Linear Algebra Notes

Eddie Qiao

bah

§1 Chapter 1 - Introduction to Vectors

§1.1 Vectors and Linear Combinations

Skipped

§1.2 Lengths and Dot Products

Skipped.

§1.3 Matrices

Skipped

§2 Chapter 2 - Solving Linear Equations

§2.1 Vectors and Linear Equations

Consider $A\mathbf{x} = \mathbf{b}$. There are two ways to view this. The column picture is that combinations of the n columns of A make $\mathbf{b}$. The row picture is that the m rows give m equations/planes that meet at $\mathbf{x}$.

§2.2 The Idea of Elimination

Standard elimination with systems of equations can be done on matrices by multiplying a row then subtracting it. Creates *pivots* along the diagonal, they are the nonzero in the row that does the elimination.

§2.3 Elimination Using Matrices

The elimination matrix is I along with a multiplier in a spot. These can all be combined into one matrix. The row exchange matrix is I with rows exchanged.

§2.4 Rules for Matrix Operations

Associativity exists. Commutativity does not. Also $AB = A$ times columns of $B =$ rows of A times $B =$ columns times rows. Matrices can be multiplied by blocks.

§2.5 Inverse Matrices

If A is invertible then:

- A has n nonzero pivots
- $\det A \neq 0$.
- $\mathbf{x} = \mathbf{0}$ is the only solution to $A\mathbf{x} = \mathbf{0}$.

$(AB)^{-1} = B^{-1}A^{-1}$. Gauss Jordan begins with $\begin{bmatrix} A & I \end{bmatrix}$ and becomes $\begin{bmatrix} I & A^{-1} \end{bmatrix}$. Each step is row exchange or elimination, so it's equivalent to multiplying the matrix by another matrix.

§2.6 Elimination = Factorization: $A = LU$

We can factor $A = LU$, where L is lower triangular and U is upper triangular. When going from A to U, with each elimination step, we add the multiplier to L in the right spot.

§2.7 Transposes and Permutations

The *transpose* swaps columns and rows. $(AB)^T = B^T A^T$ and $(A^{-1})^T = (A^T)^{-1}$. Dot product is now written as $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}$. A *symmetric matrix* has $S^T = S$, and an *orthogonal matrix* has $Q^T = Q^{-1}$. A *permutation matrix* is I with rows rearranged.

§3 Chapter 3 - Vector Spaces and Subspaces

§3.1 Spaces of Vectors

For a *vector space* S, if $\mathbf{v}, \mathbf{w} \in S$ then $c\mathbf{v} + d\mathbf{w} \in S$. A *subspace* of $\mathbb{R}^n$ is a vector space within $\mathbb{R}^n$. The *column space* $C(A)$ is all the combinations of columns of A, has all vectors $A\mathbf{x}$, $A\mathbf{x} = \mathbf{b}$ is solvable when $b \in C(A)$.

§3.2 The Nullspace of A: Solving $A\mathbf{x} = 0$ and $R\mathbf{x} = 0$

The *nullspace* $N(A)$ is the solutions $\mathbf{x}$ to $A\mathbf{x} = \mathbf{0}$. *Reduced row echelon form* R has all pivots $= 1$ with zeros around it. If column j of R is free (no pivot), there's a "special solution" to $A\mathbf{x} = \mathbf{0}$ with $x_j = 1$. *Rank* r is the number of pivots and the number of nonzero rows in R. When $m < n$ (wide), there are always nonzero solutions to $A\mathbf{x} = \mathbf{0}$.

§3.3 The Complete Solution to $A\mathbf{x} = \mathbf{b}$

The solution to $A\mathbf{x} = \mathbf{b}$ is

$$\mathbf{x} = (\text{one particular solution } \mathbf{x}_p) + (\text{any } x_n \text{ in the nullspace}).$$

Elimination goes from $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ to $\begin{bmatrix} R & \mathbf{d} \end{bmatrix}$. A solution exists when zero rows of R have zeros in $\mathbf{d}$. A has *full column rank* when $r = n$, so $N(A) = \mathbf{0}$ and there are no free variables. A has *full row rank* when $r = m$, so $C(A) = \mathbb{R}^m$ and $A\mathbf{x} = \mathbf{b}$ is always solvable.

We illustrate all of this through an example. Consider

$$A = \begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 6 \\ 7 \end{bmatrix}.$$

After augmenting $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ and doing elimination, we have

$$R = \begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix}.$$

For a solution to exist, zero rows in R must be zeros in $\mathbf{d}$. The particular solution sets the free variables to 0, so it's $\mathbf{x}_p = (1, 0, 6, 0)$. The $n - r$ special solutions solve $A\mathbf{x}_n = \mathbf{0}$. The complete solution is

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_n = \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 0 \\ -4 \\ 1 \end{bmatrix}.$$

§3.4 Independence, Basis, and Dimension

If $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are *independent*, then $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = 0 \implies c_i = 0$. If $m < n$ (wide), then it must have *dependent columns*. $\mathbf{v}_1, \ldots, \mathbf{v}_k$ *span* the space $S =$ all combinations of the $\mathbf{v}$'s. They are a *basis* for S if they're independent and span S. The *dimension* of a space S is the number of vectors in the basis.

§3.5 Dimensions of the Four Subspaces

Column space $C(A)$ and row space $C(A^T)$ have dimension r. Nullspace $N(A)$ has dimension $n - r$ and left nullspace $N(A^T)$ has dimension $m - r$.

§4 Chapter 4 - Orthogonality

§4.1 Orthogonality of the Four Subspaces

Vectors are *orthogonal* if $\mathbf{v}^T\mathbf{w} = 0$. Subspaces V and W are orthogonal if $\mathbf{v}^T\mathbf{w} = 0$ for all $v \in V, w \in W$. Row space and nullspace are orthogonal, column space and left nullspace are orthogonal. In fact, they're *orthogonal complements*: every $x \in \mathbb{R}^n$ can be written $\mathbf{x} = \mathbf{x}_{\text{row}} + \mathbf{x}_{\text{null}}$.

§4.2 Projections

The *projection* of $\mathbf{b}$ on a subspace S is the closest vector $\mathbf{p} \in S$, and $\mathbf{b} - \mathbf{p}$ is orthogonal to S. A^TA is invertible only if A has independent columns. Projection of $\mathbf{b}$ on the column space of A is

$$\mathbf{p} = A(A^TA)^{-1}A^T\mathbf{b}.$$

The *projection matrix* is

$$P = A(A^TA)^{-1}A^T,$$

so we have $\mathbf{p} = P\mathbf{b}$ and $P^2 = P = P^T$.

§4.3 Least Squares Approximation

Letting $\mathbf{p} = A\hat{\mathbf{x}}$, we can solve $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. When $A\mathbf{x} = \mathbf{b}$ has no solution, $\hat{\mathbf{x}}$ is the least-squares solution (minimize $\|\mathbf{b} - A\hat{\mathbf{x}}\|^2$). One can also produce this by setting the partial derivatives of the norm to zero. For fitting $(t_1, b_1), \ldots, (t_m, b_m)$, A has columns $(1, \ldots, 1)$ and $(t_1, \ldots, t_m)$. We then directly apply least-squares.

§4.4 Orthonormal Bases and Gram-Schmidt

Columns $q_1, \dots, q_n$ are orthonormal if $q_i^T q_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$, so $Q^T Q = I$. The *Gram-Schmidt* process takes independent $\mathbf{a}_i$ to orthonormal $\mathbf{q}_i$. Begin with $\mathbf{q}_1 = \mathbf{a_1} / \|a_1\|$, then for each remaining one, subtract its projections onto each previous one. Each a_i is a combination of q_1 to q_i, so $A = QR$ for orthogonal Q and triangular R.

§5 Chapter 5 - Determinants

§5.1 The Properties of Determinants

Determinants can be determined by three properties:

1. $\det I = 1$.
2. It changes signs when two rows are exchanged.
3. It's a linear function of each row (all other rows stay fixed).

This can be used to find a bunch of other properties. For example:

- A is singular iff $\det A = 0$
- $\det A$ is $\pm$(product of the pivots)
- $\det AB = \det A \cdot \det B$
- $\det A^T = \det A$

§5.2 Permutations and Cofactors

We have

$$\det A = \sum (\det P) a_{1\alpha} a_{2\beta} \dots a_{n\omega},$$

so it's the sum of all $n!$ column permutations P. The determinant is also the sum of cofactors

$$\det A = a_{i1} C_{i1} + \cdots + a_{in} C_{in},$$

where each cofactor is the determinant of a smaller matrix. This is essentially just Evan's expansion by minors.

§5.3 Cramer's Rule, Inverses, and Volumes

The key idea behind *Cramer's Rule* is that if $A\mathbf{x} = \mathbf{b}$, then

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{bmatrix} = B_1.$$

Taking determinants, we find $x_1 = \frac{\det B_1}{\det A}$, where B_1 is A with the 1st column replaced by $\mathbf{b}$. Creating equations for $AA^{-1} = I$, we find that

$$A^{-1} = \frac{C^T}{\det A},$$

where C is the *cofactor matrix*. There was also information about cross product and area, but I knew this.

§6 Chapter 6 - Eigenvalues and Eigenvectors

§6.1 Introduction to Eigenvalues

When $A\mathbf{x} = \lambda\mathbf{x}$, λ is the eigenvalue and $\mathbf{x}$ is the eigenvector.

$$\begin{aligned} \det A &= \lambda_1\lambda_2 \ldots \lambda_n \\ \text{trace} &= \lambda_1 + \lambda_2 + \cdots + \lambda_n = a_{11} + a_{22} + \cdots + a_{nn} \end{aligned}$$

§6.2 Diagonalizing a Matrix

$A\mathbf{x} = \lambda\mathbf{x}$ can be rewritten $AX = X\Lambda$, where X has the eigenvectors as columns and Λ has the eigenvalues along the diagonal. When there are n independent eigenvectors, $A = X\Lambda X^{-1}$. This factorization is nice when considering A^n. To solve the recursive relation $\mathbf{u}_{k+1} = A\mathbf{u}_k$, we have $\mathbf{u}_k = A^k\mathbf{u}_0$. After writing $\mathbf{u}_0 = c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n$, we find $\mathbf{u}_k = c_1(\lambda_1)^k\mathbf{x}_1 + \cdots + c_n(\lambda_n)^k\mathbf{x}_n$. If no eigenvalues are equal, then there are n independent eigenvalues and X is invertible. All $C = B^{-1}AB$ have the same eigenvalues as A, these matrices are *similar*.

§6.3 Systems of Differential Equations

We have the *matrix exponential*

$$e^{At} = I + At + \cdots + \frac{(At)^n}{n!} = Xe^{\Lambda t}X^{-1}.$$

The solutions to $\frac{d\mathbf{u}}{dt} = A\mathbf{u}$ are all $\mathbf{u}(t) = e^{\lambda t}\mathbf{x}$. If we're given $\mathbf{u}(0)$, write $\mathbf{u}(0) = c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n$ and

$$\mathbf{u}(t) = e^{At}\mathbf{u}(0) = Xe^{\Lambda t}X^{-1}\mathbf{u}(0) = c_1e^{\lambda_1 t}\mathbf{x}_1 + \cdots + c_ne^{\lambda_n t}\mathbf{x}_n.$$

When all eigenvalues have real part < 0, then A is *stable*, meaning $\mathbf{u}(t) \to 0$. For second-order systems, we convert it to first-order. For example, with $u'' + Bu' + Cu = 0$, we have

$$\begin{bmatrix} u \\ u' \end{bmatrix}' = \begin{bmatrix} 0 & 1 \\ -C & -B \end{bmatrix} \begin{bmatrix} u \\ u' \end{bmatrix}.$$

§6.4 Symmetric Matrices

Symmetric matrices S have n real eigenvalues and n orthogonal eigenvectors. Therefore, it can be diagonalized $S = Q\Lambda Q^{-1} = Q\Lambda Q^T$. The number of positive eigenvalues of S is equal to the number of positive pivots. For antisymmetric matrices $A = -A^T$, the λ's are imaginary and $\mathbf{q}$'s are orthonormal.

§6.5 Positive Definite Matrices

If a symmetric matrix S has one of these five, it has them all:

1. All n pivots are positive
2. All n upper left determinants are positive
3. All n eigenvalues are positive
4. $\mathbf{x}^TS\mathbf{x}$ is positive for $\mathbf{x} \neq \mathbf{0}$ (energy)

5. $S = A^T A$ for some A with independent columns

These are *positive definite* matrices. *Positive semidefinite* allows for those in 1-4 to be 0. $\mathbf{x}^T S\mathbf{x} = 1$ gives an ellipse in $\mathbb{R}^n$ when S is symmetric positive definite.

§7 Chapter 7 - The Singular Value Decomposition (SVD)

§7.1 Image Processing by Linear Algebra

I don't know what Gil Strang was thinking when he put this section before 7.2.

§7.2 Bases and Matrices in the SVD

Ok this was the hardest chapter by far (so far).

§7.2.1 Overview of SVD

The goal of *Singular Value Decomposition (SVD)* is to produce "good" bases for the four fundamental subspaces. In these notes, we first describe SVD fully then show how it can be constructed. When performing SVD, we write $A = U\Sigma V^T$, where U and V are orthonormal matrices. In particular,

$$\begin{aligned}
&\mathbf{u}_1, \ldots, \mathbf{u}_r \text{ is an OB for the column space}\\
&\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m \text{ is an OB for the left nullspace}\\
&\mathbf{v}_1, \ldots, \mathbf{v}_r \text{ is an OB for the row space}\\
&\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \text{ is an OB for the nullspace}
\end{aligned}$$

Here, we have

$$\begin{aligned}
A\mathbf{v}_1 &= \sigma_1\mathbf{u}_1\\
&\vdots\\
A\mathbf{v}_r &= \sigma_r\mathbf{u}_r.
\end{aligned}$$

This means

$$A = U\Sigma V^T = \mathbf{u}_1\sigma_1\mathbf{v}_1^T + \cdots + \mathbf{u}_r\sigma_r\mathbf{v}_r^T.$$

Conventionally, $\sigma \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$. This is the idea behind image compression in Section 7.1—one can reduce a complex image down to the sum of rank-1 components, which are significantly easier to store. An aspect that makes SVD useful is that the singular values are stable, unlike the eigenvalues.

§7.2.2 Constructing an SVD — All at Once

As for how to construct an SVD, we choose the $\mathbf{v}$'s to be orthonormal eigenvectors of $A^T A$. To recover $\mathbf{u}_1$ to $\mathbf{u}_r$, we use $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$. Note that

$$\mathbf{u}_i^T\mathbf{u}_j = \left(\frac{A\mathbf{v}_i}{\sigma_i}\right)^T\left(\frac{A\mathbf{v}_j}{\sigma_j}\right) = \frac{\mathbf{v}_i^T A^T A\mathbf{v}_j}{\sigma_i\sigma_j} = \frac{\sigma_j^2}{\sigma_i\sigma_j}\mathbf{v}_i^T\mathbf{v}_j = 0.$$

Note that the $\mathbf{u}$'s are also the eigenvectors of AA^T. To complete the $\mathbf{v}$'s and $\mathbf{u}$'s, we can take any orthonormal bases of $N(A)$ and $N(A^T)$.

§7.2.3 Constructing an SVD — One by One

We present another way of constructing an SVD in the order $\sigma_1, \dots, \sigma_r$. Our motivation for doing this is twofold:

1. We need two orthnormal eigenvectors for a double eigenvalue of S. Previously, we didn't know how to recover these.

2. We want a way to select the largest term of the SVD.

First choose

$$\lambda_1 = \max \frac{\mathbf{x}^T S\mathbf{x}}{\mathbf{x}^T\mathbf{x}}, \qquad \sigma_1 = \max \frac{\|A\mathbf{x}\|}{\|x\|}.$$

For the first, the winning vector is $\mathbf{x} = \mathbf{q}_1$ with $S\mathbf{q}_1 = \lambda_1\mathbf{q}_1$. For the second, the wining vector is $\mathbf{x} = \mathbf{v}_1$ with $A\mathbf{v}_1 = \sigma_1\mathbf{u}_1$. To construct the rest, we perform the same procedure with the restriction that $\mathbf{q}_1^T\mathbf{x} = 0$ for the first and $\mathbf{v}_1^T\mathbf{x} = 0$ for the second. We prove this works. This first section is not covered in Strang, but due to this lecture.

Since S is symmetric, let $S = Q\Lambda Q^T$, so we have

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T(Q\Lambda Q^T)\mathbf{x} = (\mathbf{x}^T Q)\Lambda(\mathbf{x}^T Q)^T = \mathbf{y}^T\Lambda\mathbf{y}$$

for $\mathbf{y} = Q^T\mathbf{x}$. Note that $\|\mathbf{y}\|^2 = \mathbf{y}^T\mathbf{y} = (Q^T\mathbf{x})^T(Q^T\mathbf{x}) = 1$. Thus, we have a new expression to maximize. Note that

$$\mathbf{y}^T\Lambda\mathbf{y} = \sum_{i=1}^n \lambda_i y_i^2.$$

Since $\lambda_1 \ge \cdots \ge \lambda_r$, this is maximized when $y_1 = \pm 1$, $y_2 = \cdots = y_r = 0$. The corresponding $\mathbf{x}$ is $\mathbf{x} = Q(\pm\mathbf{e}_1) = \pm\mathbf{q}_1$. Note that To choose σ_1, note that

$$\left(\frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}\right)^2 = \frac{\mathbf{x}^T A^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \frac{\mathbf{x}^T(A^T A)\mathbf{x}}{\mathbf{x}^T\mathbf{x}}.$$

Since A^TA is a symmetric matrix, σ_1 is the largest eigenvalue of A^TA, matching our observations in Section 7.2.2. We now see how this process works to choose the remaining values. Consider an orthogonal matrix Q_1 with $\mathbf{q}_1$ in the first column. Using $S\mathbf{q}_1 = \lambda_1\mathbf{q}_1$, we have

$$SQ_1 = S\begin{bmatrix}\mathbf{q}_1 & \mathbf{q_2}\dots\mathbf{q}_n\end{bmatrix} = \begin{bmatrix}\mathbf{q}_1 & \mathbf{q_2}\dots\mathbf{q}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & \mathbf{w}^T \\ \mathbf{0} & S_{n-1}\end{bmatrix} = Q_1\begin{bmatrix}\lambda_1 & \mathbf{w}^T \\ \mathbf{0} & S_{n-1}\end{bmatrix}.$$

All we need is the first column to be $S\mathbf{q}_1 = \lambda_1\mathbf{q}_1$, giving us that matrix. Multiplying by Q_1^T, we find $Q_1^TSQ_1 = \begin{bmatrix}\lambda_1 & \mathbf{w}^T \\ \mathbf{0} & S_{n-1}\end{bmatrix}$. Since the left is symmetric, we find $\mathbf{w} = \mathbf{0}$ and $S_{n-1} = S_{n-1}^T$. Thus, we've reduced the maximization problem to size $n-1$, and the restriction is gone. We now find the largest eigenvalue of S_{n-1}. Continuing this, we recover all the eigenvectors and eigenvalues.

§7.2.4 Computing Eigenvalues of S and Singular Values of A

This section concerns how these values are actually computed. We know the singular values σ_i of A are the square roots of the eigenvalues λ_i of A^TA. Recall that similar matrices have the same λ's. I claim that σ's of A and $Q_1^TAQ_2$ are the same. Note

$$(Q_1^TAQ_2)^TQ_1^TAQ_2 = Q_2^TA^TAQ_2,$$

which has the same eigenvectors as A^TA. This allows us to reach a bidiagonal matrix $Q_1^TAQ_2$. Going to a diagonal matrix requires more perspicacity that will (maybe) be covered later.