb{R}^{m \times n}$, $m > n$)

Compute inverse $x = A^{-1} b$

Problem: equivalent to solving $Au^{(j)} = e_j$ for $j = 1, \dots, n$,

$$e_j = (0, \dots, 1, \dots, 0).$$

Cramer's rule

$$x_i = \frac{\det\left(\left[A_1 \ \cdots \ A_{i-1} \ b \ A_{i+1} \ \cdots \ A_n\right]\right)}{\det(A)}.$$

Problem: determinants are $O(n!)$ to compute

Gaussian elimination:

Suppose A is triangular, $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & \ddots & & \vdots \\ 0 & & \ddots & \\ & & & a_{nn} \end{bmatrix}$.

Then we can back-substitute

$$\begin{aligned} x_n &= a_{nn}^{-1} b_n \\ x_{n-1} &= a_{n-1,n-1}^{-1} (b_{n-1} - a_{n-1,n} x_n) \\ &\vdots \\ x_1 &= a_{11}^{-1} (b_1 - a_{12} x_2 - \cdots - a_{1,n} x_n) \end{aligned}$$

Recall that an upper triangular matrix is invertible iff $a_{ii} \neq 0$ $\forall i$.

We want to convert A to an upper triangular matrix by row operations

Let $A = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$, where r_i is the ith row.

1) Pick … $\neq 0$ // the …

Let $A = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$, where r_i is the ith row.

1) Pick some $a_{i1} \neq 0$, called the pivot π_1.
2) Permute rows i and 1 so that the pivot is $a_{11} = \pi_1$.
3) $r_i' = r_i - \frac{a_{i1}}{\pi_1}$ for $i = 2, \ldots, n$.
4) Repeat on the submatrix $A(2:n; 2:n)$.

$$\begin{bmatrix} \pi_1 & * \; * \; \cdots \; * \\ 0 & \\ \vdots & \boxed{A(2:n, 2:n)} \\ 0 & \end{bmatrix}$$

If we ever cannot pick a pivot, that implies a column is 0, contradicting invertibility. We can however still continue to an upper triangular matrix by going to the next submatrix; however, some pivot will be 0.

Thm 1.7.1 Let $A \in \mathbb{R}^{n \times n}$. Then $\exists$ invertible M s.t. $U = MA$ is upper triangular. The pivots are invertible iff A is invertible.

proof. Let $P(i,j) = I - e_{ii} - e_{jj} + e_{ij} + e_{ji}$, where e_{ij} is the matrix that is all 0's except at row i, col j.

Then left-multiplication by $P(i,j)$ swaps rows i and j.

Note $P(i,j)\,P(i,j) = I$.

Let $E_{i,j;\beta} = I + \beta e_{i,j}$.

Then left-multiplication by $E_{i,j;\beta}$ adds β times row j to row i.

Note $E_{i,j;\beta}\, E_{i,j;-\beta} = I$.

Thus, Gaussian elimination is equivalent to left-multiplying by some sequence of invertible matrices, proving the claim. $\square$

LU-factorization

Prop 1.7.1 Let $A \in \mathbb{R}^{n \times n}$ invertible. Then $A = LU$, where L is lower triangular with units on the diagonal and U is invertible upper triangular iff every matrix $A(1:k, 1:k)$ is invertible for $k = 1, \ldots, n$.

proof. If every $A(1:k, 1:k)$ is invertible, then Gaussian elimination doesn't need any pivoting, by the following argument.

Case 1: Note that $A(1:1, 1:1)$ is invertible, so $a_{11} \neq 0$, so we don't need to pivot for the first step.

Case 1: doesn't need any pivoting, by the following argument.

Note that $A(1{:}1, 1{:}1)$ is invertible, so $a_{11} \neq 0$, so we don't need to pivot in the first step.

Now assume no pivoting was needed in the first $k-1$ steps.

Then $MA = A_k$, where A_k is the matrix after $k-1$ rounds of Gaussian elim.,

and $M = E_\ell \cdots E_1$, where $E_1, \dots, E_\ell$ are $E_{i,j;\beta}$, $j<i$

$\Rightarrow M = E_1^{-1} \cdots E_\ell^{-1}$, but $E_{i,j;\beta}^{-1} = E_{i,j;-\beta}$ are unit lower triangular

$\Rightarrow M = L$ a lower triangular matrix.

Then $MA = A_k$ can be written $\begin{pmatrix} L_1 & 0 \\ L_3 & L_4 \end{pmatrix} \begin{pmatrix} A(1{:}k, 1{:}k) & A_2 \\ A_3 & A_4 \end{pmatrix} = \left(\begin{array}{c|c} U_1 & B_2 \\ \hline B_3 & B_4 \end{array}\right)$,

where L_1 and L_4 are unit lower triangular, U_1 is upper triangular, and B_3 has 0's in all but the last (kth) column by design.

But $U_1 = L_1 A(1{:}k, 1{:}k)$, and both L_1 and $A(1{:}k, 1{:}k)$ are invertible,

so U_1 is invertible, implying $(U_1)_{kk} \neq 0$.

Thus, we don't need to pivot in step k, completing induction.

So $MA = U$, after the entire Gaussian elimination, where

$M = E_\ell \cdots E_1$, where $E_1, \dots, E_\ell$ are of the form $E_{i,j;\beta}$, $j<i$.

Again, then $M = E_1^{-1} \cdots E_\ell^{-1}$, but $E_{i,j;\beta}^{-1} = E_{i,j;-\beta}$ are unit lower triangular.

$\Rightarrow A = LU$.

Case 2: Suppose $A = LU$. Then $A = \begin{bmatrix} A(1{:}k, 1{:}k) & A_2 \\ A_3 & A_4 \end{bmatrix} = \begin{bmatrix} L_1 & 0 \\ L_3 & L_4 \end{bmatrix} \begin{bmatrix} U_1 & U_2 \\ 0 & U_4 \end{bmatrix} = \begin{bmatrix} L_1 U_1 & L_1 U_2 \\ L_3 U_1 & L_3 U_2 + L_4 U_4 \end{bmatrix}$,

where L_1, L_4 are unit lower triangular and U_1, U_4 are upper triangular.

$\Rightarrow A(1{:}k, 1{:}k) = L_1 U_1$.

But U_1 must be invertible since U is invertible

$\Rightarrow A(1{:}k, 1{:}k)$ is invertible, $\forall k$.

Corollary 1.7.1: Let $A \in \mathbb{R}^{n \times n}$. If $A(1{:}k, 1{:}k)$ is invertible $\forall k$, then Gaussian elimination requires no pivoting and yields an LU factorization $A = LU$.

Why do we care about LU-factorizations?

If $A = LU$, then we can solve $Ax = b$ by

$Lw = b$ and $Ux = w$ using back-substitution.

Def 1.7.2 A has an LDU factorization if $A = LDU'$, where L is unit lower triangular,

Def 1.7.2 A has an LDU factorization if $A = LDU'$, where
L is unit lower triangular,
U is unit upper triangular, and
D is a diagonal matrix.

Note: If $A = LU$ is an LU-factorization, and
then $A = LDU'$, where $D = \text{diag}(U)$ and $U' = D^{-1}U$.

$PA = LU$ Factorization

Def. 1.7.3 A permutation matrix is a square matrix with a single 1 in every column and row.

Note: For any permutation matrix P, $P = \prod_{k=1}^{\ell} P(i_k, j_k)$, a product of transpositions.

Prop. 1.7.2 Let $A \in \mathbb{R}^{n \times n}$ invertible. Then $\exists$ permutation matrix P s.t.
$PA(1:k, 1:k)$ is invertible $\forall$ k.

proof. Base case of $n = 1$ is trivial. Proof by induction.

Since A is invertible, its columns are linearly ind.
$\Rightarrow$ the first $n-1$ columns are " "
Consider $A(1:n, 1:n-1)$, which has rank $n-1$.
Then there exists $P_1(A(1:n, 1:n-1))$ s.t. $(P_1 A(1:n, 1:n-1))(1:n-1, 1:n-1)$
is an invertible $(n-1)\times(n-1)$ matrix.

$\Rightarrow$ by induction hypothesis, we can permute that $(n-1)\times(n-1)$ matrix
so that submatrices are invertible.

We can extend that permutation to a permutation P_2 of the rows
of $P_1 A$ by leaving the nth row unchanged.

Then $P_2 P_1 A$ has the desired property. $\square$

Thm 1.7.2
(1) Given invertible $A \in \mathbb{R}^{n \times n}$, there exists a $PA = LU$ factorization.
(2) If $P = I$, then $A = LU$ is a unique factorization
$A = LDU'$ is a unique factorization.

(3) We can get $A = LU$ via Gaussian elimination if such a factorization exists.
(4) We can get $PA = LU$ via Gaussian elimination by applying the same row transpositions to I.

proof. See book.

Prop. 1.7.3 If $A = LU$ is an LU-factorization and A is real and symmetric, then $A = LDL^T$, where L is a lower-triangular unit diagonal matrix and D is the diagonal matrix of pivots.
Furthermore, the factorization is unique.

proof.

$$A = LDU$$
$$A^T = U^T D L$$
$$A = A^T \implies LDU = U^T D L^T$$
$$\implies L = U^T,\ U = L^T \quad \text{because of uniqueness of } LDU$$
$$\implies A = LDL^T \text{ is unique.} \quad \boxed{/\!/}$$

SPD matrices and Cholesky decomposition

Recall: A $n \times n$ real symmetric matrix is positive definite iff
$$x^T A x > 0 \quad \forall x \in \mathbb{R}^n,\ x \neq 0.$$

(Similarly: A $n \times n$ complex Hermitian matrix $(A^* = A)$ is positive definite iff
$$z^* A z > 0 \quad \forall z \in \mathbb{C}^n,\ z \neq 0 \ )$$

Properties of real pos. def. matrix A:

Let S^n_{++} be the set of symmetric positive definite matrices

1) A is invertible
2) $a_{ii} > 0$ for $i = 1, \dots, n$
3) $\forall$ real invertible Z, $Z^T A Z \in S^n_{++}$ iff $A \in S^n_{++}$
4) S^n_{++} is convex
 i.e. if $A, B \in S^n_{++}$, then $\forall\ 0 \le \lambda \le 1$, $(1-\lambda)A + \lambda B \in S^n_{++}$
(5) S^n_{++} is a cone;
 i.e. if $\lambda > 0$, and $A \in S^n_{++}$ then $\lambda A \in S^n_{++}$

Prop. 1.7.5 If $A \in S^n_{++}$, then $A(1:k, 1:k) \in S^n_{++}$ for $k = 1, \dots, n$.

proof. Let $w \in \mathbb{R}^k$. Let $x = \begin{bmatrix} w \\ \vec{0} \end{bmatrix} \in \mathbb{R}^n$.

Then $w^T A(1:k, 1:k) w = x^T A x > 0$. $\boxed{/\!/}$

Theorem 1.7.4 Let $A \in S^n_{++}$. Then $\exists$ real lower-triangular B s.t. $A = BB^T$.
Furthermore, B can be chosen so its diagonal elements are strictly positive.

Theorem 1.7.4 Let $A \in S_{++}$. Then $\exists$ real lower-triangular B s.t. $A = BB'$. Furthermore, B can be chosen so its diagonal elements are strictly positive.

proof. Induction on the dim n of A.

Base case: $n=1$. Then $a_{11} > 0$. Let $\alpha = \sqrt{a_{11}}$, $B = (\alpha)$.

Induction: $n \geq 2$. $a_{11} > 0$, so

$$A = \begin{pmatrix} a_{11} & W^T \\ W & C \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha & 0 \\ W/\alpha & I \end{pmatrix}}_{B_1} \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & \underbrace{C - WW^T/a_{11}}_{\text{Schur complement}} \end{pmatrix}}_{A_1} \underbrace{\begin{pmatrix} \alpha & W^T/\alpha \\ 0 & I \end{pmatrix}}_{B_1^T} \equiv B_1 A_1 B_1^T, \quad \alpha = \sqrt{a_{11}}.$$

Note A_1 is also symmetric positive definite because $A_1 = B_1^{-1} A (B_1^{-1})^T$.

$\Rightarrow C - \dfrac{WW^T}{a_{11}}$ is symm pos. def. (by restriction to vectors with $x_1 = 0$, $x \neq 0$.)

Thus, we can apply the induction hypo to the $(n-1)\times(n-1)$ $C - \dfrac{WW^T}{a_{11}}$.

$\Rightarrow C - WW^T/a_{11} = LL^T$, L unique with pos. diagonal entries

So, we get

$$A = \begin{pmatrix} \alpha & 0 \\ W/\alpha & I \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & LL^T \end{pmatrix} \begin{pmatrix} \alpha & W^T/\alpha \\ 0 & I \end{pmatrix}$$

$$= \begin{pmatrix} \alpha & 0 \\ W/\alpha & I \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & L \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & L^T \end{pmatrix} \begin{pmatrix} \alpha & W^T/\alpha \\ 0 & I \end{pmatrix}$$

$$= \begin{pmatrix} \alpha & 0 \\ W/\alpha & L \end{pmatrix} \begin{pmatrix} \alpha & W^T/\alpha \\ 0 & L^T \end{pmatrix}$$

Let $B = \begin{pmatrix} \alpha & 0 \\ W/\alpha & L \end{pmatrix}$, a unique lower-triangular matrix with pos. diagonal and $A = BB^T$. $\boxed{\diagup}$

Note: Cholesky is a special case of LU and uniqueness can be proven throug LU decompositions.

However, Cholesky requires half the number of operatos / space and is also numerically stable.

Prop. 7.6 The following are equivalent for a symmetric $n \times n$ matrix

(1) $A \in S_{++}^n$

(2) Sylvester's criterion All principal minors are positive. $(\det(A(1{:}k, 1{:}k)) > 0)$

(1) $A \in S^n_{++}$

(2) Sylvester's criterion. All principal minors are positive. $\left(\det(A(1{:}k, 1{:}k)) > 0\right)$

(3) A has an LU factorization and all pivots are positive.

(4) A has an LDL^T factorization and all pivots in D are positive.

Aside: We have reviewed primarily square matrices.
Applying an analogue of Gaussian elimination to rectangular matrices gives Reduced Row Echelon Form, which you should remember from undergrad.