Non-Linear Optimization and Optimality Conditions

Optimization Methods in Management Science
Master in Management
HEC Lausanne

Dr. Rodrigue Oeuvray

Fall 2019 Semester

Outline

- General formulation
- Unconstrained optimization
- Optimality conditions for unconstrained optimization
- Constrained optimization
- Lagrange multipliers
- Optimality conditions for constrained optimization

Formulation

A constrained optimization problem can be formulated as:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})$$

$$\text{s.t.} \quad \left.\begin{array}{c} g_1(\boldsymbol{x}) \\ \vdots \\ g_m(\boldsymbol{x}) \end{array}\right\} \begin{array}{c} \leq \\ = \\ \geq \end{array} \left\{\begin{array}{c} b_1 \\ \vdots \\ b_m \end{array}\right.$$

- There is no universal algorithm to solve this kind of problem
- The choice of the algorithm depends on the assumptions about $f, g_1, \ldots, g_m$
- Even in the simplest cases, they are generally several approaches that can be used

Local and Global Optimum

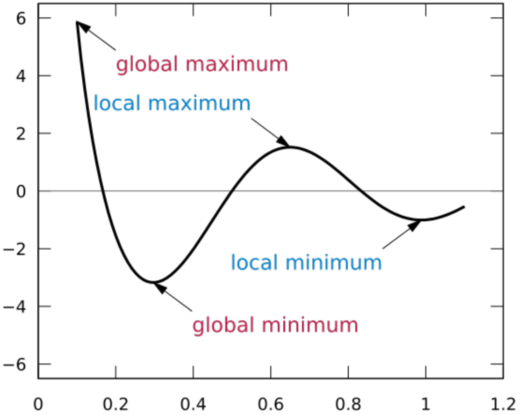
Let Y be the set of points satisfying all the constraints of the previous problem

- $\bar{x}$ is a local minimum (resp. a local maximum) if it exists ϵ for which $f(\bar{x}) \leq (\geq) f(x) \quad \forall x$ in Y such that $\|x - \bar{x}\| \leq \epsilon$
- $\bar{x}$ is a global minimum (resp. a global maximum) if $f(\bar{x}) \leq (\geq) f(x)$ $\forall x$ in Y

Important Remark

Most of the algorithms **converge** to a **local** optimum. Without strong assumptions about the objective function and its constraints, it is generally very difficult to get a global optimum

Local Versus Global Optimum



Local and global maxima and minima for $\cos(3\pi x)/x, 0.1 \leq x \leq 1.1$

Unconstrained Optimization

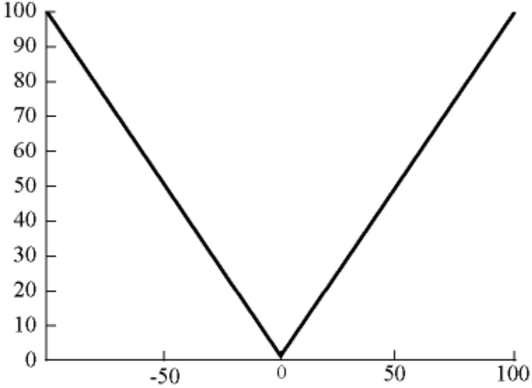
We consider the following problem $\mathcal{P}$:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})$$

and we assume that f is at least twice continuously differentiable on $\mathbb{R}^n$

- A twice continuously differentiable function means that we can compute the gradient $\nabla f(x)$ and the second partial derivatives and that these functions are still continuous
- Concretely, it means that the objective function is **smooth**
- In most of the applications that we consider in this course, this assumption is satisfied

Example of a Non-Smooth Function



This function is continuous but not differentiable at $x = 0$

Critical Point

- A critical point or a stationary point of a differentiable function $\mathbb{R}^n \to \mathbb{R}$ is a point where its gradient is null
- We remind that the gradient of f at a point $\boldsymbol{x} = (x_1, \ldots, x_n)$ is the vector $\nabla f(\boldsymbol{x})$ of its partial derivatives:

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

- A saddle point is a criticial point which is not a maximum, nor a minimum
- **A critical point can be a local minimum, a local maximum, or a saddle point**

Example of a Saddle Point

A saddle point (in red) on the graph of $z = x^2 - y^2$:

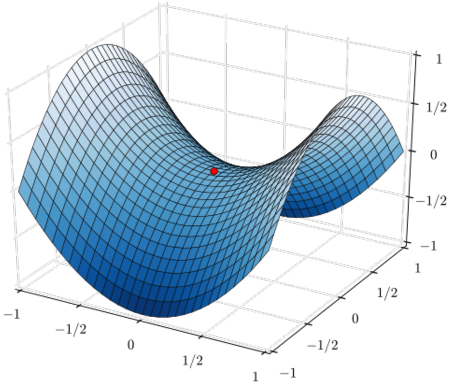


Image source: https://commons.wikimedia.org/w/index.php?curid=20570051

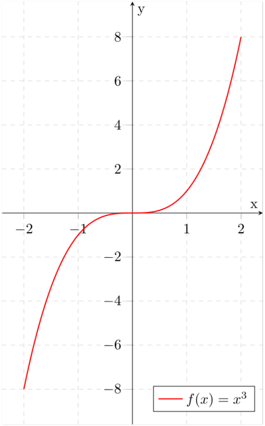
Conditions for Optimality for Unconstrained Optimization

Theorem (First Order Necessary Conditions for Optimality)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at a point $\bar{\boldsymbol{x}} \in \mathbb{R}^n$. If $\bar{\boldsymbol{x}}$ is a local optimum, then $\nabla f(\bar{\boldsymbol{x}}) = 0$

Obviously, this is not a sufficient condition ! If $\nabla f(\bar{\boldsymbol{x}}) = 0$, then we cannot conclude that it is a local optimum

$\nabla f(\bar{x}) = 0$ is Not a Sufficient Condition



$f'(0) = 0$ but 0 is not an optimum !

Hessian Matrix

Let $f : \mathbb{R}^n \to \mathbb{R}$ a twice differentiable function. The function denoted by $\nabla^2 f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ is the hessian matrix of f at $\boldsymbol{x}$ and is defined by:

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

This is a **symmetric** matrix

Example

Let's consider the following function:

$$f(x_1, x_2) = 50x_1^2 - x_2^3$$

Its gradient is given by:

$$\nabla f(x_1, x_2) = \begin{pmatrix} 100x_1 \\ -3x_2^2 \end{pmatrix}$$

and its hessian is:

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 100 & 0 \\ 0 & -6x_2 \end{pmatrix}$$

Definiteness of Symmetric Matrices

- A $n \times n$ symmetric real matrix $\boldsymbol{Q}$ is said to be positive definite (resp. resp. negative definite) if the scalar $\boldsymbol{z}^T\boldsymbol{Q}\boldsymbol{z}$ is > 0 (resp. < 0) for every non-null vector $\boldsymbol{z}$ of n real numbers
- A $n \times n$ symmetric real matrix $\boldsymbol{Q}$ is said to be positive semi-definite (resp. negative semi-definite) if the scalar $\boldsymbol{z}^T\boldsymbol{Q}\boldsymbol{z}$ is ≥ 0 (resp. ≤ 0) for every vector $\boldsymbol{z}$ of n real numbers
- Positive and negative definite matrices are always **invertible**. This is not the case for positive and negative semi-definite matrices

Examples

- positive definite matrices:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

- positive semi-definite matrices:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$$

- negative definite matrices:

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}$$

- negative semi-definite matrices:

$$\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix}$$

A Positive Semi-Definite Matrix

Let's show that the following matrix is positive semi-definite:

$$\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$$

We compute

$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2(x_1 + x_2)^2$$

The latter expression is always non-negative and is null when $x_2 = -x_1$.
We conclude that this matrix is **positive semi-definite**

A Negative Definite Matrix

Let's show that the following matrix is negative definite:

$$\begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}$$

We compute

$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -2x_1^2 + 2x_1 x_2 - 2x_2^2$$

Then:

$$-2x_1^2 + 2x_1 x_2 - 2x_2^2 = -x_1^2 - (x_1 - x_2)^2 - x_2^2 < 0 \quad \text{if } x_1, x_2 \neq 0.$$

We conclude that this matrix is **negative definite**

Characterizations

Important Result

A $n \times n$ **symmetric** matrix $\boldsymbol{Q}$ has always n **real** eigenvalues

- A symmetric positive definite matrix has n eigenvalues > 0
- A symmetric positive semi-definite matrix has n eigenvalues ≥ 0
- A symmetric negative definite matrix has n eigenvalues < 0
- A symmetric negative semi-definite matrix has n eigenvalues ≤ 0

Diagonal Matrices

The eigenvalues of a **diagonal** matrix lie on its diagonal. So, a diagonal matrix is

- positive semi-definite if and only if all of its diagonal elements are ≥ 0

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

- positive definite if and only if all of its diagonal elements are > 0

$$\begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$$

- negative semi-definite if and only if all of its diagonal elements are ≤ 0

$$\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$$

- negative definite if and only if all of its diagonal elements are < 0

$$\begin{pmatrix} -4 & 0 \\ 0 & -1 \end{pmatrix}$$

Conditions for Optimality for Unconstrained Optimization

Theorem (Second Order Optimality Conditions)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable at a point $\bar{\mathbf{x}} \in \mathbb{R}^n$

(1) *(Necessity) If $\bar{\mathbf{x}}$ is a local minimum (resp. local maximum), then $\nabla f(\bar{\mathbf{x}}) = 0$ and $\nabla^2 f(\bar{\mathbf{x}})$ is **positive semi-definite** (resp. **negative semi-definite**)*

(2) *(Sufficiency) If $\nabla f(\bar{\mathbf{x}}) = 0$ and $\nabla^2 f(\bar{\mathbf{x}})$ is **positive definite** (resp. **negative definite**) , then $\bar{\mathbf{x}}$ is a local minimum (resp. local maximum)*

Example (Cont'd)

The gradient of the following function:

$$f(x_1, x_2) = 50x_1^2 - x_2^3$$

is given by:

$$\nabla f(x_1, x_2) = \begin{pmatrix} 100x_1 \\ -3x_2^2 \end{pmatrix}$$

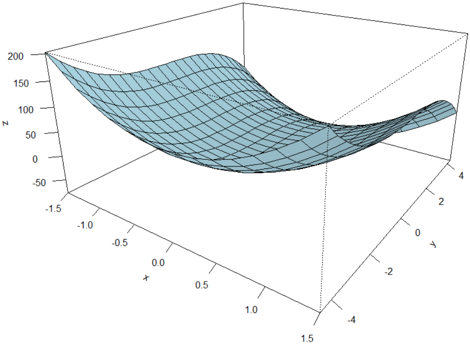
and its hessian is:

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 100 & 0 \\ 0 & -6x_2 \end{pmatrix}$$

The point $\boldsymbol{p} = (0, 0)$ is a critical point and the hessian at $\boldsymbol{p}$ is a **positive semi-definite matrix**. The value of the objective function at $\boldsymbol{p}$ is 0

Example (Cont'd)

But $\boldsymbol{p} = (0, 0)$ is not a **local optimum** !



This just shows that having a point with a null gradient and with an hessian which is positive **semi-definite** is not enough to have a local minimum

Example (Cont'd)

To proove that $\boldsymbol{p}$ ia not a local optimum, let's choose a direction $\boldsymbol{d} = (0 \;\; 1)^T$ and let's move from $\boldsymbol{p}$ into that direction with a step $\alpha > 0$. Then:

$$0 = f(0, 0) > f(0, \alpha) = \text{-}\alpha^3$$

and we conclude that $\boldsymbol{p}$ is not a local minimum. On the other hand, if we consider the following direction $\boldsymbol{d} = (0 \;\; \text{-}1)^T$ with a step $\alpha > 0$, then we get:

$$0 = f(0, 0) < f(0, \text{-}\alpha) = \alpha^3$$

and $\boldsymbol{p}$ is not a local maximum neither. Indeed, this is a **saddle** point

Quadratic Function

A function $f : \mathbb{R}^b \to \mathbb{R}$ is called quadratic if it can be written as

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{g}^T \boldsymbol{x} + c,$$

where $\boldsymbol{Q}$ is a $n \times n$ symmetric matrix, $\boldsymbol{g} \in \mathbb{R}^n$ and $c \in \mathbb{R}$

Important Result

The gradient and the hessian of a quadratic function:

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{g}^T \boldsymbol{x} + c,$$

are given by:

$$\nabla f(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{g}, \text{ and } \nabla^2 f(\boldsymbol{x}) = \boldsymbol{Q}$$

Quadratic Function (Con't)

- It is not restrictive to assume that $\boldsymbol{Q}$ is symmetric. If it is not the case, we can simply replace $\boldsymbol{Q}$ by the symmetric matrix $\frac{1}{2}(\boldsymbol{Q}+\boldsymbol{Q}^T)$
- Explanation:

$$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} = (\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x})^T = \boldsymbol{x}^T\boldsymbol{Q}^T\boldsymbol{x} = \frac{1}{2}(\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{Q}^T\boldsymbol{x}) = \boldsymbol{x}^T\frac{1}{2}(\boldsymbol{Q}+\boldsymbol{Q}^T)\boldsymbol{x}$$

Convex and Strictly Convex Function

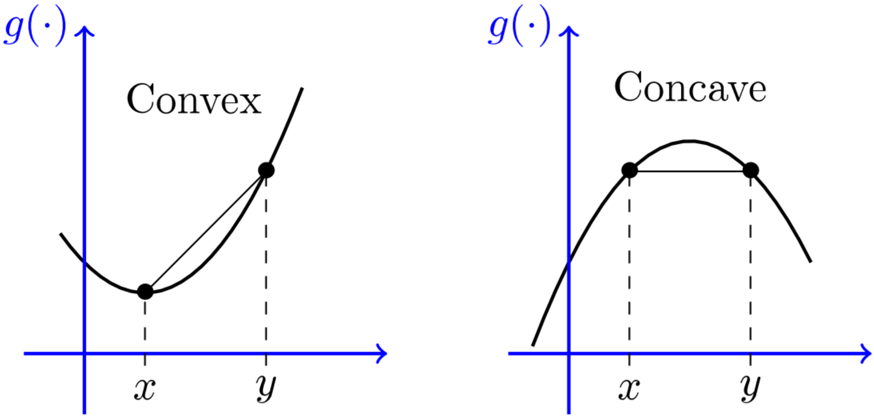
- A function $f : \mathbb{R}^n \to \mathbb{R}$ is called **convex** (resp. concave) if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and for all $\lambda \in [0,1]$ we have

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq (resp. \geq)\ \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$$

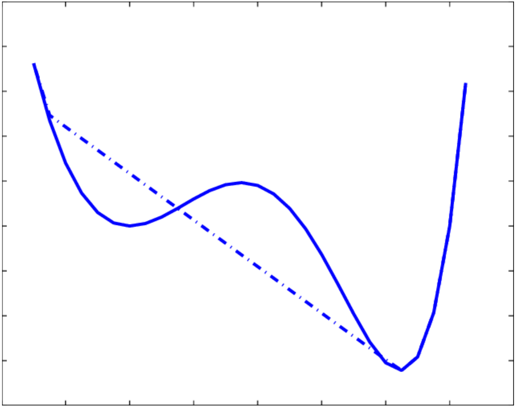
- A function $f : \mathbb{R}^n \to \mathbb{R}$ is called **strictly convex** (resp. strictly concave) if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, x \neq y$, and for all $\lambda \in]0,1[$ we have

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) < (resp. >)\ \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$$

Convex Versus Concave



Non-Convex and Non-Concave



Example of a non-convex and non-concave function

Characterization of Convex/Strictly Convex Functions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function. Then

- f is **convex if and only if** its hessian is **positive semidefinite** for all $\boldsymbol{x} \in \mathbb{R}^n$
- f is **strictly convex if** the hessian of f is **positive definite** for all $\boldsymbol{x} \in \mathbb{R}^n$
- f is **concave if and only if** its hessian is **negative semidefinite** for all $\boldsymbol{x} \in \mathbb{R}^n$
- f is **strictly concave if** the hessian of f is **negative definite** for all $\boldsymbol{x} \in \mathbb{R}^n$

Unconstrained Optimization: Sufficient Conditions for a Global Optimum

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function and $\mathbf{x}^* \in \mathbb{R}^n$ a local minimum of f. If f is convex, then $\mathbf{x}^*$ is a global minimum of f. Moreover, if f is strictly convex, $\mathbf{x}^*$ is the unique global minimum of f

The concept of convexity is **essential** in optimization. When the objective function is not convex, it is often difficult to identify a global optimum

Optimality Conditions for a Quadratic Function - 1

Let's consider the following problem:

$$min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{g}^T\boldsymbol{x} + c,$$

where $\boldsymbol{Q}$ is a $n \times n$ symmetric matrix, $\boldsymbol{g} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. The behavior of a quadratic function depends on the eigenvalues of its hessian $\boldsymbol{Q}$. Since $\boldsymbol{Q}$ is symmetric, it is known that $\boldsymbol{Q}$ has n real eigenvalues

$$\mu_1 \leq \mu_2 \leq \ldots \mu_n,$$

which are associated with n **orthonormal** (mutually orthogonal with unit norm) eigenvectors $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_n$

Optimality Conditions for a Quadratic Function - 2

Important Result

There are four cases:

(i) If $\mu_1 > 0$, then $\boldsymbol{Q}$ is a positive definite matrix, f is strictly convex and its unique minimizer is the unique solution of $\boldsymbol{Qx} = -\boldsymbol{g}$

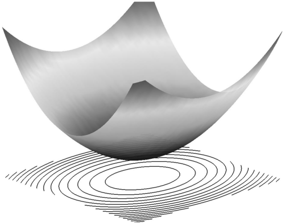
(ii) If $\mu_1 < 0$, then $f(\boldsymbol{x}) \to -\infty$ along the direction $\boldsymbol{\nu}_1$ and the problem has no finite solution

Consider now the cases in which there are null eigenvalues. Let them be $\mu_1 = \mu_2 = \ldots \mu_k$. Thus $\boldsymbol{Q}$ is a positive semi-definite matrix

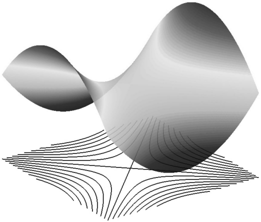
(iii) If $\boldsymbol{g}^T \boldsymbol{\nu}_i = 0$ for $i = 1, ..., k$ then f has a k-dimensional set of minimizers

(iv) If $\boldsymbol{g}^T \boldsymbol{\nu}_i < 0$ for some $i = 1, ..., k$ then f is unbounded below and the problem has no finite solution

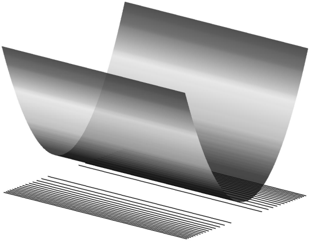
Optimality Conditions for a Quadratic Function - 3



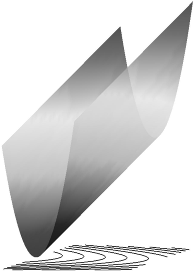
Case (i)



Case (ii)



Case (iii)



Case (iv)

Optimality Conditions for a Quadratic Function - 4

Important Result

So, when $\boldsymbol{Q}$ is symmetric positive defintie, then the solution of

$$min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{g}^T\boldsymbol{x} + c$$

is unique and is simply given by the solution of the following linear system $\boldsymbol{Q}\boldsymbol{x} = -\boldsymbol{g}$

Application: OLS Regression

We would like to determine $\hat{\boldsymbol{\beta}}$ such that it minimizes the square errors of the following linear system

$$\boldsymbol{y} = \boldsymbol{A}\beta + \boldsymbol{\epsilon},$$

where $\boldsymbol{y}, \boldsymbol{\epsilon} \in \mathbb{R}^m$, $\boldsymbol{Y}$ is a $m \times n$ matrix and $\boldsymbol{\beta} \in \mathbb{R}^n$. Moreover, we assume that $m > n$. Concretely, we want to compute the vector $\hat{\boldsymbol{\beta}}$ that minimizes

$$\|\boldsymbol{y} - \boldsymbol{A}\beta\|^2$$

OLS Regression

Let's develop the following expression:

$$\|\boldsymbol{y} - \boldsymbol{A}\hat{\beta}\|^2 = (\boldsymbol{y} - \boldsymbol{A}\hat{\beta})^T(\boldsymbol{y} - \boldsymbol{A}\hat{\beta}) = \hat{\beta}^T\boldsymbol{A}^T\boldsymbol{A}\hat{\beta} - 2\boldsymbol{y}^T\boldsymbol{A}\hat{\beta} + \boldsymbol{y}^T\boldsymbol{y}$$

This is a **quadratic** function in $\hat{\beta}$. **First order optimal conditions** (the gradient is null) imply that

$$2\boldsymbol{A}^T\boldsymbol{A}\hat{\beta} - 2\boldsymbol{A}^T\boldsymbol{y} = 0 \iff \boldsymbol{A}^T\boldsymbol{A}\hat{\beta} = \boldsymbol{A}^T\boldsymbol{y}$$

Assuming that $\boldsymbol{A}^T\boldsymbol{A}$ is positive definite, then there is a unique solution to this problem given by

$$\hat{\beta} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{y}$$

Optimization with Equality Constraints

General formulation:

$$\begin{array}{rrcl} & \multicolumn{3}{c}{\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})} \\ & g_1(\boldsymbol{x}) & = & b_1 \\ \text{s.t.} & & \vdots & \\ & g_m(\boldsymbol{x}) & = & b_m \end{array}$$

Question

Is it possible to transform the objective function f to integrate all the constraints into one function called the lagrangian so that we can apply standard calculus on it ? Only partially. Indeed, all the local minima of f are **stationary** points of the lagrangian. This approach is called the method of Lagrange multipliers

Lagrange Multipliers: Informal Presentation

- First form the lagrangian function $L(\boldsymbol{x}, \boldsymbol{\lambda})$. L is our new objective function corresponding to f augmented by the addition of the constraint functions

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{j=1}^{m} \lambda_j (g_j(\boldsymbol{x}) - b_j)$$

- Each constraint function is multiplied by a variable $\lambda_j \in \mathbb{R}$, called a lagrange multiplier

Lagrange Multipliers: Informal Presentation (Cont'd)

- The Lagrange function effectively transforms a problem in n variables and m constraints into an unconstrained optimization problem with $n + m$ variables
- If $\bar{\boldsymbol{x}}$ is a local minimum of the original constrained problem, then there exists a vector $\bar{\boldsymbol{\lambda}}$ such that $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}})$ is a **stationary** point for the Lagrange function
- However, not all stationary points yield a solution of the original problem. Thus, the method of Lagrange multipliers yields a **necessary** condition for optimality in constrained problems provided a mild assumption known as **constraint qualification** is satisfied
- Sufficient conditions for a minimum or maximum also exist but at the cost of additional assumptions

Lagrange Multipliers: Informal Presentation (Cont'd)

The **first order conditions** on $L(\boldsymbol{x}, \boldsymbol{\lambda})$ give a system of $n + m$ equations:

$$\frac{\partial L(\boldsymbol{x})}{\partial x_i} = \frac{\partial f(\boldsymbol{x})}{\partial x_i} + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j(\boldsymbol{x})}{\partial x_i} = 0, \quad i = 1, \ldots, n$$

$$\frac{\partial L(\boldsymbol{x})}{\partial \lambda_i} = g_j(\boldsymbol{x}) - b_j = 0, \quad j = 1, \ldots, m$$

Question

Are the first order conditions sufficient to guarantee that the solution of this system is a **local minimum** ?

No, it isn't ! Indeed, the first-order conditions are sufficient conditions **only in particular cases**

Example: Maximizing the Area of a Rectangle

Problem

We would like to find the dimensions x and y of the sides of a rectangle so that its area is maximized but its perimeter remains equal to p

Formulation:

$$\begin{array}{ll} \max_{x,y} & A(x,y) = xy \\ s.t. & 2(x+y) - p = 0 \end{array} \iff \begin{array}{ll} -\min_{x,y} & A(x,y) = -xy \\ s.t. & 2(x+y) - p = 0 \end{array}$$

Example: Maximizing the Area of a Rectangle

The lagrange function is defined by:

$$L(x, y, \lambda) = -xy + \lambda(2(x + y) - p)$$

First order conditions are:

$$\begin{aligned}
\frac{\partial L(x, y, \lambda)}{\partial x} &= -y + 2\lambda &= 0\\
\frac{\partial L(x, y, \lambda)}{\partial y} &= -x + 2\lambda &= 0\\
\frac{\partial L(x, y, \lambda)}{\partial \lambda} &= 2x + 2y - p &= 0
\end{aligned}$$

Example: Maximizing the Area of a Rectangle (Cont'd)

Solving the system yields $x = p/4, y = p/4$, and $\lambda = p/8$. With these values, the area $A(x, y)$ of the rectangle is $p^2/16$.

Question

Are we sure that it is really a maximum and, if it is the case, a global maximum ?

Example: Maximizing the Area of a Rectangle (Cont'd)

Let's call x the shorter side, y the longest one, and let $\epsilon \in [0, p/4]$. One can parametrize any rectangle satisfying the perimeter constraint as:

$$x = p/4 - \epsilon \quad \text{and } y = p/4 + \epsilon$$

The area of such rectangle is

$$A(x, y) = (p/4 - \epsilon)(p/4 + \epsilon) = p^2/16 - \epsilon^2 \leq p^2/16$$

We have a strict inequality as soon as $\epsilon > 0$. We conclude that $x = y = p/4$ is the **unique global** maximizer of the problem !

Quadratic Programming

We consider the following quadratic problem $\mathcal{Q}$:

$$\begin{aligned} min_{\boldsymbol{x}\in\mathbb{R}^n} \; f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{g}^T\boldsymbol{x} + c \\ s.t. \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned}$$

where $\boldsymbol{Q}$ is a symmetric $n \times n$ matrix and $\boldsymbol{A}$ a $m \times n$ matrix of full rank with $m \leq n$

Quadratic Programming (Cont'd)

Its lagrangian function is given by:

$$\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{g}^T\boldsymbol{x} + c + \boldsymbol{\lambda}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$$

First order optimal conditions imply that:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{g} + \boldsymbol{A}^T\boldsymbol{\lambda} = 0 \quad \text{and} \quad \nabla_{\boldsymbol{\lambda}} L(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = 0$$

These equations can be rewritten as:

$$\begin{pmatrix} \boldsymbol{Q} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{g} \\ \boldsymbol{b} \end{pmatrix} \tag{1}$$

This is a linear system with $n + m$ variables and $n + m$ equations

Positive Definite Quadratic Programming

Theorem

Let's assume that $\boldsymbol{Q}$ is positive definite. Then Problem $\mathcal{Q}$ has a unique global minimzer $\boldsymbol{x}^*$ given by the solution of the linear system (1)

The solution of the system of equations (1) is given by:

$$\boldsymbol{x}^* = -\boldsymbol{Q}^{-1}(\boldsymbol{A}^T\boldsymbol{\lambda}^* + \boldsymbol{g}) \text{ and } \boldsymbol{\lambda}^* = -(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{g} + \boldsymbol{b})$$

However, in practice, we don't use this formula except when the size of the problem is small because the inversion of a matrix is computationally intensive. A factorization is generally prefered to solve this system

Lagrange Duality

We now consider the general case:

$$\begin{aligned} min_{\boldsymbol{x} \in \mathbb{R}^n} \quad & f(\boldsymbol{x}) \\ s.t. \quad & g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \dots, k \\ & h_i(\boldsymbol{x}) = 0, \quad i = 1, \dots, l \end{aligned}$$

We assume that the functions f, g_i, h_i are continuously differentiable. This is the primal problem $\mathcal{P}$. Its lagrangian is defined by

$$L(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = f(\boldsymbol{x}) + \sum_{i=1}^{k} \alpha_i g_i(\boldsymbol{x}) + \sum_{i=1}^{l} \beta_i h_i(\boldsymbol{x}).$$

The α_i and β_i are called lagrange multipliers

Lagrange Duality (Cont'd)

Consider the quantity

$$\theta_{\mathcal{P}} = \max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\beta}} L(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta})$$

- If $g_i(\boldsymbol{x}) < 0$, then $max_{\alpha_i} \alpha_i g_i(\boldsymbol{x}) = 0 \quad \forall \alpha_i \geq 0$
- If $g_i(\boldsymbol{x}) = 0$, then $\alpha_i g_i(\boldsymbol{x}) = 0 \quad \forall \alpha_i \geq 0$
- If $g_i(\boldsymbol{x}) > 0$, then $max_{\alpha_i \geq 0} \alpha g_i(\boldsymbol{x}) = +\infty$
- If $h_i(\boldsymbol{x}) = 0$, then $\beta_i h_i(\boldsymbol{x}) = 0$
- If $h_i(\boldsymbol{x}) \neq 0$, then $max_{\beta_i} \beta_i h_i(\boldsymbol{x}) = +\infty$

This simply shows that :

$$\theta_{\mathcal{P}}(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & \text{if } \boldsymbol{x} \text{ satisfies primal constraints} \\ +\infty & \text{otherwise} \end{cases}$$

Lagrange Duality (Cont'd)

- So, if we consider

$$min_{\mathbf{x}} \theta_{\mathcal{P}}(\mathbf{x}) = min_{\mathbf{x}} max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\beta}} L(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}),$$

  then this is the same problem as the original one but formulated without any constraint

- However, this formulation is not very helpful... The objective function is not even continuous !

- A more interesting approach is based on the **dual** formulation !

Lagrange Duality (Cont'd)

We define $\theta_{\mathcal{D}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ as :

$$\theta_{\mathcal{D}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta})$$

$\mathcal{D}$ stands for "dual". The dual optimization problem is defined by:

$$max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\beta}} \theta_{\mathcal{D}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

It can be easily shown that

$$d^* = max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\beta}} \theta_{\mathcal{D}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leq min_{\mathbf{x}} \theta_{\mathcal{P}}(\mathbf{x}) = p^*$$

However, under certains conditions, we can have $d^* = p^*$ so that we can **solve the dual problem in place of the original problem**

Lagrange Duality (Cont'd)

Theorem

*Suppose that f and the g_i's are convex, and that h_i are affine. Suppose further that there exists some feasible $\mathbf{x}$ so that $g_i(\mathbf{x}) < 0$, $\forall i$. Then there exists $\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$ so that $\mathbf{x}^*$ is the solution to the primal problem $\mathcal{P}$,$\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$ are the solutions of the dual problem, and $p^* = d^* = L(\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$. Moreover, $\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$ satisfy the **Karush-Kuhn-Tucker (KKT)** conditions given by:*

$$\begin{aligned}
\frac{\partial}{\partial x_i} L(\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) &= 0, \quad i = 1, \ldots, n \\
\frac{\partial}{\partial \beta_i} L(\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) &= 0, \quad i = 1, \ldots, l \,(\iff h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, l) \\
\alpha_i^* g_i(\mathbf{x}^*) &= 0, \quad i = 1, \ldots, k \text{ (dual complementary cond.)} \\
g_i(\mathbf{x}^*) &\leq 0, \quad i = 1, \ldots, k \\
\alpha_i^* &\geq 0, \quad i = 1 \ldots, k
\end{aligned}$$

Moreover, if some $\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$ satisfy the KKT conditions, then they are also a solution to the primal and dual problems