R401: Statistical and Mathematical Foundations

Lecture 14: Unconstrained Optimization. Static Optimization with Equality Constraints. Lagrange Multipliers

Andrey Vassilev

2016/2017

General Principles and Caveats for the Optimization Module

- Emphasis on practicality over rigour
- Consequently, algorithmic approach and “recipes” rather than proofs
- Also, existence and relevant properties of various objects are often implicitly assumed
- Pathologies and mathematical peculiarities discussed only in special cases

Lecture Contents

1. Warm-up: Basic Unconstrained Optimization in $\mathbb{R}^1$
2. Unconstrained Optimization in $\mathbb{R}^n$
3. Static Optimization with Equality Constraints. Lagrange Multipliers

Warm-up: Basic Unconstrained Optimization in $\mathbb{R}^1$

Fact 1

For a function $f : \mathbb{R} \to \mathbb{R}$ differentiable at a point x, a necessary condition for a local extreme point (i.e. a maximum or a minimum) at x is

$$f'(x) = 0.$$

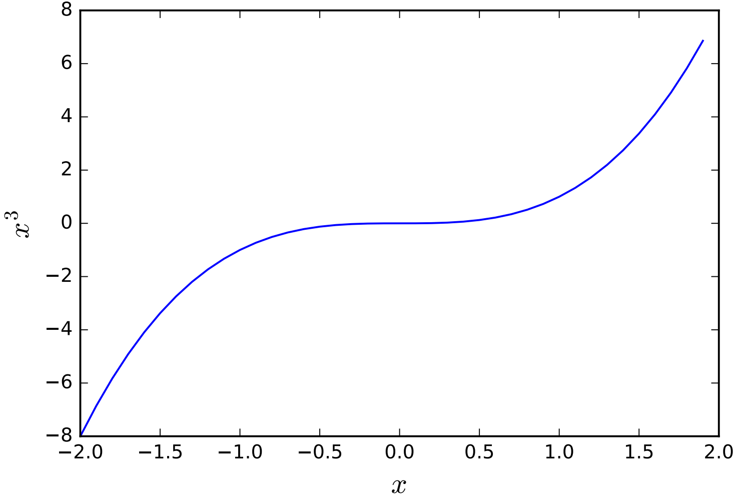
Example 1

If $f(x) = ax^2 + bx + c$, then $f'(x) = 2ax + b$ and the condition $f'(x) = 0$ yields the familiar $x = -\frac{b}{2a}$ (recall your high-school days). Depending on the sign of a, this is a maximum or a minimum (What is the relationship?).

Example 2

If $f(x) = x^3$, then $f'(x) = 3x^2$ and $f'(x) = 0 \Rightarrow x = 0$.
Does the function attain a maximum or a minimum at $x = 0$?

Warm-up: Basic Unconstrained Optimization in $\mathbb{R}^1$



Warm-up: Basic Unconstrained Optimization in $\mathbb{R}^1$

Example 2 (cont.)

The answer is "neither"! The point $x = 0$ is not a local extreme point of $f(x) = x^3$.

This illustrates the pitfalls of using necessary conditions – they supply only candidates that need to be checked further.

The above examples generalize in the following manner:

Fact 2

Let a function f be n times differentiable at a point x and

$$f'(x) = f''(x) = \ldots = f^{(n-1)}(x) = 0, \qquad f^{(n)} \neq 0.$$

1. If n is odd, the point x is not an extreme point of $f(x)$.
2. If n is even and $f^{(n)}(x) > 0$, the point x is a minimum.
3. If n is even and $f^{(n)}(x) < 0$, the point x is a maximum.

Unconstrained Optimization in $\mathbb{R}^n$

Necessary conditions

Fact 3

For a function $f : \mathbb{R}^n \to \mathbb{R}$, differentiable at a point $\mathbf{x}$, a necessary condition for $\mathbf{x}$ to be a local extreme point is

$$f'(\mathbf{x}) = \mathbf{0},$$

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \; \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ and } f'(\mathbf{x}) = \begin{pmatrix} \dfrac{\partial f(x_1, \ldots, x_n)}{\partial x_1} \\ \vdots \\ \dfrac{\partial f(x_1, \ldots, x_n)}{\partial x_n} \end{pmatrix}$$

Note: A point where the gradient of a function f vanishes is called a *critical point* or a *stationary point*. This also applies to functions on $\mathbb{R}^1$.

Unconstrained Optimization in $\mathbb{R}^n$

Example 3

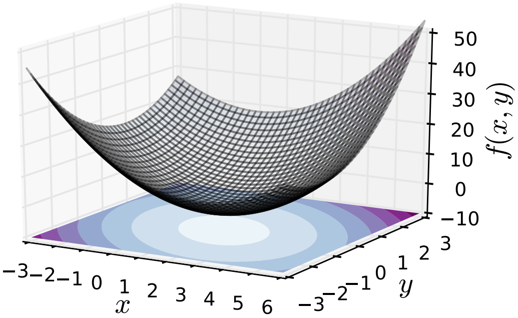
$$f(x,y) = x^2 + 2y^2 - 3x + xy$$

$$\frac{\partial f}{\partial x} = 2x - 3 + y = 0 \quad \Rightarrow \quad x = \frac{3-y}{2}$$

$$\frac{\partial f}{\partial y} = 4y + x = 0 \quad \Rightarrow \quad y = -\frac{x}{4}$$

$$x = \frac{12}{7},\ y = -\frac{3}{7}$$

Unconstrained Optimization in $\mathbb{R}^n$



Unconstrained Optimization in $\mathbb{R}^n$

The necessity of the condition $f'(\mathbf{x}) = \mathbf{0}$ has implications that are similar to the univariate case:

Example 4

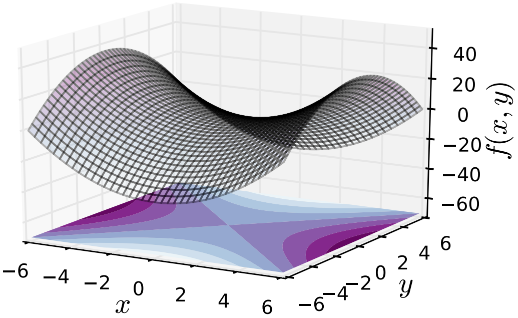
Consider the function $f(x, y) = x^2 - y^2$. The NCs yield the following candidate:

$$\frac{\partial f}{\partial x} = 2x = 0 \quad \Rightarrow \quad x = 0,$$

$$\frac{\partial f}{\partial y} = -2y = 0 \quad \Rightarrow \quad y = 0.$$

Let's look at the graph of the function in a neighbourhood of the point $(0, 0)'$.

Unconstrained Optimization in $\mathbb{R}^n$



Unconstrained Optimization in $\mathbb{R}^n$

Example 4 (cont.)

The critical point $\mathbf{x} = (0, 0)'$ is an example of a *saddle point*. The function f (obviously) does not attain an extremum at $\mathbf{x}$.

Example 4 illustrates the need to refine the approach for checking candidate points in the n-dimensional case. To this end, we have to review several concepts.

A symmetric square matrix A is called *positive semidefinite* if, for any vector $\mathbf{x}$, we have

$$\mathbf{x}' A\mathbf{x} \geq 0.$$

If the inequality is strict for any non-zero vector $\mathbf{x}$, the matrix is called *positive definite*.
Similarly, a symmetric square matrix A is called *negative semidefinite* if, for any vector $\mathbf{x}$, we have $\mathbf{x}' A\mathbf{x} \leq 0$, and *negative definite* in case of strict inequality for $\mathbf{x} \neq \mathbf{0}$.

Unconstrained Optimization in $\mathbb{R}^n$

Incidentally, for a given square symmetric matrix A, the function $Q(\mathbf{x}) = \mathbf{x}' A \mathbf{x}$ is called a *quadratic form*. Quadratic forms are also referred to as "positive/negative (semi)definite", depending on the properties of the respective matrix.

Recall that, for an $n \times n$ matrix A, a *principal minor* of order k ($1 \leq k \leq n$), denoted by Δ_k, is the determinant of the submatrix obtained by deleting $n - k$ rows of the matrix and the correspondingly numbered columns, e.g.

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & \cdots & a_{n,n} \end{bmatrix}$$

Note: The notation Δ_k does not identify a unique principal minor of order k.

Unconstrained Optimization in $\mathbb{R}^n$

Incidentally, for a given square symmetric matrix A, the function $Q(\mathbf{x}) = \mathbf{x}' A \mathbf{x}$ is called a *quadratic form*. Quadratic forms are also referred to as "positive/negative (semi)definite", depending on the properties of the respective matrix.

Recall that, for an $n \times n$ matrix A, a *principal minor* of order k ($1 \leq k \leq n$), denoted by Δ_k, is the determinant of the submatrix obtained by deleting $n - k$ rows of the matrix and the correspondingly numbered columns, e.g.

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & \cdots & a_{n,n} \end{bmatrix}$$

Note: The notation Δ_k does not identify a unique principal minor of order k.

Unconstrained Optimization in $\mathbb{R}^n$

The k-th *leading principal minor* of a matrix A ($1 \leq k \leq n$), denoted by D_k, is the determinant of the submatrix

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1} & a_{k,2} & \cdots & a_{k,k} \end{bmatrix},$$

i.e. the principal minor obtained by deleting the last $n-k$ rows and columns and, respectively, keeping the first k.

Unconstrained Optimization in $\mathbb{R}^n$

Fact 4 (Sylvester's criterion)

Let A be a symmetric matrix. Then:

1. A is positive definite if and only if $D_k > 0,\ k = 1, \ldots, n$.
2. A is positive semidefinite if and only if $\Delta_k \geq 0$ for all principal minors of order $k = 1, \ldots, n$.
3. A is negative definite if and only if $(-1)^k D_k > 0,\ k = 1, \ldots, n$.
4. A is negative semidefinite if and only if $(-1)^k \Delta_k \geq 0$ for all principal minors of order $k = 1, \ldots, n$.

Note that the necessary and sufficient conditions for "semidefiniteness" involve all principal minors (and hence are cumbersome to check), not just the leading principal minors.

Unconstrained Optimization in $\mathbb{R}^n$

Let a function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ be twice differentiable. The matrix of second partial derivatives, evaluated at a point $\mathbf{x}$, i.e.

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

is called the *Hessian (matrix)* of f at $\mathbf{x}$.

Unconstrained Optimization in $\mathbb{R}^n$

- The Hessian is denoted $\mathbf{f}''(\mathbf{x})$.
- The Hessian is symmetric.
- Sometimes the partial derivative $\dfrac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$ is written as $f''_{ij}(\mathbf{x})$.
- A leading principal minor of order k of the Hessian is denoted $D_k(\mathbf{x})$.
- An arbitrary principal minor of order k of the Hessian is denoted $\Delta_k(\mathbf{x})$.

Unconstrained Optimization in $\mathbb{R}^n$

Fact 5

Let a (twice) differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ have a critical point at $\mathbf{x}^*$.

1. If the Hessian $\mathbf{f}''(\mathbf{x}^*)$ is positive definite or, equivalently, $D_k(\mathbf{x}^*) > 0,\ k = 1, \ldots, n$, then $\mathbf{x}^*$ is a *local minimum point*.
2. If the Hessian $\mathbf{f}''(\mathbf{x}^*)$ is negative definite or, equivalently, $(-1)^k D_k(\mathbf{x}^*) > 0,\ k = 1, \ldots, n$, then $\mathbf{x}^*$ is a *local maximum point*.
3. If $D_n(\mathbf{x}^*) \neq 0$ and neither 1) nor 2) is satisfied, then $\mathbf{x}^*$ is a *saddle point*.

Unconstrained Optimization in $\mathbb{R}^n$

Fact 6

Let a (twice) differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ have an extreme point at $\mathbf{x}^*$.

1. If $\mathbf{x}^*$ is a local minimum point, then the Hessian $\mathbf{f}''(\mathbf{x}^*)$ is positive semidefinite or, equivalently, $\Delta_k(\mathbf{x}^*) \geq 0$ for all principal minors of order $k = 1, \ldots, n$.
2. If $\mathbf{x}^*$ is a local maximum point, then the Hessian $\mathbf{f}''(\mathbf{x}^*)$ is negative semidefinite or, equivalently, $(-1)^k \Delta_k(\mathbf{x}^*) \geq 0$ for all principal minors of order $k = 1, \ldots, n$.

Unconstrained Optimization in $\mathbb{R}^n$

Example 5 (Verification of Example 3)

Recall that:

$$f(x,y) = x^2 + 2y^2 - 3x + xy$$

$$\frac{\partial f}{\partial x} = 2x - 3 + y, \quad \frac{\partial f}{\partial y} = 4y + x.$$

We now have:

$$\frac{\partial^2 f}{\partial x^2} = 2, \quad \frac{\partial^2 f}{\partial y^2} = 4, \quad \frac{\partial^2 f}{\partial x \partial y} = 1, \quad \frac{\partial^2 f}{\partial y \partial x} = 1.$$

$$D_1 = \det[2] = 2 > 0, \quad D_2 = \det \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} = 2 \cdot 4 - 1 \cdot 1 = 7 > 0.$$

Since $D_1 > 0,\ D_2 > 0,$ the critical point $x = \frac{12}{7},\ y = -\frac{3}{7}$ is a minimum.

Homework: Apply the same procedure to Example 4.

Static Optimization with Equality Constraints

Formulation

Now we look at problems of the form

$$f(x_1, \ldots, x_n) \to \min(\max) \tag{1}$$

s.t.

$$\begin{aligned} g_1(x_1, \ldots, x_n) &= b_1 \\ g_2(x_1, \ldots, x_n) &= b_2 \\ \cdots \\ g_m(x_1, \ldots, x_n) &= b_m \end{aligned} \tag{2}$$

where $m < n$. (Can you explain the last requirement?)

Note: In what follows, all required properties of the objects in (1) and (2) like differentiability are implicitly assumed.

Static Optimization with Equality Constraints

Formulation

Using vector notation for compactness, the objective function is:

$$f(\mathbf{x}) \to \min(\max)$$

We introduce

$$\mathbf{g}(\mathbf{x}) := [g_1(\mathbf{x}), \ldots, g_m(\mathbf{x})]', \quad \mathbf{b} = [b_1, \ldots, b_m]'$$

and the constraints are written as

$$\mathbf{g}(\mathbf{x}) = \mathbf{b}.$$

Static Optimization with Equality Constraints

The Lagrangian

The standard approach to solving (1)-(2) starts by defining a *Lagrangian*:

$$\mathcal{L}(\mathbf{x}) = f(\mathbf{x}) - \lambda_1(g_1(\mathbf{x}) - b_1) - \cdots - \lambda_m(g_m(\mathbf{x}) - b_m).$$

The numbers $\lambda_1, \ldots, \lambda_m$ are called *Lagrange multipliers*.

This can also be written in vector notation:

$$\mathcal{L}(\mathbf{x}) = f(\mathbf{x}) - \boldsymbol{\lambda}'(\mathbf{g}(\mathbf{x}) - \mathbf{b}),$$

where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_m]'$ is the vector of Lagrange multipliers.

Static Optimization with Equality Constraints

We can use the Lagrangian to produce necessary conditions for optimality in the following manner:

Algorithm

1. Form the Lagrangian as above
2. Differentiate it w.r.t. the variables we are optimizing over, i.e.

$$\frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial f(\mathbf{x})}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial g_i(\mathbf{x})}{\partial x_i},\ i = 1, \ldots, n$$

3. Set the resulting derivatives equal to zero, i.e.

$$\frac{\partial \mathcal{L}}{\partial x_i} = 0,\ i = 1, \ldots, n$$

4. The equations in the preceding step, together with the constraints (2), form a system of $n + m$ equations which is solved for the unknowns x_i and λ_j

Static Optimization with Equality Constraints

Remarks

- Sometimes the Lagrangian is equivalently formulated as

$$\mathcal{L}(\mathbf{x}) = f(\mathbf{x}) - \lambda_1 g_1(\mathbf{x}) - \cdots - \lambda_m g_m(\mathbf{x}).$$

  It obviously makes no difference as to the result of the differentiation step.
- Let the algorithm yield a candidate $\mathbf{x}^*$. Roughly, if the Lagrangian is convex in $\mathbf{x}$, then the candidate $\mathbf{x}^*$ is a minimum. If the Lagrangian is concave in $\mathbf{x}$, then the candidate $\mathbf{x}^*$ is a maximum. (See SHSS, p. 117, for the precise formulation.)
- A Lagrange multiplier is interpreted as a *shadow price*, i.e. the gain (or loss) arising from relaxing the associated constraint.