Background: linear, quadratic and nonlinear programming

Outline

Nonlinear optimization theory
- Convexity and quadratic programming
- The optimality conditions: the KKT conditions

Exercises

Learning objectives (1/2)

At the end of this chapter you will be able to:

- recognise a convex problem and use the fact that every local minimum is global;
- decide whether a QP is convex from the matrix $\boldsymbol{Q}$;

Learning objectives (2/2)

At the end of this chapter you will be able to:

- write the KKT conditions of a problem and use them to check a solution;
- read the multipliers as shadow prices and estimate them by perturbation.

Prerequisites

- From the module *Convex Optimization*, Chapter 3 «Convex and concave functions»: convex and concave functions and their second-order characterisation
- From the module *Convex Optimization*, Chapter 4 «Quadratic functions»: quadratic functions and positive semidefinite matrices
- From the module *LP Modelling*, Chapter 2 «Background: linear programming and sensitivity analysis»: LP duality, complementary slackness and shadow prices

What this chapter is about

- We collect the **theoretical tools** of the course: convexity, quadratic programming, KKT conditions
- Linear programming theory and the sensitivity protocol: LP Modelling
- Convex sets and functions: Convex Optimization

Nonlinear optimization theory

Introduction: Nonlinear optimization

- We extend the tools of LP to **quadratic** and **nonlinear** problems
- **Convexity** is the property that allows us to certify the optimum
- The **KKT conditions** generalize complementary slackness
- The KKT multipliers generalize the shadow prices: we will still have an economic reading

Convex problems

- A problem $\min f(x_1, x_2, \ldots, x_n)$ over a feasible set X is **convex** if
  - f is convex
  - X is convex

- Let us see the property we shall use all the time

Why convexity matters

- In a convex problem every local minimum is global: optimality can be *certified*
- In non-convex problems (for instance a pricing problem with a bilinear term $p \cdot q$) a local method guarantees only a local optimum
- Non-convex QPs can nonetheless be solved to global optimality with dedicated techniques, but the computational cost grows

Quadratic programming

- A **QP** has objective

$$\frac{1}{2}\sum_i \sum_j q_{ij}\, x_i x_j + \sum_j c_j x_j$$

  and linear constraints
- It is convex if and only if the matrix $\boldsymbol{Q} = (q_{ij})$ is positive semidefinite ($\boldsymbol{Q} \succeq 0$)
- In this course QPs appear every time a cost or a risk grows more than linearly with the decision

Example: a 2×2 QP (1/3)

A minimization QP with a separable quadratic objective (the second variable "costs" twice as much as the first, at the margin) and a threshold constraint on the sum:

$$\begin{array}{rlll} \min & x_1^2 + 2\,x_2^2 & & \\ \text{subject to} & x_1 + \quad x_2 & \geq & 6 \\ & x_1, \quad\; x_2 & \geq & 0 \end{array}$$

Here $\boldsymbol{Q} = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \succ 0$: the QP is convex and the optimum found is global.

Example: a 2×2 QP (2/3)

Step 1 — solving we obtain:

$$\tilde{x}_1 = 4, \qquad \tilde{x}_2 = 2, \qquad \tilde{f} = 24, \qquad \lambda = 8$$

Step 2 — verification of the shadow price.

- The constraint is active ($4 + 2 = 6$): the objective grows with the variables, hence the sum does not exceed what is needed
- The partial derivatives coincide:

$$2\tilde{x}_1 = 8 = 4\tilde{x}_2 = \lambda \quad \checkmark$$

- If they were different it would be worth shifting quantity from one variable to the other; the common derivative is precisely the multiplier

Example: a 2×2 QP (3/3)

Step 3 — verification by perturbation.

With right-hand side 7 we obtain $(\tilde{x}_1, \tilde{x}_2) = (14/3,\ 7/3)$ with

$$\tilde{f} = 32.67 \approx 24 + 8$$

- On top of that there is the curvature term
- Here the shadow price is exact only at the margin, not for finite increments

The Lagrangian

- The optimality conditions of constrained nonlinear problems are known as the **KKT conditions** (after Karush, Kuhn and Tucker)
- Problem: $\min f(x_1, \ldots, x_n)$ subject to
  - $g_i(x_1, \ldots, x_n) \leq 0$ for every $i \in \{1, 2, \ldots, m\}$
  - $h_j(x_1, \ldots, x_n) = 0$ for every $j \in \{1, 2, \ldots, q\}$
- The **Lagrangian** is

$$L = f(x_1, \ldots, x_n) + \sum_{i=1}^{m} \lambda_i \, g_i(x_1, \ldots, x_n) + \sum_{j=1}^{q} \nu_j \, h_j(x_1, \ldots, x_n), \qquad \lambda_i \geq 0$$

The KKT conditions

KKT conditions

At a regular optimal point $(\tilde{x}_1, \ldots, \tilde{x}_n)$ there exist multipliers $\lambda_i \geq 0$ and ν_j such that

$$\frac{\partial L}{\partial x_k}(\tilde{x}_1, \ldots, \tilde{x}_n) = 0 \quad \forall k \in \{1, 2, \ldots, n\} \quad \text{(stationarity)},$$

$$\lambda_i \, g_i(\tilde{x}_1, \ldots, \tilde{x}_n) = 0 \quad \forall i \in \{1, 2, \ldots, m\} \quad \text{(complementarity)}.$$

If the problem is convex, the KKT conditions are also *sufficient*: a point that satisfies them is a global optimum.

The KKT conditions are complementary slackness, generalized

- The two parts of the chapter mirror each other: for an LP the KKT conditions reduce *exactly* to complementary slackness
- The stationarity of the Lagrangian is equivalent to the feasibility of the dual:
  - the multipliers λ_i are the dual variables π_i
  - the multipliers of the sign constraints are the reduced costs
- The complementarity $\lambda_i \, g_i(\tilde{x}_1, \ldots, \tilde{x}_n) = 0$ is the complementary slackness condition $\bar{\pi}_i \cdot \bar{s}_i = 0$
- Complementary slackness is the KKT conditions in the linear case

Example: the KKT conditions on the QP (1/3)

Let us take up again the QP just solved: min $x_1^2 + 2x_2^2$ subject to $x_1 + x_2 \geq 6$, $x_1, x_2 \geq 0$.

Step 1. The constraint in standard form:
$g(x_1, x_2) = 6 - x_1 - x_2 \leq 0$

Step 2. Stationarity of the Lagrangian
$L = x_1^2 + 2x_2^2 + \lambda(6 - x_1 - x_2)$:

$$\frac{\partial L}{\partial x_1} = 2x_1 - \lambda = 0, \qquad \frac{\partial L}{\partial x_2} = 4x_2 - \lambda = 0 \quad \Rightarrow \quad x_1 = \frac{\lambda}{2},\ x_2 = \frac{\lambda}{4}$$

(the constraints $x_1, x_2 \geq 0$ will turn out to be inactive: their multipliers are null by complementarity)

Example: the KKT conditions on the QP (2/3)

Step 3. Complementarity:

- if the threshold constraint were inactive ($\lambda = 0$) we would have $x_1 = x_2 = 0$, which violates $x_1 + x_2 \geq 6$: impossible
- hence the constraint is active:

$$\frac{\lambda}{2} + \frac{\lambda}{4} = 6 \;\Rightarrow\; \lambda = 8, \quad \tilde{x}_1 = 4, \quad \tilde{x}_2 = 2, \quad \tilde{f} = 24$$

- the same numbers found by elementary means, now obtained with the general procedure

Example: the KKT conditions on the QP (3/3)

Step 4. Reading of the multiplier:

- with right-hand side $6 + \varepsilon$ the optimal value grows by about $\lambda\varepsilon = 8\varepsilon$
- verification: with right-hand side d the solution is $\tilde{x}_1 = 2d/3$, $\tilde{x}_2 = d/3$, hence

$$\tilde{f}(d) = \tfrac{2}{3}d^2 \;\Rightarrow\; \tilde{f}(6+\varepsilon) = 24 + 8\varepsilon + \tfrac{2}{3}\varepsilon^2 \quad \checkmark$$

Multipliers and shadow prices

- The KKT multipliers generalize the shadow prices of the LP: they measure the marginal variation of the optimal value with respect to the right-hand side of the constraint
- When the duals are not available (general nonlinear models) we estimate the multiplier *by perturbation*:
  - the right-hand side is increased by ε
  - the problem is re-optimized
  - the incremental ratio is computed
- This numerical verification is itself a recurring exercise

Summary: Nonlinear optimization

Convexity:

- In a convex problem every local minimum is global
- A QP is convex if and only if $\boldsymbol{Q} \succeq 0$

KKT conditions:

- Stationarity of the Lagrangian and complementarity
- Sufficient in convex problems
- In the linear case they are complementary slackness
- The multipliers are shadow prices; they can also be estimated by perturbation

Exercise: KKT

Solve with the KKT conditions:

$$\min\ (x-3)^2+(y-1)^2 \quad \text{subject to} \quad x+2y \le 2, \quad x,y \ge 0$$

- Say which constraints are active at the optimum
- Interpret the multipliers

Solution (1/2)

- The point $(3, 1)$ is not feasible; its projection onto the line $x + 2y = 2$ is $(2.4, -0.2)$, with $y < 0$: hence $y \geq 0$ will be active too
- Candidate: $x + 2y = 2$, $y = 0 \Rightarrow (\tilde{x}, \tilde{y}) = (2, 0)$, $\tilde{f} = 1 + 1 = 2$
- Lagrangian: $L = (x - 3)^2 + (y - 1)^2 + \lambda (x + 2y - 2) - \mu y$

$$\frac{\partial L}{\partial x} = 2(x - 3) + \lambda = -2 + \lambda = 0 \qquad \Rightarrow \ \lambda = 2$$

$$\frac{\partial L}{\partial y} = 2(y - 1) + 2\lambda - \mu = -2 + 4 - \mu = 0 \qquad \Rightarrow \ \mu = 2$$

Solution (2/2)

- $\lambda = 2 \geq 0$, $\mu = 2 \geq 0$; the constraint $x \geq 0$ is not active ($\tilde{x} = 2$) and has null multiplier: the KKT conditions hold
- The problem is convex: $(2, 0)$ is the global optimum
- Active: $x + 2y \leq 2$ and $y \geq 0$
- $\lambda = 2$: relaxing the constraint to $x + 2y \leq 2 + \varepsilon$ the optimal value decreases by about 2ε
- $\mu = 2$: if $y \geq -\varepsilon$ were allowed the optimal value would decrease by about 2ε

Exercise: convexity

For which values of the parameter γ is the function

$$f(x_1, x_2) = x_1^2 + \gamma x_1 x_2 + x_2^2$$

convex?

(Write $\boldsymbol{Q}$ and impose $\boldsymbol{Q} \succeq 0$.)

Solution

- With $f = \frac{1}{2} \sum_i \sum_j q_{ij}\, x_i x_j$:

$$\boldsymbol{Q} = \begin{pmatrix} 2 & \gamma \\ \gamma & 2 \end{pmatrix}$$

- A symmetric 2×2 matrix is positive semidefinite if and only if the diagonal entries and the determinant are non-negative:

$$2 \geq 0, \qquad \det \boldsymbol{Q} = 4 - \gamma^2 \geq 0$$

- Hence f is convex if and only if $-2 \leq \gamma \leq 2$

Final summary (1/2)

Nonlinear optimization:

- Convexity: every local minimum is global
- A QP is convex if and only if $\boldsymbol{Q} \succeq 0$
- KKT conditions: complementary slackness generalized

Final summary (2/2)

In the lab:

- Multipliers are verified by perturbation, with a small increment
- In convex problems the KKT conditions certify the global optimum