

Lecture 5 - Quadratic programming

Dinesh Krishnamoorthy

Department of Engineering Cybernetics, NTNU

Spring 2026

Quadratic programming (QP)

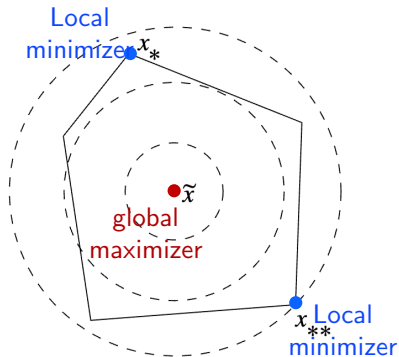
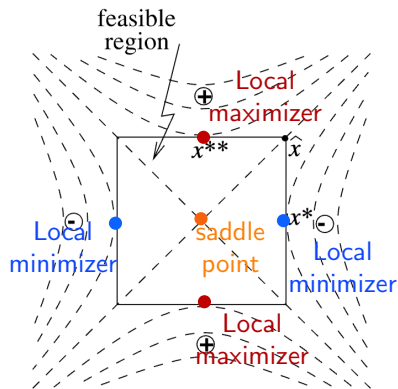
An optimization problem is an QP problem if $f(x)$ is quadratic and $g(x), h(x)$ are affine

$$\begin{aligned} \min_x f(x) &:= \frac{1}{2}x^\top Gx + c^\top x \\ \text{s.t. } h(x) &:= A_{eq}x - b_{eq} = 0, \\ g(x) &:= Ax - b \leq 0 \end{aligned}$$

- ▶ This QP is **convex** if $G \succeq 0$, and strictly convex if $G \succ 0$.
- ▶ Nonconvex QP if G is indefinite - can be more challenging because they have several stationary points and local minima.

Example: Nonconvex QP

Nonconvex QP if G is indefinite - can be more challenging because they have several stationary points and local minima.



Why convex QPs are interesting?

- ▶ It is the “easiest” nonlinear programming problem, with efficient solvers
- ▶ QP is the building block for SQP (sequential quadratic problem) - used to solve general nonlinear programs (NLP)
- ▶ QPs are often used in control - Model Predictive Control (MPC)
- ▶ Also used in finance (Portfolio optimization), Machine learning (Mean square loss minimization)

Data fitting (Empirical Risk Minimization)

- ▶ Given the data set $\mathcal{D} := \{(u_i, y_i)\}_{i=1}^N$
- ▶ Choose a parametric model structure $y = g(u, \mathbf{x})$
- ▶ Find the parameter $\mathbf{x} \in \mathbb{R}^p$ that gives the best fit

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \ell(\mathbf{x}; \mathcal{D})$$

Data fitting: Linear Regression

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{N} \sum_{i=1}^N \|y_i - g(\mathbf{x}, u_i)\|_2^2$$

- ▶ $\mathbf{x} = [m, c]^\top$
- ▶ $g(u_i, \mathbf{x}) = mu_i + c = \phi_i^\top \mathbf{x}$
- ▶ $\phi_i := [u_i, 1]^\top$
- ▶ $\ell(\mathbf{x}; \mathcal{D}) := \frac{1}{N} \sum_{i=1}^N (y_i - \phi_i^\top \mathbf{x})^2$

This is Least Squares Estimate (LSE), which has an analytical solution.

$$\hat{\mathbf{x}} = \left[\sum_{i=1}^N \phi_i \phi_i^\top \right]^{-1} \sum_{i=1}^N \phi_i y_i$$

Least Squares Estimate: Analytical Solution

$$\min_{\mathbf{x}} \frac{1}{N} \sum_{i=1}^N (y_i - \phi_i^\top \mathbf{x})^2$$

► First order necessary condition[†]:

$$\frac{2}{N} \sum_{i=1}^N \phi_i (y_i - \phi_i^\top \mathbf{x}) = 0 \Rightarrow \frac{1}{N} \sum_{i=1}^N \phi_i \phi_i^\top \mathbf{x} = \frac{1}{N} \sum_{i=1}^N \phi_i y_i$$

$$\Rightarrow \hat{\mathbf{x}} = \left[\frac{1}{N} \sum_{i=1}^N \phi_i \phi_i^\top \right]^{-1} \frac{1}{N} \sum_{i=1}^N \phi_i y_i$$

[†]**Question:** Is this also a sufficient condition? why?

Yes, since this is a convex unconstrained optimization problem.

Least squares estimation

Note that:

$$\min_x \frac{1}{N} \sum_{i=1}^N \|r_i\|_2^2 \neq \min_x \sum_{i=1}^N \|r_i\|_2^2 \neq \min_x \frac{1}{2} \sum_{i=1}^N \|r_i\|_2^2$$

but

$$\arg \min_x \frac{1}{N} \sum_{i=1}^N \|r_i\|_2^2 = \arg \min_x \sum_{i=1}^N \|r_i\|_2^2 = \arg \min_x \frac{1}{2} \sum_{i=1}^N \|r_i\|_2^2$$

Least Squares Estimate: Analytical Solution

Analytical solution can also be conveniently written in matrix form:

$$Y := \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^N \quad \Phi := \begin{bmatrix} u_1 & 1 \\ u_2 & 1 \\ \vdots & \vdots \\ u_N & 1 \end{bmatrix} \in \mathbb{R}^{N \times p}$$

$$\min_{x \in \mathbb{R}^p} \|Y - \Phi x\|^2$$

► First order necessary condition:

$$\Phi^\top Y - \Phi^\top \Phi x = 0$$

$$\Rightarrow \hat{x} = [\Phi^\top \Phi]^{-1} \Phi^\top Y$$

Equality constrained QPs

$$\begin{aligned} \min_x f(x) &:= \frac{1}{2}x^\top Gx + c^\top x \\ \text{s.t. } h(x) &:= Ax - b = 0, \end{aligned}$$

$$\text{Lagrangian: } \mathcal{L}(x, \lambda) = \frac{1}{2}x^\top Gx + c^\top x + \lambda^\top (Ax - b)$$

KKT conditions

$$\begin{aligned} Gx^* + c + A^\top \lambda^* &= 0 \\ Ax^* - b &= 0 \end{aligned}$$

In matrix form:

$$\underbrace{\begin{bmatrix} G & A^\top \\ A & 0 \end{bmatrix}}_{\text{KKT Matrix}} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

Nonsingularity of KKT Matrix

Let A have full row rank, and assume that the reduced Hessian $Z^T G Z$ is positive definite, where $AZ = 0$. Then the KKT matrix

$$K = \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix}$$

is non-singular, and there is a unique vector pair (x^*, λ^*) satisfying

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

and x^* is the unique global solution of the QP.

1. A has full row rank $\iff$ LICQ
2. G is positive definite on the nullspace of A , i.e.,

$$\exists Z \neq 0, AZ = 0 \text{ s.t. } Z^T G Z \succ 0$$

Condition 2 is exactly the strong second-order sufficient condition for the QP.

Why $\text{null}(A)$?

- ▶ Given some feasible point x , we want to find the descent direction p that gives $x^* = x + p$
- ▶ The feasible directions must satisfy the linearized constraint

$$Ap = 0$$

- ▶ This means, the only directions that matter for curvature are those in the nullspace of A .

$$\begin{aligned} \min_{p \in \mathbb{R}^n} \quad & \frac{1}{2}(x + p)^\top G(x + p) + c^\top(x + p) \\ \text{s.t.} \quad & \underbrace{A(x + p) - b = 0}_{Ap=0} \end{aligned}$$

Solving the KKT System: Direct Methods

Consider the KKT linear system for equality-constrained QP:

$$\begin{bmatrix} G & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}.$$

Direct (factorization-based) methods:

- ▶ Solve the full KKT system using a **sparse LDL^T** or **LDL^T with pivoting**.
- ▶ Schur complement reduction:

$$(AG^{-1}A^\top)\lambda = b + AG^{-1}c,$$

then recover

$$x = -G^{-1}(c + A^\top\lambda).$$

(block elimination, augmented system solves).

Pros: Robust, accurate, predictable.

Cons: Expensive for large-scale problems.

Solving the KKT System: Iterative Methods

Large-scale KKT systems are often solved using Krylov-based iterative solvers.

Preconditioning is essential.

- ▶ KKT matrices are indefinite, and often ill-conditioned → convergence issues
- ▶ A preconditioner transforms the system into

$$P^{-1} \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = P^{-1} \begin{bmatrix} -c \\ b \end{bmatrix}$$

such that $P^{-1}K \approx I$

Pros: Scales to very large problems; exploits sparsity.

Cons: Convergence depends on conditioning and preconditioning quality.

Inequality constrained QP

$$\min_x f(x) := \frac{1}{2}x^\top Gx + c^\top x$$

$$\text{s.t. } g(x) := Ax - b \leq 0$$

$$\mathcal{L}(x, \mu) = \frac{1}{2}x^\top Gx + c^\top x + \sum_{i \in \mathcal{I}} [\mu_i (a_i^\top x - b_i)]$$

KKT conditions

$$Gx^* + c + \sum_i \mu_i^* a_i = 0$$

$$a_i^\top x^* - b = 0, \quad \forall i \in \mathbb{A}(x^*)$$

$$a_i^\top x^* - b \leq 0, \quad \forall i \in \mathcal{I} \setminus \mathbb{A}(x^*)$$

$$\mu_i^* \geq 0, \quad \forall i \in \mathcal{I} \cap \mathbb{A}(x^*)$$

For convex QP (i.e., $G \succeq 0$), these are sufficient and necessary conditions.

Active-set methods

Key idea:

1. Start with a set of active constraints, known as working set $\mathcal{W}_k$
2. Solve equality constrained QP

$$\begin{aligned} \min_x \quad & \frac{1}{2} \mathbf{x}^\top \mathbf{G} \mathbf{x} + \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}_i^\top \mathbf{x} - b_i = 0, \quad \forall i \in \mathcal{W}_k \end{aligned}$$

3. Check KKT conditions
 - ▶ Terminate, if KKT ok!
 - ▶ Update the working set $\mathcal{W}_k$, and go to step 2.

The main challenge: Determination of the active set.

Active set methods

- ▶ Iteration k with current working set of active constraints $\mathcal{W}_k$, and current feasible estimate x_k
- ▶ Define $A_k = [\dots, a_i^\top, \dots]^\top, i \in \mathcal{W}_k$ and $b_k = [\dots, b_i, \dots]^\top, i \in \mathcal{W}_k$
- ▶ Solve the EQP to find $x^* = x_k + p$

$$\begin{aligned} \min_{p \in \mathbb{R}^n} \quad & \frac{1}{2}(x_k + p)^\top G(x_k + p) + c^\top(x_k + p) & \Leftrightarrow \min_{p \in \mathbb{R}^n} \quad & \frac{1}{2}p^\top Gp + (Gx_k + c)^\top p \\ \text{s.t.} \quad & A_k(x_k + p) - b_k = 0 & \text{s.t.} \quad & A_k p = 0 \end{aligned}$$

Active set strategy

- ▶ If solution $p_k = 0$, check if all Lagrange multipliers are non-negative
 - ▶ Terminate if yes.
 - ▶ Remove one constraint with negative Lagrange multiplier from $\mathcal{W}_k$
- ▶ If solution $p_k \neq 0$
 - ▶ Take full step $x = x + p_k$ if feasible w.r.t. constraints $i \notin \mathcal{W}_k$
 - ▶ Else $x = x + \alpha p_k$ where for some $i \notin \mathcal{W}_k$

$$\alpha_k \leq \frac{b_i - a_i^\top x_k}{a_i^\top p_k} < 1$$

Add the blocking constraint i to the working set $\mathcal{W}_k$

Active set strategy

- ▶ If $\mathbf{p}_k \neq 0$, are there any constraints that are blocking the full step

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{p}_k?$$

- ▶ Yes? $\rightarrow$ Add blocking constraint to $\mathcal{W}_k$
- ▶ No? $\rightarrow$ Keep same $\mathcal{W}_k$
- ▶ If $\mathbf{p}_k = 0$, determine Lagrange multipliers μ_A . Is $\mu_A < 0$?
 - ▶ Yes? $\rightarrow$ Delete inequality constraint with largest negative multiplier value from $\mathcal{W}_k$
 - ▶ No? $\rightarrow$ stop with $\mathbf{w}_{k+1} = \mathbf{w}_k$

Example

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & \frac{1}{2}(x_1 - 1)^2 + (x_2 - 3)^2 \\ \text{s.t.} \quad & -x_1 + x_2 \leq 0 \\ & -x_2 \leq 0 \end{aligned}$$

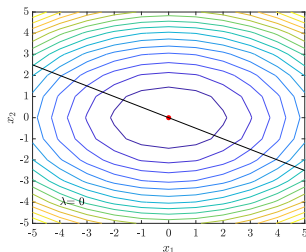
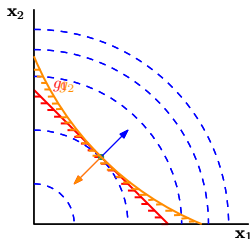
Solution: See lecture notes.

Active-set strategy

- ▶ When solving EQPs, update only the factorization that is needed
- ▶ Active set method converges in a finite number of iterations
- ▶ Potentially slow, unless you have good initial guess
- ▶ When using QP to solve MPC, we typically have good initial guesses!

Degeneracy

1. active constraint gradients are linearly dependent (i.e., LICQ fails)
2. Weakly active constraints with $\mu_i^* = 0$ for some $i \in \mathbb{A}(x^*)$ (i.e., strict complementarity fails)



Degeneracy can cause problems

- ▶ KKT matrices can become rank deficient
- ▶ Algorithms cannot decide whether to include the constraint or not, and can zig-zag between the active constraint sets.

Summary: Solving Quadratic Programs

- ▶ Equality-constrained QPs form the core subproblem in many nonlinear optimization algorithms (e.g., SQP).
- ▶ Their solutions are characterized by a linear KKT system involving the Hessian of the Lagrangian and the constraint Jacobian.
- ▶ Under LICQ and positive (semi)definiteness of the Hessian on the constraint nullspace, the KKT system has a unique solution.
- ▶ Active-set methods solve general QPs by iteratively guessing the set of active constraints and solving a sequence of equality-constrained QPs.
- ▶ At each iteration, constraints may be added or removed based on feasibility and Lagrange multiplier signs.
- ▶ The method terminates when the current iterate satisfies feasibility, stationarity, and complementary slackness.
- ▶ For convex QPs, active-set methods are globally convergent and finite-terminating.
- ▶ These ideas underpin modern SQP and MPC algorithms used in control engineering.