MATH529 – Fundamentals of Optimization
Unconstrained Optimization II

Marco A. Montes de Oca

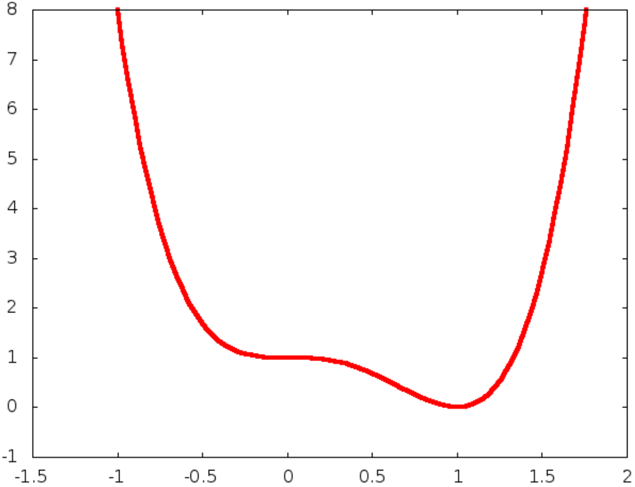
Mathematical Sciences, University of Delaware, USA

Recap

Example

Find the local and global minimizers and maximizers on $\mathbb{R}$ of $f(x) = 3x^4 - 4x^3 + 1$.

Graph of $f(x) = 3x^4 - 4x^3 + 1$.



Optimization via Calculus

Two theorems summarize the basic facts about **global** optimization of one variable functions.

Theorem (1st order condition (Necessary, but not sufficient))

Suppose that $f(x)$ is a differentiable function on $\mathbb{R}$ (or the function's domain I). If $x^\star$ is a global minimizer of $f(x)$, then $f'(x^\star) = 0$.

Optimization via Calculus

Two theorems summarize the basic facts about **global** optimization of one variable functions.

Theorem (1st order condition (Necessary, but not sufficient))

Suppose that $f(x)$ is a differentiable function on $\mathbb{R}$ (or the function's domain I). If $x^\star$ is a global minimizer of $f(x)$, then $f'(x^\star) = 0$.

Theorem (2nd order condition (Sufficient, but not necessary))

Suppose that $f(x)$, $f'(x)$, and $f''(x)$ are all continuous on $\mathbb{R}$ (or I) and that $x^\star$ is a critical point of $f(x)$.

a) If $f''(x) \geq 0$ for all $x \in \mathbb{R}$ (or I), then $x^\star$ is a global minimizer of $f(x)$ on $\mathbb{R}$ (or I).
b) If $f''(x) > 0$ for all $x \in I$ such that $x \neq x^\star$, then $x^\star$ is a strict global minimizer of $f(x)$ on $\mathbb{R}$ (or I).

Local optimization is easier to verify.

Theorem (1st order condition (Necessary, but not sufficient))

Suppose that $f(x)$ is a differentiable function on $\mathbb{R}$ (or I). If $x^\star$ is a local minimizer of $f(x)$, then $f'(x^\star) = 0$.

Optimization via Calculus

Local optimization is easier to verify.

Theorem (1st order condition (Necessary, but not sufficient))

Suppose that $f(x)$ *is a differentiable function on* $\mathbb{R}$ *(or* I*). If* $x^\star$ *is a local minimizer of* $f(x)$*, then* $f'(x^\star) = 0$.

Theorem (2nd order condition (Sufficient, but not necessary))

Suppose that $f(x)$, $f'(x)$, *and* $f''(x)$ *are all continuous on* $\mathbb{R}$ *(or* I*) and that* $x^\star$ *is a critical point of* $f(x)$.

If $f''(x^\star) > 0$, *then* $x^\star$ *is a strict local minimizer of* $f(x)$.

Exercise

Find the local and global minimizers and maximizers on $I = (-1, 1)$ of $f(x) = \ln(1 - x^2)$.

What about functions of many variables?

What about functions of many variables?

Extend theorems that allow us to identify and classifly local minimizers of one variable functions to multivariable cases.

Notation:

A vector in $\mathbb{R}^n$ is an ordered n-tuple $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}$ of real numbers called *components* of $\mathbf{x}$.

If $\mathbf{x}$ and $\mathbf{y}$ are vectors in $\mathbb{R}^n$, then their *dot product* or *inner product* is defined by

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} = (x_1, x_2, x_3, \ldots, x_n) \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{n} x_i y_i$$

where $\mathbf{x}^T$ is the *transpose* of $\mathbf{x}$.

Optimization via Calculus

Notation:
If $f(\mathbf{x})$ is a function of n variables with continuous first and second partial derivatives on $\mathbb{R}^n$, then the *gradient* of $f(\mathbf{x})$ is the vector

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Optimization via Calculus

Notation: The *Hessian* of $f(\mathbf{x})$, denoted by $\nabla^2 f$ or Hf, is the symmetric $n \times n$ matrix

$$\nabla^2 f = Hf = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3^2} & \cdots & \frac{\partial^2 f}{\partial x_3 \partial x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \frac{\partial^2 f}{\partial x_n \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Optimization via Calculus

Definition

Suppose $f(\mathbf{x})$ is a real-valued function defined on a subset D of $\mathbb{R}^n$. A point $\mathbf{x}^\star$ in D is:

- A **global minimizer** for $f(\mathbf{x})$ on D if $f(\mathbf{x}^\star) \leq f(\mathbf{x})$ for all $\mathbf{x} \in D$;

Definition

Suppose $f(\mathbf{x})$ is a real-valued function defined on a subset D of $\mathbb{R}^n$. A point $\mathbf{x}^\star$ in D is:

- A **global minimizer** for $f(\mathbf{x})$ on D if $f(\mathbf{x}^\star) \leq f(\mathbf{x})$ for all $\mathbf{x} \in D$;
- A **strict global minimizer** for $f(\mathbf{x})$ on D if $f(\mathbf{x}^\star) < f(\mathbf{x})$ for all $\mathbf{x} \in D$ such that $\mathbf{x} \neq \mathbf{x}^\star$;

Optimization via Calculus

Definition

Suppose $f(\mathbf{x})$ is a real-valued function defined on a subset D of $\mathbb{R}^n$. A point $\mathbf{x}^\star$ in D is:

- A **global minimizer** for $f(\mathbf{x})$ on D if $f(\mathbf{x}^\star) \leq f(\mathbf{x})$ for all $\mathbf{x} \in D$;
- A **strict global minimizer** for $f(\mathbf{x})$ on D if $f(\mathbf{x}^\star) < f(\mathbf{x})$ for all $\mathbf{x} \in D$ such that $\mathbf{x} \neq \mathbf{x}^\star$;
- A **local minimizer** for $f(\mathbf{x})$ if there is a positive number δ such that $f(\mathbf{x}^\star) \leq f(\mathbf{x})$ for all $\mathbf{x} \in D$ for which $\mathbf{x}$ in $B(\mathbf{x}^\star, \delta)$;

Optimization via Calculus

Definition

Suppose $f(\mathbf{x})$ is a real-valued function defined on a subset D of $\mathbb{R}^n$. A point $\mathbf{x}^\star$ in D is:

- A **global minimizer** for $f(\mathbf{x})$ on D if $f(\mathbf{x}^\star) \leq f(\mathbf{x})$ for all $\mathbf{x} \in D$;
- A **strict global minimizer** for $f(\mathbf{x})$ on D if $f(\mathbf{x}^\star) < f(\mathbf{x})$ for all $\mathbf{x} \in D$ such that $\mathbf{x} \neq \mathbf{x}^\star$;
- A **local minimizer** for $f(\mathbf{x})$ if there is a positive number δ such that $f(\mathbf{x}^\star) \leq f(\mathbf{x})$ for all $\mathbf{x} \in D$ for which $\mathbf{x}$ in $B(\mathbf{x}^\star, \delta)$;
- A **strict local minimizer** for $f(\mathbf{x})$ if there is a positive number δ such that $f(\mathbf{x}^\star) < f(\mathbf{x})$ for all $\mathbf{x} \in D$ for which $\mathbf{x}$ in $B(\mathbf{x}^\star, \delta)$ and $\mathbf{x} \neq \mathbf{x}^\star$;

Optimization via Calculus

Definition

Suppose $f(\mathbf{x})$ is a real-valued function defined on a subset D of $\mathbb{R}^n$. A point $\mathbf{x}^\star$ in D is:

- A **global minimizer** for $f(\mathbf{x})$ on D if $f(\mathbf{x}^\star) \leq f(\mathbf{x})$ for all $\mathbf{x} \in D$;
- A **strict global minimizer** for $f(\mathbf{x})$ on D if $f(\mathbf{x}^\star) < f(\mathbf{x})$ for all $\mathbf{x} \in D$ such that $\mathbf{x} \neq \mathbf{x}^\star$;
- A **local minimizer** for $f(\mathbf{x})$ if there is a positive number δ such that $f(\mathbf{x}^\star) \leq f(\mathbf{x})$ for all $\mathbf{x} \in D$ for which $\mathbf{x}$ in $B(\mathbf{x}^\star, \delta)$;
- A **strict local minimizer** for $f(\mathbf{x})$ if there is a positive number δ such that $f(\mathbf{x}^\star) < f(\mathbf{x})$ for all $\mathbf{x} \in D$ for which $\mathbf{x}$ in $B(\mathbf{x}^\star, \delta)$ and $\mathbf{x} \neq \mathbf{x}^\star$;
- A **critical point** (also called a **stationary point**) of $f(\mathbf{x})$ if the first partial derivatives of $f(\mathbf{x})$ exist at $\mathbf{x}^\star$ and $\frac{\partial f}{\partial x_i} = 0$, for $i = 1, 2, 3 \ldots, n$.

Theorem (Multivariable Taylor's formula)

Suppose that $\mathbf{x}$, $\mathbf{x}^\star$ *are points in* $\mathbb{R}^n$ *and that* $f(\mathbf{x})$ *is a real-valued function of* n *variables with continuous first and second partial derivatives on some open set containing the line segment* $[\mathbf{x}^\star, \mathbf{x}] = \{\mathbf{w} \in \mathbb{R}^n : \mathbf{w} = \mathbf{x}^\star + t(\mathbf{x} - \mathbf{x}^\star)\,,\, 0 \leq t \leq 1\}$ *joining* $\mathbf{x}^\star$ *and* $\mathbf{x}$. *Then, there exists a* $\mathbf{z} \in [\mathbf{x}^\star, \mathbf{x}]$ *such that*

$$f(\mathbf{x}) = f(\mathbf{x}^\star) + \nabla f(\mathbf{x}^\star)^T(\mathbf{x} - \mathbf{x}^\star) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^\star)^T Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^\star)$$

Optimization via Calculus

Theorem (Local minimizer identification)

Suppose that $f(\mathbf{x})$ is a real-valued function for which all first partial derivatives of $f(\mathbf{x})$ exist on a subset $D \in \mathbb{R}^n$. If $\mathbf{x}^\star$ is an interior point of D that is a local minimizer of $f(\mathbf{x})$, then $\nabla f(\mathbf{x}^\star) = \mathbf{0}$.

Theorem (Classification of minimizers (maximizers))

Suppose that $\mathbf{x}^\star$ *is a critical point of a function* $f(\mathbf{x})$ *with continuous first and second partial derivatives on* $\mathbb{R}^n$*. Then:*

- $\mathbf{x}^\star$ *is a* **global minimizer** *of* $f(\mathbf{x})$ *if* $(\mathbf{x}-\mathbf{x}^\star)^T Hf(\mathbf{z})(\mathbf{x}-\mathbf{x}^\star) \geq 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$ *and all* $\mathbf{z} \in [\mathbf{x}^\star, \mathbf{x}]$*;*
- $\mathbf{x}^\star$ *is a* **strict global minimizer** *of* $f(\mathbf{x})$ *if* $(\mathbf{x}-\mathbf{x}^\star)^T Hf(\mathbf{z})(\mathbf{x}-\mathbf{x}^\star) > 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$ *such that* $\mathbf{x} \neq \mathbf{x}^\star$ *and for all* $\mathbf{z} \in [\mathbf{x}^\star, \mathbf{x}]$*;*
- $\mathbf{x}^\star$ *is a* **global maximizer** *of* $f(\mathbf{x})$ *if* $(\mathbf{x}-\mathbf{x}^\star)^T Hf(\mathbf{z})(\mathbf{x}-\mathbf{x}^\star) \leq 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$ *and all* $\mathbf{z} \in [\mathbf{x}^\star, \mathbf{x}]$*;*
- $\mathbf{x}^\star$ *is a* **strict global maximizer** *of* $f(\mathbf{x})$ *if* $(\mathbf{x}-\mathbf{x}^\star)^T Hf(\mathbf{z})(\mathbf{x}-\mathbf{x}^\star) < 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$ *such that* $\mathbf{x} \neq \mathbf{x}^\star$ *and for all* $\mathbf{z} \in [\mathbf{x}^\star, \mathbf{x}]$*;*

Optimization via Calculus

Practical ways to use the previous theorem:

Conditions that involve the form $(\mathbf{x} - \mathbf{x}^\star)^T Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^\star)$, or in general $\mathbf{v}^T A \mathbf{v}$, where A is a symmetric square matrix, call for methods to **identify whether A (in our case the Hessian of the objective function) is positive or negative (semi)definite**.

Quadratic forms:

Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

The quadratic form $Q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = a_{11}x_1^2 + a_{12}x_1x_2 + a_{13}x_1x_3 + \ldots + a_{ij}x_ix_j + \ldots + a_{ii}x_i^2 + \ldots + a_{nn}x_n^2$.

Example

Write the quadratic form associated with the following matrix:

$$A = \begin{pmatrix} -1 & 0 & 2 & -3 \\ 0 & 2 & 1/2 & -1 \\ 2 & 1/2 & 0 & 4 \\ -3 & -1 & 4 & 5 \end{pmatrix}.$$

Determining whether a quadratic form $Q_A(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbb{R}^n$.
Example in class . . . .