Math 484: Nonlinear programming

Chapter 1: Lecture 3

Questions for $\mathbb{R}^n$

Now, we will consider functions $f : \mathbb{R}^n \to \mathbb{R}$, that take a vector in $\mathbb{R}^n$ as input and give a real number as output.

We would like to use our tools for one-variable functions, so we will be interested in restricting functions to a line.

Lines in $\mathbb{R}^n$

A line in $\mathbb{R}^n$ is given by a *position* and a *direction*.

If ℓ is a line that contains the point $\mathbf{x} \in \mathbb{R}^n$ and the direction vector $\mathbf{u} \in \mathbb{R}^n$, then ℓ has the representation

$$\ell = \{\mathbf{x} + t\mathbf{u} : t \in \mathbb{R}\}.$$

Restricting a function to a line

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function, and let

$$\ell = \{\mathbf{x} + t\,\mathbf{u} : t \in \mathbb{R}\}$$

be a line. We define the *restriction of f to ℓ* as the function

$$\begin{aligned} \phi_{\mathbf{x},\mathbf{u}} \;:\; & \mathbb{R} \to \mathbb{R} \\ & t \mapsto f(\mathbf{x} + t\,\mathbf{u}). \end{aligned}$$

In other words, $\phi_{\mathbf{x},\mathbf{u}}(t) = f(\mathbf{x} + t\,\mathbf{u})$.

Global minimizers in $\mathbb{R}^n$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. We say that $\mathbf{x}^* \in \mathbb{R}^n$ is a *global minimizer* if $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$.

Lemma

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. A point $\mathbf{x}^* \in \mathbb{R}^n$ is a global minimizer of f if and only if for every vector $\mathbf{u} \in \mathbb{R}^n$, $t = 0$ is the global minimizer of $\phi_{\mathbf{x}^*,\mathbf{u}}(t) = f(\mathbf{x}^* + t\,\mathbf{u})$.

Theorem (Chain rule)

If $f : \mathbb{R}^n \to \mathbb{R}$ *and* $g : \mathbb{R} \to \mathbb{R}^n$ *are functions, then* $f \circ g : \mathbb{R} \to \mathbb{R}$. *Furthermore,*

$$\frac{d}{dt}(f(\mathbf{g}(t)) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{g}(t))\frac{d}{dt}g_i(t).$$

Theorem (Chain rule)

If $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}^n$ are functions, then $f \circ g : \mathbb{R} \to \mathbb{R}$. Furthermore,

$$\frac{d}{dt}(f(\mathbf{g}(t)) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{g}(t))\frac{d}{dt}g_i(t).$$

Consider the function $\phi_{\mathbf{x},\mathbf{u}}(t) = f(\mathbf{x} + t\,\mathbf{u})$.

$$\phi'_{\mathbf{x},\mathbf{u}}(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x} + t\,\mathbf{u})u_i.$$

We define

$$\nabla f(\mathbf{x})^T = \left[\frac{\partial f}{\partial x_1}(\mathbf{x}) \ \frac{\partial f}{\partial x_2}(\mathbf{x}) \ \cdots \ \frac{\partial f}{\partial x_n}(\mathbf{x})\right]$$

Consider the function $\phi_{\mathbf{x},\mathbf{u}}(t) = f(\mathbf{x} + t\,\mathbf{u})$.

$$\phi'_{\mathbf{x},\mathbf{u}}(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x} + t\,\mathbf{u}) u_i = \nabla f(\mathbf{x} + t\,\mathbf{u}) \cdot \mathbf{u}\,.$$

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ *be a function. If* ∇f *is continuous and* $\mathbf{x}^*$ *is a global minimizer of* f*, then* $\nabla f(\mathbf{x}^*) = 0$*.*

If $\nabla f(\mathbf{x}^*) = 0$, then we say that $\mathbf{x}^*$ is a *critical point* of f.

Second derivative test in $\mathbb{R}^n$

Consider the function $\phi_{\mathbf{x},\mathbf{u}}(t) = f(\mathbf{x} + t\,\mathbf{u})$.

$$\phi'_{\mathbf{x},\mathbf{u}}(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x} + t\,\mathbf{u}) u_i = \nabla f(\mathbf{x} + t\,\mathbf{u}) \cdot \mathbf{u}\,.$$

$$\phi''_{\mathbf{x},\mathbf{u}}(t) = \sum_{i=1}^{n} u_i \left(\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x} + t\,\mathbf{u}) u_j \right).$$

$$\phi''_{\mathbf{x},\mathbf{u}}(t) = \sum_{i=1}^{n} u_i \left(\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x} + t\,\mathbf{u}) u_j \right).$$

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

where

$$Hf(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 x_n}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 x_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_n x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) \end{bmatrix}$$

is the *Hessian matrix* of f.

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function for which Hf is continuous. If $\nabla f(\mathbf{x}^*) = 0$ and

$$\mathbf{u}^T\, Hf(\mathbf{x})\, \mathbf{u} \geq 0 \textit{ for all } \mathbf{u} \in \mathbb{R}^n \textit{ and } \mathbf{x} \in \mathbb{R}^n,$$

then $\mathbf{x}^*$ is a global minimizer of f.

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

Minimizing over other sets

Suppose that $D \subseteq \mathbb{R}^n$, and $f : D \to \mathbb{R}$ is a function. Can we use our tools to optimize f?

We will want two assumptions:

- We do not want D to contain boundary points. (D should be open.)
- Given $\mathbf{x} \in D$, we want to be able to "see" every $\mathbf{y} \in D$ along a straight line. (D should be convex.)

Given an open set $D \subseteq \mathbb{R}^n$ and a function $f : D \to \mathbb{R}$, we say that $\mathbf{x}^* \in D$ is a *local minimizer* of f if there exists $r > 0$ such that $f(\mathbf{x}^*) \leq f(\mathbf{x})$ whenever $\|\mathbf{x}^* - \mathbf{x}\| < r$.

In other words, $\mathbf{x}^*$ is a global minimizer of the function f restricted to the open ball $B(\mathbf{x}^*, r)$.

Theorem

Let $D \subseteq \mathbb{R}^n$ be an open set. Let $f : D \to \mathbb{R}$ be a function for which ∇f is continuous, and let $\mathbf{x}^* \in D$. If $\mathbf{x}^*$ is a local minimizer of f, then $\mathbf{x}^*$ is a critical point of f (i.e. $\nabla f(\mathbf{x}^*) = 0$).

Theorem

Let $D \subseteq \mathbb{R}^n$ be an open set. Let $f : D \to \mathbb{R}$ be a function, and suppose $\mathbf{x}^* \in D$ is a critical point of f. If there exists a value $r > 0$ such that

$$\mathbf{u}^T Hf(\mathbf{x})\,\mathbf{u} \geq 0 \text{ for all } \mathbf{u} \in \mathbb{R}^n \text{ and } \mathbf{x} \in B(\mathbf{x}^*, r),$$

then $\mathbf{x}^*$ is a local minimizer of f.