Jim Lambers
MAT 419/519
Summer Session 2011-12
Lecture 4 Notes

These notes correspond to Sections 1.4 and 1.5 in the text.

Coercive Functions and Global Minimizers

We now know how to prove that a critical point of a function $f(\mathbf{x})$ is a global minimizer if the Hessian of $f(\mathbf{x})$ is positive semidefinite on all of $\mathbb{R}^n$ (or a strict global minimizer if the Hessian is positive definite), but we need a way to establish global minimizers even if the Hessian is not necessarily positive semidefinite on all of $\mathbb{R}^n$. Fortunately, this is easily accomplished in certain cases.

Previously, we defined a set $D \subseteq \mathbb{R}^n$ to be *closed* if its complement D^c in $\mathbb{R}^n$ is open. Intuitively, a closed set includes its boundary, if it has one, whereas an open set does not. This sets the stage for the following definitions.

Definition We then say that a set $D \subseteq \mathbb{R}^n$ is *bounded* if there exists a constant $M > 0$ such that $\|\mathbf{x}\| < M$ for all $\mathbf{x} \in D$. The set D is said to be *compact* if it is closed and bounded.

Example A closed interval $[a, b]$ is bounded, and is therefore also compact. The circle and its interior $\{(x, y) \,|\, x^2 + y^2 \leq 1\}$ is a closed set, and is also bounded, and therefore it is compact. The interval $[0, \infty)$ is closed, as its complement $(-\infty, 0)$ is open, but it is not bounded, so it is not compact either. □

The following theorem is an essential result concerning the existence of global minimizers and maximizers on compact sets.

Theorem Let D be a compact subset of $\mathbb{R}^n$. If $f(\mathbf{x})$ is a continuous function on D, then $f(\mathbf{x})$ has a global maximizer and a global minimizer on D.

We now describe functions for which global minimizers can be found even on sets that are not bounded or not closed.

Definition A continuous function $f(\mathbf{x})$ that is defined on all of $\mathbb{R}^n$ is *coercive* if

$$\lim_{\|\mathbf{x}\| \to \infty} f(\mathbf{x}) = +\infty.$$

That is, for any constant $M > 0$ there exists a constant $R_M > 0$ such that $\|f(\mathbf{x}\| > M$ whenever $\|\mathbf{x}\| > R_M$.

Example The function $f(x, y) = x^2 + y^2$ is coercive, as

$$\lim_{\|\mathbf{x}\| \to \infty} f(x, y) = \lim_{\|\mathbf{x}\| \to \infty} \|\mathbf{x}\|^2 = +\infty.$$

$\square$

Example A linear function is never coercive. For instance, a linear function on $\mathbb{R}^2$ has the form

$$f(x, y) = ax + by + c,$$

for constants a, b and c, and is equal to c along the line defined by the equation $ax + by = 0$. Since $\|\mathbf{x}\| \to \infty$ along this line, but $f(x, y) = c$ along this line, $f(x, y)$ is not coercive. $\square$

Example The function $f(x, y) = x^2 - 2xy + y^2$ is not coercive, as $f(x, y) = (x - y)^2$, which means that $f(x, y) = 0$ on the line $y = x$, along which $\|\mathbf{x}\|$ becomes infinite. $\square$

As these examples show, in order for a function to be coercive, it must approach $+\infty$ along *any* path within $\mathbb{R}^n$ on which $\|\mathbf{x}\|$ becomes infinite.

The following theorem indicates the usefulness of knowing whether a function is coercive.

Theorem Let $f(\mathbf{x})$ be a continuous function defined on all of $\mathbb{R}^n$. If $f(\mathbf{x})$ is coercive, then $f(\mathbf{x})$ has a global minimizer. Furthermore, if the first partial derivatives of $f(\mathbf{x})$ exist on all of $\mathbb{R}^n$, then any global minimizers of $f(\mathbf{x})$ can be found among the critical points of $f(\mathbf{x})$.

This theorem can be proved by using the fact that $f(\mathbf{x})$ is coercive to find a compact subset of $\mathbb{R}^n$ on which $f(\mathbf{x})$ must have a global minimizer, by the preceding theorem.

To find the global minimizer of a coercive function $f(\mathbf{x})$, it is sufficient to find the critical points of $f(\mathbf{x})$, and then evaluate $f(\mathbf{x})$ at each of these points. The critical points for which $f(\mathbf{x})$ assumes the smallest values are then the global minimizers.

Example Let $f(x, y) = x^4 - 4xy + y^4$. Then we have

$$\nabla f(x, y) = (4x^3 - 4y, -4x + 4y^3),$$

and therefore critical points satisfy $y = x^3$ and $x = y^3$. That is, $x = x^9$, which means $x = 0$, $x = 1$ or $x = -1$. This yields the critical points $(0, 0)$, $(1, 1)$ and $(-1, -1)$. However, as the Hessian is

$$Hf(x, y) = \begin{bmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{bmatrix},$$

which is *indefinite* at $(0, 0)$, as its determinant is equal to -16.

We instead determine whether $f(x, y)$ is coercive. We have

$$f(x, y) = (x^4 + y^4) \left(1 - \frac{4xy}{x^4 + y^4} \right).$$

As $\|(x,y)\| \to \infty$, $4xy/(x^4+y^4) \to 0$, while $x^4+y^4 \to +\infty$. It follows that

$$\lim_{\|(x,y)\|\to\infty} f(x,y) = \lim_{\|(x,y)\|\to\infty} (x^4+y^4)(1-0) = +\infty.$$

Therefore, $f(x,y)$ is coercive, and the three critical points are candidates for global minimizers. Evaluating $f(x,y)$ at $(0,0)$, $(1,1)$ and $(-1,1)$, we obtain

$$f(0,0) = 0, \quad f(1,1) = -2, \quad f(-1,-1) = -2,$$

and therefore $(1,1)$ and $(-1,-1)$ are global minimizers. $\Box$

Eigenvalues and Positive Definite Matrices

We now use concepts from linear algebra to obtain simpler, more intuitive criteria for determining whether a symmetric matrix, such as the Hessian of a function at a point, is positive or negative definite or semidefinite.

Definition Let A be an $n \times n$ symmetric matrix. A nonzero vector $\mathbf{x} \in \mathbb{R}^n$ is an *eigenvector* of A if there exists a scalar λ such that

$$A\mathbf{x} = \lambda\mathbf{x}.$$

The scalar λ is called an *eigenvalue* of A corresponding to $\mathbf{x}$.

From the equation

$$A\mathbf{x} - \lambda\mathbf{x} = (A - \lambda I)\mathbf{x} = \mathbf{0},$$

and the fact that $\mathbf{x} \neq \mathbf{0}$, it follows that the matrix $A - \lambda I$ is not invertible (that is, it is *singular*). Therefore, any eigenvalue λ of A satisfies

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

This determinant is a polynomial of degree n in λ, which is called the *characteristic polynomial*. Therefore, the eigenvalues can be found by computing the characteristic polynomial, and then computing its roots.

For a general matrix A, the eigenvalues may be real or complex, because a polynomial with real coefficients can have complex roots, but the eigenvalues of a symmetric matrix A are real. Furthermore, if A is symmetric, there exists an *orthogonal* matrix P, meaning that $P^TP = I$, such that

$$A = PDP^T,$$

where D is a *diagonal* matrix whose diagonal entries are the eigenvalues of A. The columns of P are *orthonormal* vectors, meaning that they are orthogonal and are of magnitude 1. They are also the eigenvectors of A.

It follows that

$$\begin{aligned} Q_A(\mathbf{x}) &= \mathbf{x}\cdot A\mathbf{x} \\ &= \mathbf{x}^T A\mathbf{x} \\ &= \mathbf{x}^T PDP^T\mathbf{x} \\ &= (P^T\mathbf{x})^T D(P^T\mathbf{x}) \\ &= \mathbf{y}^T D\mathbf{y} \\ &= \sum_{j=1}^n \lambda_i y_i^2, \end{aligned}$$

where $\mathbf{y} = P^T\mathbf{x}$ and

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}.$$

This simple form of $Q_A(\mathbf{x})$ yields the following conditions for classifying A.

Theorem Let A be a symmetric matrix. Then

1. A is positive definite if and only if all of its eigenvalues are positive;
2. A is negative definite if and only if all of its eigenvalues are negative;
3. A is positive semidefinite if and only if all of its eigenvalues are nonnegative;
4. A is negative semidefinite if and only if all of its eigenvalues are nonpositive;
5. A is indefinite if and only if at least one of its eigenvalues is positive and at least one of its eigenvalues is negative.

We now illustrate the use of these conditions for optimization.

Example Let $f(x,y,z) = x^2 + y^2 + z^2 - 4xy$. Then we have

$$\nabla f(x,y,z) = (2x-4y, 2y-4x, 2z),$$

which yields the critical point $(0,0,0)$, and

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

This matrix has the characteristic polynomial

$$\begin{aligned}
\det(Hf(x,y,z) - \lambda I) &= \det \begin{bmatrix} 2-\lambda & -4 & 0 \\ -4 & 2-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{bmatrix} \\
&= (2-\lambda)[(2-\lambda)^2 - 16] \\
&= (2-\lambda)(\lambda^2 - 4\lambda - 12) \\
&= (2-\lambda)(\lambda+2)(\lambda-6).
\end{aligned}$$

Therefore, the eigenvalues are 2, -2 and 6, which means that the Hessian is indefinite. We conclude that $(0,0,0)$ is a saddle point, and there are no global maximizers or minimizers. $\square$

Exercises

1. Chapter 1, Exercise 8
2. Chapter 1, Exercise 12
3. Chapter 1, Exercise 13
4. Chapter 1, Exercise 14