Recitation 15

All Hallows' Eve, 2024

1 Recap

1.1 Positive definite/semidefinite matrices

Positive-definite (and semidefinite) matrices

(We'll stick with real matrices, though this is easily generalized to complex-Hermitian matrices.)

A **symmetric positive-definite (SPD)** or **positive-semidefinite** (PSD) matrix $A = A^T$ satisfies:

1. $A = B^T B$ for some matrix B. If A is SPD (not just semidefinite), then B must be full column rank, i.e. $N(B) = \{\vec{0}\}$.

2. The **quadratic form** $x^T Ax$ is ≥ 0 for all x. If A is SPD, then it is > 0 for all $x \neq 0$.

3. The eigenvalues λ of A are all ≥ 0. If A is SPD, then all $\lambda > 0$.

4. If you do Gaussian elimination $A \to U$, all of the diagonal elements of U are ≥ 0. If A is SPD, all the diagonal elements (pivots) are > 0.

All three of these properties are equivalent: if any is true, then all of them are true. Property (4) was *not* shown in class, but is the quickest way to check positive-definiteness. (It also leads to an important variant of Gaussian elimination for SPD matrices, called "Cholesky factorization," that we will not cover in 18.C06.)

1.2 Connection to the SVD

For any matrix B with (full) SVD $B = U\Sigma V^T$, then we immediately see that $A = B^T B = V\Sigma^T \Sigma V^T = Q\Lambda Q^T$ has (real, ≥ 0) eigenvalues $\sigma_1^2, \sigma_2^2, \ldots, \sigma_r^2, 0, 0, \ldots$ and (orthonormal) eigenvectors $Q = V$. So, one strategy to find the SVD of any matrix B (albeit not the most efficient/robust approach) is to diagonalize and obtain the singular values from the square roots of the positive eigenvalues, the right singular vectors V from the eigenvectors, and the left singular vectors U from $Bv_i = \sigma_i u_i$.

1.3 Connection to quadratic programming (QP)

Suppose that A is an SPD matrix. Then it follows that the function $f(x) = x^T Ax - 2b^T x +$ (scalar), for any vector b and any scalar constant, is a **convex quadratic function** of x. It has a unique minimum at some $x_\star$, and finding this minimum is known as an (unconstrained) **quadratic programming (QP) problem**.

"Convexity" means informally that the function is an "upward bowl", so that it has a unique minimum (not a maximum, or multiple local minima, or saddle points). Formally, a **convex function** means that the function is $\leq$ the line segment connecting any two points. (If the function is $<$ the line segment, it is "strictly" convex.) Convexity is an extremely useful property to have in optimization because it makes finding the global minimum straightforward, with many nice algorithms and properties. (If you flip the sign, you get a "concave" function which is good for maximizing.)

Since A is SPD, we can write $A = B^T B$ from above, and a little algebra shows that minimizing $f(x)$ is exactly equivalent to minimizing $\|Bx - c\|^2$ where $b = B^T c$. That is, this is exactly equivalent to an ordinary least-squares problem, which we saw was minimized by solving the normal equations $Ax = B^T B x_\star = B^T c = b$. Hence:

- Minimizing a convex quadratic function $f(x) = x^T Ax - 2b^T x$ (plus any constant, or multiplied by any positive scalar) is exactly equivalent to solving $Ax = b$!

Another way to derive this is to compute the gradient of $f(x)$ and set it equal to zero. We can do this 18.02-style, component by component, to find: $\nabla f = 2(Ax - b)$. Thus, setting $\nabla f = \vec{0}$ again gives $Ax = b$.

The gradient is also important because it is the "uphill" direction, also called the steepest-ascent direction. Hence, $-\nabla f$, which in this case is twice $b - Ax$, is the "downhill" or steepest-descent direction. This leads to a class of algorithms to solve $Ax = b$ / quadratic minimization by "going downhill" called a **steepest-descent / gradient-descent**: take a sequence of "steps" in the downhill direction $b?Ax$.

2 Exercises

1. *Without* using the SVD, show that for any $m \times n$ real matrix A of rank r, the two semidefinite matrices AA^T and A^TA share the same positive eigenvalues $\lambda > 0$ (but have a different number of zero eigenvalues in general). In particular, suppose that v_k is one of the orthonormal (since it is SPD) eigenvectors of A^TA with a positive eigenvalue $\lambda_k = \sigma_k^2 > 0$, *defining* $\sigma_k = \sqrt{\lambda_k}$ (not assuming it is a singular value). Show that, $y_k = Av_k$ is an eigenvector of AA^T with the *same* eigenvalue σ_k^2, and if you normalize $u_k = y_k/\|y_k\|$ you get the relationship $Av_k = \sigma_k u_k$. This implies the existence of the SVD!

2. Suppose that you have a quadratic form $f(x) = x^TAx$ where A is an $m \times m$ matrix, not necessarily symmetric. Show that you can always replace A with its "symmetric part" $\frac{A+A^T}{2}$ and get the same function, i.e. $x^TAx = x^T\left(\frac{A+A^T}{2}\right)x$ for all x. (So, whenever we have a quadratic form, we can treat the matrix as symmetric without loss of generality.)

3. Indicate whether the following statements are True or False. It might be easier to think within the framework of symmetric matrices.
   (a) A non-negative matrix (i.e. a matrix with entries ≥ 0) is always positive semidefinite.
   (b) A positive semidefinite matrix is always non-negative.
   (c) A positive definite matrix is always invertible.

4. Consider the quadratic function defined as follows:
$$f(x, y) = 7x^2 + 6y + 8xy + 13y^2 + 3x$$
   (a) Express the given function in matrix–vector form $w^TAw + b^Tw$.
   (b) Determine the minimum value of f.
   (c) Does the function f has a unique minimum?

5. Show the following properties of positive semidefinite matrices.
   (a) The sum of two positive semidefinite matrices is also positive semidefinite.
   (b) Let $A \in \mathbb{R}^{m\times m}$ be a positive semidefinite matrix and $B \in \mathbb{R}^{m\times n}$ be another matrix. Show that B^TAB is also positive semidefinite.

3 Solutions

1. We are given $A^T A v_k = \sigma_k^2 v_k$ where $\sigma_k^2 > 0$. Let $y_k = Av_k$ as suggested, then $AA^T y_k = A(A^T A)v_k = \sigma_k^2 A v_k = \sigma_k^2 y_k$, so y_k is indeed an eigenvector of AA^T with the same eigenvalue σ_k^2 ... as long as the vector is $\neq 0$, which we can check by computing the length. Its length is $\|y_k\| = \sqrt{y_k^T y_k} = \sqrt{v_k^T A^T A v_k} = \sigma_k \|v_k\| = \sigma_k > 0$ (v_k has length 1 since it was given to be part of an orthonormal set), so indeed $y_k \neq 0$. Furthermore, if we normalize it to $u_k = y_k / \|y_k\| = y_k/\sigma_k = Av_k/\sigma_k$, then we get the SVD relationship $Av_k = \sigma_k u_k$.

2. $x^T \left(\frac{A+A^T}{2}\right) x = \frac{x^T Ax + x^T A^T x}{2}$, but since $x^T A^T x$ is a scalar it equals its own transpose: $x^T A^T x = (x^T A^T x)^T = x^T A x$. This combines with the other $x^T A x$ term and cancels the 2 in the denominator, so we get $x^T A x = x^T \left(\frac{A+A^T}{2}\right) x$ as desired.

3. (a) False. Consider the following counterexample: $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, which is a non-negative matrix. However, the product of the eigenvalues is $det(A) = -1$, so one of them must be negative. (In fact, the eigenvalues are ± 1.) Therefore, by property (3), A is not PSD.

   (b) False. Consider the counterexample $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$. This matrix has eigenvalues 1 and 3, so by property (3) it is SPD. However, it has negative elements.

   (c) True. If it is SPD, all the eigenvalues are positive, but a singular matrix has a nontrivial nullspace corresponding to an eigenvalue $\lambda = 0$, so it cannot be singular. Alternatively, the determinant is equal to the product of eigenvalues, which is positive (i.e., not zero), which implies that the matrix is invertible.

4. (a) The given quadratic function can be expressed as $w^T A w + b^T w$, where $w = \begin{bmatrix} x \\ y \end{bmatrix}$, A is a symmetric matrix, and b is a coefficient vector.

$$A = \begin{bmatrix} 7 & 4 \\ 4 & 13 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

   (b) There are a couple of ways to approach this.

   - The easiest method is probably to take the gradient and set it equal to zero: $\nabla_w (w^T A w + b^T w) = 2Aw + b = 0 \implies w = A^{-1}(-b/2)$. (This is slightly different from what was done in class, because b is defined differently here — you can't plug in a formula blindly just because the letters match!). That is, we need to solve $Aw = -b/2$, which we can do by Gaussian elimination. You can use Julia, or we can do it by hand:

$$\underbrace{\left[\begin{array}{cc|c} \boxed{7} & 4 & -\frac{3}{2} \\ 4 & 13 & -3 \end{array}\right]}_{[\,A\,|\,-b/2\,]} \xrightarrow{r_2 - \frac{4}{7} r_1} \underbrace{\left[\begin{array}{cc|c} \boxed{7} & 4 & -\frac{3}{2} \\ 0 & \boxed{\frac{75}{7}} & -\frac{15}{7} \end{array}\right]}_{[\,U\,|\,c\,]},$$

     Notice that since it has both pivots > 0, our matrix A is **positive definite**. Now, solving $Uw = c$ by backsubstitution, we get $w_2 = (-15/7)/(75/7) =$

$-1/5$, and $w_1 = [(-3/2) - 4(-1/5)]/7 = -1/10$. That is, the minimum occurs at $\boxed{w = \begin{pmatrix} -1/10 \\ -1/5 \end{pmatrix}}$. Plugging this into $w^T Aw + b^T w$, we find that the minimum *value* is $\boxed{-47/100}$.

- Alternatively, we can utilize the eigendecomposition of the matrix A. Recall that a symmetric A can be decomposed as $A = Q\Lambda Q^T$, where Q is an orthogonal matrix of eigenvectors and Λ is a diagonal matrix of eigenvalues. If we go through the usual process of solving for eigenvalues and eigenvectors, we find:

$$Q = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}, \Lambda = \begin{bmatrix} 15 & 0 \\ 0 & 5 \end{bmatrix}$$

Note that the eigenvalues are both positive, so we again find that A is positive definite. Now, let $v = Q^T w$, then we can rewrite the quadratic function as

$$\begin{aligned} f(x, y) &= w^T Aw + b^T w \\ &= (w^T Q)\Lambda(Q^T w) + b^T w \\ &= v^T \Lambda v + b^T(Qv) \\ &= v^T \Lambda v + (Q^T b)^T v \\ &= v^T \begin{bmatrix} 15 & 0 \\ 0 & 5 \end{bmatrix} v + \sqrt{5} \begin{bmatrix} 3 & 0 \end{bmatrix} v \\ &= 15v_1^2 + 5v_2^2 + 3\sqrt{5}v_1 \end{aligned}$$

Therefore, $f(x, y)$ is easy to solve for since it can be written as a sum of quadratic functions in a single variables. Thus, the minimum value is obtained when $v_1 = \frac{-3\sqrt{5}}{30} = \frac{-1}{2\sqrt{5}}, v_2 = 0$, which means that $w = Qv = \begin{bmatrix} -1/10 \\ -1/5 \end{bmatrix}$, the same as above.

Note that the approach with diagonalization is equivalent to re-writing f as the following:

$$\begin{aligned} f(x, y) &= 7x^2 + 6y + 8xy + 13y^2 + 3x \\ &= 3(x + 2y)^2 + (-2x + y)^2 + 3(x + 2y) \end{aligned}$$

Then we defined v_1 to be a scalar multiple of $(x+2y)$ and v_2 to be a scalar multiple of $(-2x + y)$.

(c) The function f has a unique minimum because the matrix A is positive definite (as shown in the previous part), and the minimum value occurs at a unique point. The positive definiteness of A guarantees that f is strictly convex, and therefore, it has a unique global minimum.

5. (a) Let A and B be two $m \times m$ positive semi-definite matrices. For any vector $x \in \mathbb{R}^m$, we have

$$x^T(A + B)x = x^T Ax + x^T Bx \geq 0 + 0 = 0.$$

This means $A + B$ is also positive semi-definite, by property (1).

(b) Let x be a vector in $\mathbb{R}^n$. Notice that Bx is a vector in $\mathbb{R}^m$. Since A is PSD, it follows that $x^T(B^TAB)x = (Bx)^TA(Bx) \geq 0$. This means B^TAB is also PSD, by property (1).