Math 484: Nonlinear Programming[1]	Mikhail Lavrov

Chapter 1, Lecture 5: Sylvester's criterion

January 28, 2019 — *University of Illinois at Urbana-Champaign*

1 Sylvester's criterion

Given an $n \times n$ symmetric matrix A with (i, j)-th entry $a_{ij} = a_{ji}$, let $A^{(k)}$ denote the $k \times k$ submatrix taken from the top left corner of A. That is,

$$A^{(k)} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix}.$$

In particular, $A^{(1)} = [a_{11}]$ and $A^{(n)} = A$.

Let $\Delta_k = \det(A^{(k)})$. (So $\Delta_n = \det(A)$.)

We're motivated to look at the determinant of A based on the eigenvalue test. Since

$$\det(A - xI) = (\lambda_1 - x)(\lambda_2 - x) \cdots (\lambda_n - x),$$

by setting $x = 0$ we get $\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$. When $A \succ 0$, all the eigenvalues are positive, so $\det(A) > 0$ as well.

We're motivated to look at the $k \times k$ submatrices for a different reason. Suppose we're looking at the quadratic form $\mathbf{u}^\mathsf{T} A \mathbf{u}$. If we set $u_{k+1} = u_{k+2} = \cdots = u_n = 0$, then the quadratic form for A simplifies to the quadratic form for $A^{(k)}$. Therefore we expect $A^{(k)} \succ 0$ as well, which means we must have $\Delta_k > 0$ for each k.

Sylvester's criterion says that actually, this characterizes positive definite matrices:

Theorem 1.1 (Sylvester's criterion)**.** *Let A be an $n \times n$ symmetric matrix. Then:*

- *$A \succ 0$ if and only if $\Delta_1 > 0, \Delta_2 > 0, \ldots, \Delta_n > 0$.*
- *$A \prec 0$ if and only if $(-1)^1 \Delta_1 > 0, (-1)^2 \Delta_2 > 0, \ldots, (-1)^n \Delta_n > 0$.*
- *A is indefinite if the first Δ_k that breaks both patterns is the wrong sign.*
- *Sylvester's criterion is inconclusive (A can be positive or negative semidefinite, or indefinite) if the first Δ_k that breaks both patterns is 0.*

Proof. We prove that having $\Delta_1, \ldots, \Delta_n > 0$ guarantees $A \succ 0$ by induction on n. For a 1×1 matrix A, we have $A \succ 0 \iff a_{11} > 0 \iff \Delta_1 > 0$, which is exactly Sylvester's criterion.

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

Assume that Sylvester's criterion works for $(n-1)\times(n-1)$ matrices, and that A is an $n\times n$ matrix with $\Delta_1,\dots,\Delta_n > 0$. By the induction hypothesis, $A^{(n-1)}$ is positive definite.

First, we show that A has at most one negative eigenvalue. Suppose not: suppose that A has two negative eigenvalues. Then their eigenvectors $\mathbf{u},\mathbf{v}$ are two independent vectors with $\mathbf{u}^\mathsf{T}A\mathbf{u} < 0$ and $\mathbf{v}^\mathsf{T}A\mathbf{v} < 0$. The spectral theorem means that they're orthogonal: $\mathbf{u}\cdot\mathbf{v} = 0$. Let $\mathbf{w} = v_n\mathbf{u} - u_n\mathbf{v}$.

On the one hand, $w_n = 0$, so $\mathbf{w}^\mathsf{T}A\mathbf{w}$ is plugging in $(w_1, w_2, \dots, w_{n-1})$ into $A^{(n-1)}$'s quadratic form. Since $A^{(n-1)} \succ 0$, we must have $\mathbf{w}^\mathsf{T}A\mathbf{w} > 0$. On the other hand, $\mathbf{w}^\mathsf{T}A\mathbf{w} = (v_n\mathbf{u} - u_n\mathbf{v})^\mathsf{T}A(v_n\mathbf{u} - u_n\mathbf{v})$ simplifies to $v_n^2(\mathbf{u}^\mathsf{T}A\mathbf{u}) + u_n^2(\mathbf{v}^\mathsf{T}A\mathbf{v})$, so we must have $\mathbf{w}^\mathsf{T}A\mathbf{w} < 0$. This is a contradiction, so having two negative eigenvalues is impossible.

Having just one negative eigenvalue is also impossible. Then the product $\lambda_1\lambda_2\cdots\lambda_n$ would be a product of $n-1$ positive values and 1 negative value, so $\det(A) = \lambda_1\lambda_2\cdots\lambda_n < 0$. But we know that $\det(A) = \Delta_n > 0$. Similarly, A can't have any zero eigenvalues, since $\Delta_n \neq 0$.

Therefore all the eigenvalues are positive, which means $A \succ 0$ by the eigenvalue test.

The condition for $A \prec 0$ is proven by applying Sylvester's criterion to $-A$. Multiplying every entry in A by -1 has weird effects on determinants. A determinant flips sign if you change the sign of just one of the rows, so flipping all k rows of $A^{(k)}$ multiplies Δ_k by $(-1)^k$. Therefore Sylvester's criterion for negative definite matrices asks for $\Delta_1, \Delta_2, \dots$ to alternate signs, starting from negative.

Suppose Sylvester's criterion fails because Δ_k has the wrong sign: for example, $\Delta_1,\dots,\Delta_{k-1} > 0$, but $\Delta_k < 0$. By looking carefully at our inductive proof, we see that this means that $A^{(k)}$ has exactly one negative eigenvalue. This makes it indefinite, and therefore A is indefinite as well.

However, if Sylvester's criterion fails because $\Delta_k = 0$, that means that one of $A^{(k)}$'s eigenvalues was 0. This rules out being positive or negative definite, but the matrix *might* be semidefinite. It also might not be: later on, we might encounter eigenvalues that are the wrong sign, but we'll never know, because we can't continue the proof. In this case, Sylvester's criterion is inconclusive. $\square$

In theory, some extra work lets us test if a matrix is positive semidefinite. The criterion there, which we're not going to prove, is the following.

If I is a subset of $\{1, 2, \dots, n\}$, let $A^{(I)}$ be the matrix obtained by taking rows and columns indexed by elements of i. For example, if $I = \{1, 3\}$, then

$$A^{(I)} = \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix}.$$

As before, let $\Delta_I = \det(A^{(I)})$. Then the matrix A is positive semidefinite if and only if $\Delta_I \geq 0$ for every subset I.

This is a reasonable test for small matrices. For larger matrices, we quickly run into trouble, because we'll need to compute 2^n determinants.

2 Classifying critical points locally

Sylvester's criterion lets us prove the following result:

Theorem 2.1. *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 $Hf(\mathbf{x}^*) \succ 0$, then $\mathbf{x}^*$ is a strict local minimizer.

If $Hf(\mathbf{x}^*) \prec 0$, then $\mathbf{x}^*$ is a strict local maximizer.

This adds on to our result from the previous lecture: if $Hf(\mathbf{x}^*)$ is indefinite, then $\mathbf{x}^*$ is a saddle point.

Proof. The missing ingredient provided by Sylvester's criterion is this: $\Delta_1, \Delta_2, \dots, \Delta_n$ are *continuous* functions of the entries of the matrix. So if each of them is positive for $Hf(\mathbf{x}^*)$, then we can pick a positive radius $r > 0$ such that each of them is *still* positive for $Hf(\mathbf{x})$ when $\|\mathbf{x} - \mathbf{x}^*\| < r$.

As a result, Hf is positive definite everywhere in the open ball $B(\mathbf{x}^*, r)$, and therefore $\mathbf{x}^*$ is a strict minimizer on this ball. This is precisely what it means to be a strict local minimizer. □

It is still the case that we don't know exactly what happens if $Hf(\mathbf{x}^*) \succeq 0$ or $Hf(\mathbf{x}^*) \preceq 0$. That's not entirely true: we know a little. If $Hf(\mathbf{x}^*)$ has at least one positive eigenvalue, then its eigenvector is a direction $\mathbf{u}$ such that $f(\mathbf{x}^* + t\mathbf{u})$ is strictly minimized at $t = 0$. So $\mathbf{x}^*$ is a local minimizer if it's anything. Similarly, if $Hf(\mathbf{x}^*)$ has a negative eigenvalue, then $\mathbf{x}^*$ is either a saddle point or a local maximizer.

It's only when $Hf(\mathbf{x}^*)$ is the zero matrix that none of these occur, and we can't predict anything at all about the point $\mathbf{x}^*$ based on its Hessian matrix. There are examples of this: for example, the Hessian matrix can't help us tell that $(0,0)$ is a local minimizer of $x^4 + y^4$, a local maximizer of $-x^4 - y^4$, and a saddle point of $x^4 - y^4$.

3 Example of classifying critical points

Consider the two-variable function $f(x, y) = x^2 + 3y^2 + 2xy^3$.

First, we begin by taking the gradient to find the critical points. We have

$$\nabla f(x, y) = \begin{bmatrix} 2x + 2y^3 \\ 6y + 6xy^2 \end{bmatrix}.$$

Starting with $6y + 6xy^2$ seems easier because it factors as $6y(1 + xy)$: we have either $y = 0$ or $y = -\frac{1}{x}$.

If $y = 0$, then the first equation tells us $x = 0$, so we get the critical point $(0, 0)$. If $y = -\frac{1}{x}$, then we get

$$2x - \frac{2}{x^3} = 0 \iff x = \frac{1}{x^3} \iff x^4 = 1 \iff x = \pm 1$$

so we get two further critical points $(1, -1)$ and $(-1, 1)$.

Now we must classify them. We have

$$Hf(x,y) = \begin{bmatrix} 2 & 6y^2 \\ 6y^2 & 12xy+6 \end{bmatrix}.$$

Therefore

$$Hf(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}, \qquad Hf(1,-1) = Hf(-1,1) = \begin{bmatrix} 2 & 6 \\ 6 & -6 \end{bmatrix}.$$

The first matrix is positive definite because it's a diagonal matrix and $2, 6 > 0$. So $(0,0)$ is a strict local minimizer. (Whenever we're definitely able to tell that a point is a local minimizer by looking at the Hessian matrix, it'll be a strict one.)

The second matrix is indefinite: $\Delta_1 = 2$ while $\Delta_2 = -48$. So $(1,-1)$ and $(-1,1)$ are both saddle points.