Continuous Optimization I

Definition (Optimization problem)

We consider the optimization problem:

$$\begin{array}{lll} \text{minimize} & f_0(\mathbf{x}) & \\ \text{subject to} & f_i(\mathbf{x}) = 0 & i = 1, \cdots, m \\ & h_j(\mathbf{x}) \leq 0 & j = 1, \cdots, p \end{array}$$

with $\mathbf{x} \in \mathbf{R}^d$.

Reference:

- Chapter 4 in *Mathematical Optimization in Graphics and Vision*, P.C.Pinto Carvalho and L. Velho, SIGGRAPH 2003 (available online).

Continuous Optimization II

- When $m = p = 0$, the problem is called unconstrained.

Example (Least squares)

$$\mathbf{x}_{LS} = \arg\min_{\mathbf{x}} \{f_0(\mathbf{x}) = \|\mathrm{A}\mathbf{x} - \mathbf{y}\|^2\}$$

- Otherwise, the problem is constrained.

Example (Principal Component Analysis)

$$\mathbf{x}_{pca} = \arg\max_{\mathbf{x}} \left\{ f_0(\mathbf{x}) = \mathbf{x}^T \mathrm{C}\mathbf{x} \text{ constraint to } \mathbf{x}^T\mathbf{x} - 1 = 0 \right\}$$

Reminder: Gradient vector and Hessian Matrix I

Definition

The gradient vector of the function $f : \mathbb{R}^n \to \mathbb{R}$ is:

$$\nabla f(\mathbf{x}) = \frac{\partial y}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

with $y = f(\mathbf{x})$, $y \in \mathbb{R}$ (c.f. appendix D). The Hessian matrix of the function f is defined as:

$$\mathrm{H} = \nabla^2 f(\mathbf{x}) = \frac{\partial(\nabla f(\mathbf{x}))}{\partial \mathbf{x}} = \nabla(\nabla f(\mathbf{x})) = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & & \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_n} \end{pmatrix}$$

Note: H is (square and) symmetric.

Real Symmetric Matrix I

For real symmetric matrices, it can be shown that:

- The n eigenvalues of a real symmetric matrix of order $n \times n$ are real.
- The eigenvectors corresponding to distinct eigenvalues are orthogonal. The eigenvectors corresponding to multiple roots may be orthogonalized with respect to each other.
- The n eigenvectors form a complete orthonormal basis for the Euclidean space $\mathbb{R}^n$.

Positive definite matrix I

Definition (Positive definite matrix)

A square and symmetric matrix A is said positive definite if

$$\mathbf{x}^T \mathrm{A}\mathbf{x} > 0 \quad , \forall \mathbf{x} \in \mathbb{R}^n$$

It is said to be non-negative if

$$\mathbf{x}^T \mathrm{A}\mathbf{x} \geq 0 \quad , \forall \mathbf{x} \in \mathbb{R}^n$$

Positive definite matrix II

You can check if a matrix is positive definite by computing all its n eigenvalues $\{\lambda_i\}_{i=1\cdots n}$:

- if $\forall i,\ \lambda_i > 0$, then the matrix is positive definite.
- if $\forall i\ \lambda_i \geq 0$ and $\exists i\ \lambda_i = 0$, then the matrix is non-negative.
- if $\forall i,\ \lambda_i < 0$, then the matrix is negative definite.
- Otherwise the matrix is indefinite.

Second Order Taylor's Theorem

Theorem (Taylor's Theorem)

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ *and a point* $\mathbf{x}_0 \in \mathbb{R}^n$ *and* $\mathbf{h} \in \mathbb{R}^n$.

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot \mathbf{h} + \frac{1}{2}\mathbf{h}^T \nabla^2 f(\mathbf{x}_0) \cdot \mathbf{h} + r(\mathbf{h})$$

with $\lim_{\mathbf{h} \to \mathbf{0}} \frac{r(\mathbf{h})}{|\mathbf{h}|} = 0$.

Optimality conditions I

Theorem (First Order necessary conditions)

Let $f : \mathbb{R}^n \to \mathbb{R}$ *be a function of class* C^1. *If* $\mathbf{x}_0$ *is a local minimum (or maximum) point of* f, *then* $\nabla f(\mathbf{x}_0) = 0$.

Theorem (Second Order necessary conditions)

Let $f : \mathbb{R}^n \to \mathbb{R}$ *be a function of class* C^2. *If* $\mathbf{x}_0$ *is a local minimum point of* f, *then* $\nabla f(\mathbf{x}_0) = 0$ *and* $\nabla^2 f(\mathbf{x}_0)$ *is non-negative.*

Optimality conditions II

Theorem (Second Order necessary and sufficient conditions)

Let $f : \mathbb{R}^n \to \mathbb{R}$ *be a function of class* C^2. *If* $\nabla f(\mathbf{x}_0) = 0$ *and* $\nabla^2 f(\mathbf{x}_0)$ *is positive definite, then* $\mathbf{x}_0$ *is a local minimum point of* f.

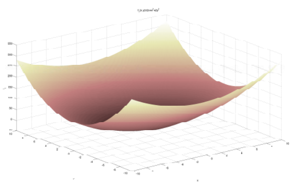
From these two theorems, we conclude that:

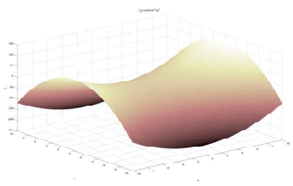
- $$(\mathbf{x}_o \text{ local minimum}) \equiv ((\nabla f(\mathbf{x}_0) = 0) \wedge (\mathrm{H} \text{ positive definite}))$$
- $$(\mathbf{x}_o \text{ local maximum}) \equiv ((\nabla f(\mathbf{x}_0) = 0) \wedge (\mathrm{H} \text{ negative definite}))$$

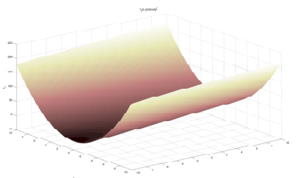
Exercises

1. Consider $f_1(x,y) = 2x + x^2 + 2y^2$.
   1. Compute the gradient of f
   2. Compute the Hessian of f
   3. Conclude if the point $(-1;0)$ is a local maximum or minimum
2. Same questions with $f_2(x,y) = 2x + x^2 - 2y^2$.
3. Same questions with $f_3(x,y) = 2x + 2y^2$.

Exercises







Convexity: Global minima

Definition (Convexity)

A function $f : \mathbb{R} \to \mathbb{R}^n$ is convex when $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ and $0 < \lambda < 1$:

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)$$

Theorem

For $f : \mathbb{R} \to \mathbb{R}^n$ *of class* C^2

$$(f \text{ convex}) \equiv (\mathrm{H} \text{ is non-negative})$$

Theorem

If $f : \mathbb{R} \to \mathbb{R}^n$ *is convex and has a local minimum at* $\mathbf{x}_0$, *then* f *has a global minimum at* $\mathbf{x}_0$.