Math 484: Nonlinear Programming[1] Mikhail Lavrov

Chapter 1, Lecture 4: Positive Definite Matrices

January 25, 2019 *University of Illinois at Urbana-Champaign*

1 Positive definite matrices and their cousins

Last time, we reduced the second-derivative test for analyzing a critical point to determining if a matrix is "positive semidefinite".

Here are the definitions. We say that a symmetric $n \times n$ matrix A is:

- positive semidefinite (written $A \succeq 0$) if $\mathbf{x}^\mathsf{T} A\mathbf{x} \geq 0$ for all $\mathbf{x}$, and
- positive definite (written $A \succ 0$) if $\mathbf{x}^\mathsf{T} A\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- negative semidefinite (written $A \preceq 0$) if $\mathbf{x}^\mathsf{T} A\mathbf{x} \leq 0$ for all $\mathbf{x}$, and
- negative definite (written $A \prec 0$) if $\mathbf{x}^\mathsf{T} A\mathbf{x} < 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- indefinite (not written in any particular way) if none of the above apply.

The expression $\mathbf{x}^\mathsf{T} A\mathbf{x}$ is a function of $\mathbf{x}$ called the quadratic form associated to A. (It's a quadratic form because it's made up of terms like x_i^2 and $x_i x_j$: quadratic terms in the components of $\mathbf{x}$.) When the conditions above are met, we can also call the quadratic form positive semidefinite, positive definite, etc.

We only make these definitions for a symmetric matrix A: one that satisfies $A^\mathsf{T} = A$. This isn't a problem for us because Hessian matrices (assuming that the second derivatives are continuous, which we do anyway) are symmetric. Also, every quadratic form $\mathbf{x}^\mathsf{T} A\mathbf{x}$ can be written down as $\mathbf{x}^\mathsf{T} B\mathbf{x}$ for some symmetric matrix B.

We know to classify a critical point of a function $f : \mathbb{R}^n \to \mathbb{R}$ as a global minimizer if the Hessian matrix of f (its matrix of second derivatives) is positive semidefinite everywhere, and as a global maximizer if the Hessian matrix is negative semidefinite everywhere. If the Hessian matrix is positive definite or negative definite, the minimizer or maximizer (respectively) is strict.

We don't yet know how to tell when a matrix has this property, so that's what we'll look at today.

1.1 Easy example: a diagonal matrix

Consider the function $f(x, y) = x^2 + 2y^2$. (Not a very exciting function.)

Setting the gradient $\nabla f(x, y) = (2x, 4y)$ to $\mathbf{0}$, we get $x = y = 0$, so $(0, 0)$ is the only critical point of f. What kind of critical point is it?

[1]This document comes from the Math 484 course webpage: https://faculty.math.illinois.edu/~mlavrov/courses/484-spring-2019.html

The Hessian matrix of f is

$$Hf(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}.$$

(It's constant; it does not depend on x and y.) What can we say about it?

Here is a proof that this matrix is positive definite. For an arbitrary $\mathbf{x} \in \mathbb{R}^2$, we have

$$\mathbf{x}^\mathsf{T} \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \mathbf{x} = 2x_1^2 + 4x_2^2$$

which is a sum of squares. We always have $x_1^2 \geq 0$ and $x_2^2 \geq 0$, so $2x_1^2 + 4x_2^2 \geq 0$. Moreover, the only way to get 0 is to set $x_1 = x_2 = 0$. So for all $\mathbf{x} \neq \mathbf{0}$, $2x_1^2 + 4x_2^2 > 0$.

So $Hf(x, y) \succ 0$ for all $(x, y) \in \mathbb{R}^2$, which means 0 is a strict global minimizer.

In general, it's easy to classify diagonal matrices. For a diagonal matrix

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix}$$

the quadratic form is just $d_1x_1^2 + d_2x_2^2 + \cdots + d_nx_n^2$ and so the signs of $d_1, \dots, d_n$ determine its behavior:

- If $d_1, \dots, d_n$ are all nonnegative, then $d_1x_1^2 + d_2x_2^2 + \cdots + d_nx_n^2$ must be nonnegative for any $\mathbf{x}$, so $D \succeq 0$: D is positive semidefinite.
- If, moreover, $d_1, \dots, d_n$ are all positive, then $d_1x_1^2 + d_2x_2^2 + \cdots + d_nx_n^2$ can only be 0 if $\mathbf{x} = \mathbf{0}$, so $D \succ 0$: D is positive definite.
- Similarly, if $d_1, \dots, d_n \leq 0$, then $D \preceq 0$, and if $d_1, \dots, d_n < 0$, then $D \prec 0$, by the same logic.
- D is indefinite if the signs of $d_1, \dots, d_n$ are mixed.

2 Using eigenvalues

In general, the matrix A will not be diagonal, so this test does not work immediately. But we can change to a different basis in which A *is* represented by a diagonal matrix. For this, we will have to review some facts about eigenvalues from linear algebra.

For an $n \times n$ matrix A, if a nonzero vector $\mathbf{x} \in \mathbb{R}^n$ satisfies $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar $\lambda \in \mathbb{R}$, we call λ an eigenvalue of A and $\mathbf{x}$ its associated eigenvector.

When A is symmetric, we are guaranteed good behavior from eigenvalues, summarized by the following result from linear algebra.

Theorem 2.1 (Spectral theorem for symmetric matrices)**.** *If A is an $n \times n$ symmetric matrix, then it can be factored as*

$$A = Q^\mathsf{T} \Lambda Q = Q^\mathsf{T} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} Q$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A and the columns of Q are the corresponding eigenvectors.[2]

Applying this theorem to the quadratic form $\mathbf{x}^\mathsf{T} A \mathbf{x}$, we get

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

so if we substitute $\mathbf{y} = Q\mathbf{x}$ (converting to a different basis), the quadratic form becomes diagonal:

$$\mathbf{x}^\mathsf{T} A \mathbf{x} = \mathbf{y}^\mathsf{T} \Lambda \mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2.$$

Now we can classify the matrix A by looking at the eigenvalues of A. This is summarized in the following theorem:

Theorem 2.2. *Let A be a symmetric $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$. Then:*

- $A \succeq 0$ *if* $\lambda_1, \lambda_2, \dots, \lambda_n \geq 0$.
- $A \succ 0$ *if* $\lambda_1, \lambda_2, \dots, \lambda_n > 0$.
- $A \preceq 0$ *if* $\lambda_1, \lambda_2, \dots, \lambda_n \leq 0$.
- $A \prec 0$ *if* $\lambda_1, \lambda_2, \dots, \lambda_n < 0$.
- *A is indefinite if it has both positive and negative eigenvalues.*

Unfortunately, this theorem isn't terribly easy to apply, because computing eigenvalues is annoying. The technique is as follows: suppose $\mathbf{x}$ is an eigenvector with eigenvalue λ. Then $A\mathbf{x} = \lambda\mathbf{x}$ means that $(A - \lambda I)\mathbf{x} = \mathbf{0}$, so $A - \lambda I$ is singular. We can find λ for which this happens by seeing when the equation

$$\det(A - \lambda I) = 0$$

is satisfied.

When A is an $n \times n$ matrix, setting $\det(A - \lambda I) = 0$ gives a degree n polynomial which we can solve to find the eigenvalues. This is hard to do when A is large, and impossible to do exactly for $n \geq 5$; computers, however, are good at finding the eigenvalues approximately.

For us, the eigenvalue test is useful theoretically, but in the next lecture, we will develop a way to test when $A \succ 0$ that involves less scary computation.

[2]More is true. In fact, the columns of Q can be chosen to be an orthonormal basis of $\mathbb{R}^n$, so that $Q^\mathsf{T} = Q^{-1}$. Additionally, the spectral theorem is needed to guarantee the existence of n real eigenvalues, which is not true for general $n \times n$ matrices. But this statement is all that we'll need.

3 Saddle points

Here is one immediate application of the eigenvalue test to the problem we're actually interested in: classifying critical points of functions.

Theorem 3.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function with continuous Hf, and let $\mathbf{x}^* \in \mathbb{R}^n$ be a critical point of f.*

If $Hf(\mathbf{x}^*)$ is indefinite, then $\mathbf{x}^*$ is neither a local minimizer nor a local maximizer: it is a saddle point of f.

Proof. Recall some facts from the previous lecture: if we define $\phi_{\mathbf{u}}$ to be a restriction of f to a line through $\mathbf{x}^*$,

$$\phi_{\mathbf{u}}(t) = f(\mathbf{x}^* + t\mathbf{u}),$$

then $\phi''_{\mathbf{u}}$ can be found using the Hessian matrix of f:

$$\phi''_{\mathbf{u}}(t) = \mathbf{u}^\mathsf{T} Hf(\mathbf{x}^* + t\mathbf{u})\mathbf{u}.$$

In particular, $\phi''_{\mathbf{u}}(0) = \mathbf{u}^\mathsf{T} Hf(\mathbf{x}^*)\mathbf{u}$.

Suppose that $Hf(\mathbf{x}^*)$ is indefinite: it has both positive and negative eigenvalues.

If we choose $\mathbf{u}$ to be an eigenvector of $Hf(\mathbf{x}^*)$ with eigenvalue $\lambda > 0$, then we have

$$\phi''_{\mathbf{u}}(0) = \mathbf{u}^\mathsf{T} Hf(\mathbf{x}^*)\mathbf{u} = \mathbf{u}^\mathsf{T}(\lambda\mathbf{u}) = \lambda\|\mathbf{u}\|^2 > 0.$$

Having $\phi''_{\mathbf{u}}(0) > 0$ tells us that 0 is a strict local minimizer of $\phi_{\mathbf{u}}$: in other words, $\mathbf{x}^*$ looks like a strict local minimizer of f when we look in the direction $\mathbf{u}$.

Now choose a different direction $\mathbf{v}$ which is also an eigenvector of $Hf(\mathbf{x}^*)$, but with eigenvalue $\mu < 0$. Then we have

$$\phi''_{\mathbf{v}}(0) = \mathbf{v}^\mathsf{T} Hf(\mathbf{x}^*)\mathbf{v} = \mathbf{v}^\mathsf{T}(\mu\mathbf{v}) = \mu\|\mathbf{v}\|^2 < 0.$$

This tells us that along the direction $\mathbf{v}$, the critical point $\mathbf{x}^*$ doesn't look like a strict local minimizer, but rather a strict local maximizer!

This means that $\mathbf{x}^*$ isn't either kind of point. Arbitrarily close to $\mathbf{x}^*$, we have both points with larger f-value (when looking in the direction of $\mathbf{u}$) and points with smaller f-value (when looking in the direction of $\mathbf{v}$). $\square$

We call such points saddle points because the center of a saddle has this property: in one direction, the saddle curves up, and in another direction, the saddle curves down.