CS 524: Introduction to Optimization
Lecture 26 : Quadratic programs

Michael Ferris

Computer Sciences Department
University of Wisconsin-Madison

November 3, 2023

Quadratic forms

- **Linear functions:** sum of terms of the form $c_i x_i$ where the c_i are parameters and x_i are variables. General form:

$$c_1 x_1 + \cdots + c_n x_n = c^\mathsf{T} x$$

- **Quadratic functions:** sum of terms of the form $q_{ij} x_i x_j$ where q_{ij} are parameters and x_i are variables. General form:

$$q_{11} x_1^2 + q_{12} x_1 x_2 + \cdots + q_{nn} x_n x_n \qquad (n^2 \text{ terms})$$

$$= \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}^\mathsf{T} \begin{bmatrix} q_{11} & \dots & q_{1n} \\ \vdots & \ddots & \vdots \\ q_{n1} & \dots & q_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x^\mathsf{T} Q x$$

Quadratic forms

Example: $4x^2 + 6xy - 2yz + y^2 - z^2$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}^\mathsf{T} \begin{bmatrix} 4 & 6 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

In general: $$\begin{bmatrix} x \\ y \\ z \end{bmatrix}^\mathsf{T} \begin{bmatrix} 4 & p_2 & q_2 \\ p_1 & 1 & r_2 \\ q_1 & r_1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad \begin{cases} p_1 + p_2 = 6 \\ q_1 + q_2 = 0 \\ r_1 + r_2 = -2 \end{cases}$$

Symmetric: $$\begin{bmatrix} x \\ y \\ z \end{bmatrix}^\mathsf{T} \begin{bmatrix} 4 & 3 & 0 \\ 3 & 1 & -1 \\ 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Quadratic forms

Any quadratic function $f(x_1, \ldots, x_n)$ can be written in the form $x^\mathsf{T} Q x$ where Q is a symmetric matrix ($Q = Q^\mathsf{T}$).

Proof: Suppose $f(x_1, \ldots, x_n) = x^\mathsf{T} R x$ where R is *not* symmetric. Since it is a scalar, we can take the transpose:

$$x^\mathsf{T} R x = \left(x^\mathsf{T} R x\right)^\mathsf{T} = x^\mathsf{T} R^\mathsf{T} x$$

Therefore:

$$x^\mathsf{T} R x = \tfrac{1}{2}\left(x^\mathsf{T} R x + x^\mathsf{T} R^\mathsf{T} x\right) = x^\mathsf{T} \tfrac{1}{2}(R + R^\mathsf{T}) x$$

So we're done, because $\frac{1}{2}(R + R^\mathsf{T})$ is symmetric!

Gradients and Hessians

- Given a function $f : \mathbb{R}^n \to \mathbb{R}$

Gradient

- The multi-dimension analog of the derivative of a function
- The gradient of f, $\nabla(f) : \mathbb{R}^n \to \mathbb{R}^n$ is

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Hessian

- The multi-dimension analog of the second derivative of a function
- The Hessian of f, $\nabla^2(f) : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ is a matrix made up of all second partial derivatives

Examples

$$f(x) = \frac{1}{2}x^T Qx + c^T x + \alpha$$

$$\nabla f(x) = Qx + c, \;\; \nabla^2 f(x) = Q$$

- The following are some examples of convex functions:
- Exponential, x^a (when $a \geq 1$ or $a < 0$ on interval $(0, \infty)$), $|x|^p$, $-\log(x)$, negative entropy: $x \log x$, norms, max of linear, x^2/y, log(sum(exp)), geometric means
- Also there is a calculus that determines when compositions, etc are convex.
- When is a quadratic convex?

Eigenvalues and eigenvectors

If $A \in \mathbb{R}^{n\times n}$ and there is a vector v and scalar λ such that

$$Av = \lambda v$$

Then v is an **eigenvector** of A and λ is the corresponding **eigenvalue**. Some facts:

- The eigenvalues and eigenvectors can be complex.
- There are n (eigenvalue, eigenvector) pairs. Define: $V = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}$ and $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. Then:

$$AV = V\Lambda$$

- If there are n *linearly independent* eigenvectors, we say that A is **diagonalizable** and then V is invertible and:

$$A = V\Lambda V^{-1}$$

Eigenvalues and eigenvectors

If $A \in \mathbb{R}^{n \times n}$ is **symmetric**, then

- All eigenvalues of A are real.
- A is diagonalizable and its eigenvectors can be chosen to be orthonormal. So: $\|v_i\| = 1$ and $v_i^\mathsf{T} v_j = 0$ for $i \neq j$.
- We can write the factorization of A as:

$$A = V \Lambda V^\mathsf{T}$$

- Matrices that satisfy $V^{-1} = V^\mathsf{T}$ are called **orthogonal**. They represent isometries (preserve distances and angles).

Eigenvalue example

Consider the quadratic: $7x^2 + 4xy + 6y^2 + 4yz + 5z^2$.
A simple question: are there values that make this negative?
equivalent to:

$$\begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 7 & 2 & 0 \\ 2 & 6 & 2 \\ 0 & 2 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Eigenvalue decomposition:

$$\begin{bmatrix} 7 & 2 & 0 \\ 2 & 6 & 2 \\ 0 & 2 & 5 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 9 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ -\frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \end{bmatrix}$$

Eigenvalues are 3, 6, 9.

Recap

Question: Is $x^T Qx$ ever negative?
Answer: Look at the eigenvalues of Q:

- $Q = V\Lambda V^T$
- Define new coordinates $z = V^T x$.
- $x^T Qx = x^T V\Lambda V^t x = z^T \Lambda z = \lambda_1 z_1^2 + \cdots + \lambda_n z_n^2$

If all $\lambda_i \geq 0$, then $x^T Qx \geq 0$ for any x.

If some $\lambda_k < 0$, set $z_k = 1$ and all other $z_i = 0$. Then find corresponding x using $x = Vz$, and $x^T Qx < 0$.

Positive (semi)definite matrices

For a matrix $Q = Q^T$, the following are equivalent:

1. $x^T Qx \geq 0$ for all $x \in \mathbb{R}^n$
2. all eigenvalues of Q satisfy $\lambda_i \geq 0$

A matrix with the property is called positive semidefinite (PSD).
The notation is $Q \succeq 0$.

For a matrix $Q = Q^T$, the following are equivalent:

1. $x^T Qx > 0$ for all $0 \neq x \in \mathbb{R}^n$
2. all eigenvalues of Q satisfy $\lambda_i > 0$
3. The determinant of all leading principal minors of Q are > 0

A matrix with the property is called positive definite (PD).
The notation is $Q \succ 0$.

Positive definite matrices

Name	Definition	Notation
Positive semidefinite	all $\lambda_i \geq 0$	$Q \succeq 0$
Positive definite	all $\lambda_i > 0$	$Q \succ 0$
Negative semidefinite	all $\lambda_i \leq 0$	$Q \preceq 0$
Negative definite	all $\lambda_i < 0$	$Q \prec 0$
Indefinite	everything else	(none)

Some properties:

- If $P \succeq 0$ then $-P \preceq 0$
- If $P \succeq 0$ and $\alpha > 0$ then $\alpha P \succeq 0$
- If $P \succeq 0$ and $Q \succeq 0$ then $P + Q \succeq 0$
- Every $R = R^{\mathsf{T}}$ can be written as $R = P - Q$ for some appropriate choice of matrices $P \succeq 0$ and $Q \succeq 0$.

The Hessian

- Written out in full:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Characterization of Convexity (see Lecture 7)

- A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex on a domain $\mathcal{D}$ if and only if $\nabla^2 f(x)$ is positive semi-definite $\forall x \in \mathcal{D}$

Quadratic Programming

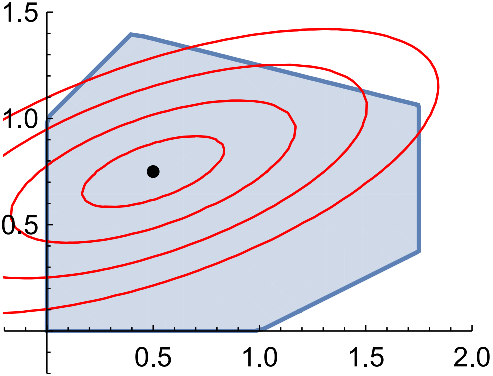
Simple description of QP:

$$\min\ f(x) := \frac{1}{2}x^T Qx + c^T x \ \text{ s.t. } \ Ax \leq b.$$

- Note from above that Q can be taken as symmetric.
- For the objective function $f(x) = \frac{1}{2}x^T Qx + c^T x$, the gradient is $\nabla f(x) = Qx + c$ (this formula valid only if Q is indeed symmetric.)
- Hessian is $\nabla^2 f(x) = Q$.
- If $Q \succeq 0$, it is a convex QP
  - feasible set is a polyhedron
  - contours of f are ellipsoids
  - solution can be on boundary or in the interior
  - relatively easy to solve (CPLEX, Gurobi, etc)
- If $Q \not\succeq 0$, it is very hard to solve in general (baron)

Quadratic programs

$$\min \; f(x) := \frac{1}{2}x^T Q x + c^T x \;\; \text{s.t.} \;\; Ax \le b.$$

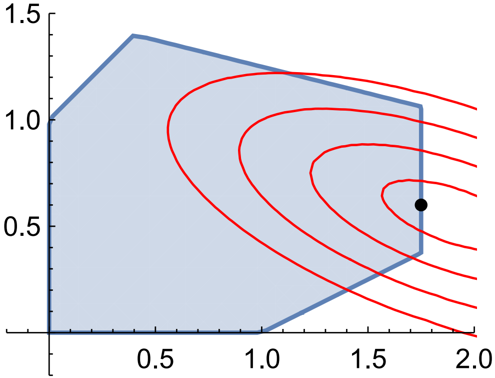


First case:
If the ellipsoid center is feasible, then it is also the optimal point.

11-25

Quadratic programs

$$\min\ f(x) := \frac{1}{2}x^T Qx + c^T x \ \text{ s.t. } \ Ax \leq b.$$



Second case:
If the ellipsoid center is infeasible, optimal point is on the boundary.
(not always at a vertex!)

11-26

Where do quadratics commonly occur?

1. As a regularization or penalty term
   - $(\text{cost}) + \lambda \|x\|^2$: standard L_2 regularizer
   - $(\text{cost}) + \lambda\, x^\mathsf{T} Q\, x$ (with $Q \succ 0$) : weighted L_2 regularizer
2. Hard norm bounds on a decision variable
   - $\|x\|^2 \le r$: a way to ensure that x doesn't get too big.
3. Allowing some tolerance in constraint satisfaction
   - $\|Ax - b\|^2 \le e$: we allow a tolerance e.
4. Energy quantities (physics/mechanics)
   - examples: $\frac{1}{2}mv^2$ (kinetic), $\frac{1}{2}kx^2$ (spring), $\frac{1}{2}CV^2$ (capacitor), $\frac{1}{2}I\omega^2$ (rotational), $\frac{1}{2}VE\varepsilon^2$ (strain).
5. Covariance constraints (statistics)

QCP: Quadratically constrained program

QCP

$$\color{red}{QCQP} \begin{cases} \min\limits_{x \in \mathbb{R}^n} & q_0(x) \\ \text{s.t} & q_k(x) \leq b_k \qquad \forall k \in \mathcal{I} \\ & q_k(x) = b_k \qquad \forall k \in \mathcal{E} \\ & x \geq 0 \end{cases}$$

- $q_k(x) = (c^k)^T x + x^T Q^k x \qquad \forall k \in \{0 \cup \mathcal{I} \cup \mathcal{E}\}$
- QCP is convex if $q_k(x)$ is a convex function of x $\forall k \in \{0 \cup \mathcal{I}\}$ and $q_k(x)$ is a linear function of x $\forall k \in \mathcal{E}$
- Of course, can have linear inequality constraints as well.

- Solve convex QCP using CPLEX, Gurobi, Mosek, Xpress, SCIP[1]
- These (commercial) solvers should be the fastest. Most also have *MIQCP* capabilities
- Nonconvex QCP can be attempted by conopt, knitro, baron and lindoglobal

[1] SCIP is very good, but not commercial