Optimization

Class 6: Second-Derivative Test

1 Optimization via calculus

- One-variable: $f(x)$
  - first-derivative test: $f'(\bar{x}) = 0$ identifies stationary points $\bar{x}$.
  - second-derivative test:
    - $f''(\bar{x}) > 0 \Rightarrow \bar{x}$ is a local minimizer
    - $f''(\bar{x}) < 0 \Rightarrow \bar{x}$ is a local maximizer
    - $f''(\bar{x}) = 0 \Rightarrow$ nothing.
- Two-variables $f(x, y)$
  - first-derivative test: $\nabla f(\bar{x}, \bar{y}) = \vec{0}$ identifies stationary points $(\bar{x}, \bar{y})$.
  - second-derivative test: (a relatively elaborate series of tests involving the second partial derivatives)
    - $f_{xx}(\bar{x}, \bar{y}) f_{yy}(\bar{x}, \bar{y}) > f_{xy}(\bar{x}, \bar{y})^2$
      - $f_{xx}(\bar{x}, \bar{y}) > 0 \Rightarrow (\bar{x}, \bar{y})$ is a local minimizer
      - $f_{xx}(\bar{x}, \bar{y}) < 0 \Rightarrow (\bar{x}, \bar{y})$ is a local maximizer
    - $f_{xx}(\bar{x}, \bar{y}) f_{yy}(\bar{x}, \bar{y}) < f_{xy}(\bar{x}, \bar{y})^2 \Rightarrow (\bar{x}, \bar{y})$ is a saddle point (not maximizer or minimizer).
    - $f_{xx}(\bar{x}, \bar{y}) f_{yy}(\bar{x}, \bar{y}) = f_{xy}(\bar{x}, \bar{y})^2 \Rightarrow$ nothing.
- More-variables $f(x_1, x_2, x_3, \ldots, x_n)$
  - first-derivative test: $\nabla f(x_1, x_2, x_3, \ldots, x_n) = \vec{0}$ identifies stationary points $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_n)$.
  - second-derivative test: ?????

2 Multi-variable second-derivative test

We need a second-derivative test for $f(x, y)$ that is easily scaled-up to $f(x_1, x_2, x_3, \ldots, x_n)$.

2.1 Remainder form of Taylor quadratic

Recall the Taylor quadratic $TQ(x, y)$:

$$f(x,y) \approx \underbrace{\frac{1}{2}\begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix}^T \nabla^2 f(\bar{x},\bar{y}) \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix} + \nabla f(\bar{x},\bar{y})^T \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix} + f(\bar{x},\bar{y})}_{TQ(x,y)}.$$

The "remainder form" gives a perfect approximation of $f(x,y)$, at the price of some mystery:

$$f(x,y) = \underbrace{\frac{1}{2}\begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix}^T \overbrace{\nabla^2 f(\xi_x,\xi_y)}^{\text{at mystery point}} \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix} + \nabla f(\bar{x},\bar{y})^T \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix} + f(\bar{x},\bar{y})}_{\text{remainder form of Taylor quadratic}}.$$

The mystery point (ξ_x,ξ_y) is somewhere between (x,y) and the base point $(\bar{x},\bar{y})$.

2.2 Impractical second-derivative test

Group Exercise 3.2.1. Confirm the following second-derivative test:

> A stationary point $(\bar{x},\bar{y})$ is a minimizer (maximizer) if the vector-matrix-vector multiplication
> $$\begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix}^T \nabla^2 f(\xi_x,\xi_y) \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix}$$
> is always positive (negative) for $(x,y) \neq (\bar{x},\bar{y})$. As shorthand, we call such matrices *positive-definite* (*negative-definite*).

Group Exercise 3.2.1. (solution)

For $(x,y) \neq (\bar{x},\bar{y})$:

$$f(x,y) - f(\bar{x},\bar{y}) = \frac{1}{2}\begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix}^T \nabla^2 f(\xi_x,\xi_y) \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix} > 0 \quad \Rightarrow \quad f(x,y) > f(\bar{x},\bar{y}).$$

The last inequality implies directly that $(\bar{x},\bar{y})$ is a minimizer. (The opposite starting inequality leads to the conclusion of maximizer.)

Notice this gives *global* optimizers....but depends on a mysterious point.

2.3 Compromise for practicality

Instead, we test the positive-definiteness (or negative-definiteness) of the second-derivative matrix $\nabla^2 f(\bar{x},\bar{y})$ at the base point $(\bar{x},\bar{y})$ (which we know for certain). Then, as long as the mystery point is not too far from the base point, the continuity of the second-derivative matrix $\nabla^2 f$ guarantees that the two different vector-matrix-vector multiplications have the same sign.

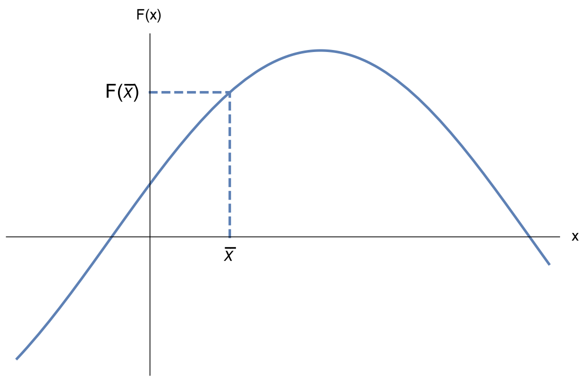

Figure 1: Continuity: nearby inputs ⇒ nearby outputs

Since the mystery point (ξ_x, ξ_y) is somewhere between (x, y) and the base point $(\bar{x}, \bar{y})$, we can ensure that it is close to the base point by ensuring that (x, y) is. So, the best we can do with this approach is to guarantee *local* optimizers.

Q: What does

$$\begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix}^T \nabla^2 f(\bar{x}, \bar{y}) \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix} = 0$$

say about the sign of

$$\begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix}^T \nabla^2 f(\xi_x, \xi_y) \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix}$$

when (x, y) is close to the base point?

A: Nothing. Continuity only says the second output value is close the first. Since the first is zero, the second could be close but negative, close but positive, or even zero itself.

Q: In general, what is impractical about testing for positive-definiteness directly:

$$\begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix}^T \nabla^2 f(\bar{x}, \bar{y}) \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix} > 0 \quad \text{for all } (x, y) \neq (\bar{x}, \bar{y})?$$

A: In general, it is impractical to test the vector-matrix-vector multiplication above for *all* pairs $(x, y) \neq (\bar{x}, \bar{y})$. Fortunately, the *eigenvalues* of the second-derivative matrix determine its "definiteness".

2.4 Eigenvalues

For any matrix

$$A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$$

that is *symmetric* above and below its diagonal, the *eigenvalues* are the solutions ϵ to the quadratic equation

$$(a - \epsilon)(d - \epsilon) = b^2.$$

Mathematica is happy to find eigenvalues (even for other, larger matrices). Here's a nice fact that holds for symmetric $n \times n$ matrices:

- eigenvalues of A are positive $\Leftrightarrow$ A is positive-definite.
- eigenvalues of A are negative $\Leftrightarrow$ A is negative-definite.

2.5 Practical second-derivative test

> If the eigenvalues of $\nabla^2 f$ at a stationary point are positive (negative), then the stationary point is a local minimizer (maximizer).

This works for any number of variables $(x_1, x_2, x_3, \ldots, x_n)$, and turns out to be equivalent in the case of two-variables to the (relatively elaborate) second-derivative test that you saw in calculus.