Quadratic Programming

GIAN Short Course on Optimization: Applications, Algorithms, and Computation

Sven Leyffer

Argonne National Laboratory

September 12-24, 2016

Outline

1. Introduction to Quadratic Programming
   - Applications of QP in Portfolio Selection
   - Applications of QP in Machine Learning

2. Active-Set Method for Quadratic Programming
   - Equality-Constrained QPs
   - General Quadratic Programs

3. Methods for Solving EQPs
   - Generalized Elimination for EQPs
   - Lagrangian Methods for EQPs

Introduction to Quadratic Programming

Quadratic Program (QP)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \tfrac{1}{2}x^T G x + g^T x \\ \text{subject to} \quad & a_i^T x = b_i \quad i \in \mathcal{E} \\ & a_i^T x \geq b_i \quad i \in \mathcal{I}, \end{aligned}$$

where

- $G \in \mathbb{R}^{n \times n}$ is a symmetric matrix
  ... can reformulate QP to have a symmetric Hessian
- $\mathcal{E}$ and $\mathcal{I}$ sets of equality/inequality constraints

Quadratic Program (QP)

- Like LPs, can be solved in finite number of steps
- Important class of problems:
  - Many applications, e.g. quadratic assignment problem
  - Main computational component of SQP:
    Sequential Quadratic Programming for nonlinear optimization

Introduction to Quadratic Programming

Quadratic Program (QP)

$$\begin{array}{lll} \underset{x}{\text{minimize}} & \frac{1}{2}x^T G x + g^T x & \\ \text{subject to} & a_i^T x = b_i & i \in \mathcal{E} \\ & a_i^T x \geq b_i & i \in \mathcal{I}, \end{array}$$

No assumption on eigenvalues of G

- If $G \succeq 0$ positive semi-definite, then QP is convex
  $\Rightarrow$ can find global minimum (if it exists)
- If G indefinite, then QP may be globally solvable, or not:
  - If $A_{\mathcal{E}}$ full rank, then $\exists Z_{\mathcal{E}}$ null-space basis
    Convex, if "reduced Hessian" positive semi-definite:

$$Z_{\mathcal{E}}^T G Z_{\mathcal{E}} \succeq 0, \quad \text{where} \quad Z_{\mathcal{E}}^T A_{\mathcal{E}} = 0 \qquad \text{then globally solvable}$$

... eliminate some variables using the equations

Introduction to Quadratic Programming

Quadratic Program (QP)

$$\begin{array}{lll} \underset{x}{\text{minimize}} & \frac{1}{2}x^TGx + g^Tx & \\ \text{subject to} & a_i^Tx = b_i & i \in \mathcal{E} \\ & a_i^Tx \geq b_i & i \in \mathcal{I}, \end{array}$$

- Feasible set may be empty ... use phase-I methods from LP.
- Feasible set can be unbounded $\Rightarrow$ QP may be unbounded ... detect during the line-search ... $G \succ 0$ implies boundedness
- Polyhedral feasible set ... but solution may not be at vertex:

$$\underset{x}{\text{minimize}} \; x^2 \quad \text{subject to} \; -1 \leq x \leq 1$$

Applications of QP in Portfolio Selection

Investment decisions across collection of financial assets (e.g. stocks)

- Return and risk on investment are unknown (random vars)
- Historical data provides
  - Expected rate of return of investment
  - Covariance of rates of returns for investments

Markowitz Investment Model

- Balances risk and return (multi-objective)
- Choose mix of investment
  - minimize risk (covariance)
  - subject to minimum expected return

  Goal: Find how much to invest in each asset

Simple model, there exist more sophisticated models

Applications of QP in Portfolio Selection

Problem Data

- n number of available assets
- r desired minimum growth of portfolio
- β available capital for investment
- m_i expected rate of return of asset i
- C covariance matrix of asset returns
  ... models correlation between assets

Problem Variables

- x_i amount of investment in asset i
- Assume $x_i \geq 0$ and $x_i \in \mathbb{R}$ real

Applications of QP in Portfolio Selection

Problem Objective

- Minimize risk of investment

$$\underset{x}{\text{minimize}} \quad x^T C x$$

Problem Constraints

- Minimum rate of return on investment

$$\sum_{i=1}^{n} m_i x_i \geq r$$

- Upper bound on total investment

$$\sum_{i=1}^{n} x_i \leq \beta$$

Applications of QP in Machine Learning

Least squares problem

- Solve system of equations with more equations than variables
- Classical problem in data fitting / regression analysis

$$\underset{x}{\text{minimize}} \ \|Ax - b\|_2^2$$

... dates back to Legendre (1805)

- Solution from normal equations or augmented system (preferred)

$$A^T A x = A^T b \qquad \Leftrightarrow \qquad \begin{bmatrix} 0 & A^T \\ A & -I \end{bmatrix} \begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} 0 \\ b \end{pmatrix}$$

- Writing least-squares as a QP:

$$\|Ax - b\|_2^2 = (Ax - b)^T (Ax - b) = x^T A^T A x - 2 b^T A x + b^T b$$

Applications of QP in Machine Learning

Snag: Least-squares solution, x, typically dense

Interested in sparse solution, x, with few nonzeros $\Rightarrow \ell_1$ norm

1. LASSO: least absolute shrinkage and selection operator

$$\underset{x}{\text{minimize}} \; \|Ax - b\|_2^2 \quad \text{subject to } \|x\|_1 \leq \tau$$

   - ℓ_1-norm constraint act like a "sparsifier"
   - Least-squares problem with limit on number of nonzeros

2. Regularized least-squares

$$\underset{x}{\text{minimize}} \; \|Ax - b\|_2^2 + \lambda \|x\|_1$$

   - Related to basis pursuit denoising
   - ℓ_1-norm penalty act like a "sparsifier"

Writing LASSO as QP Problem

LASSO: least absolute shrinkage and selection operator

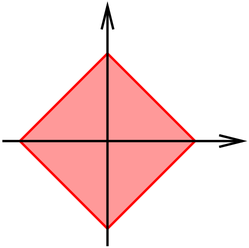
$$\underset{x}{\text{minimize}} \ \|Ax - b\|_2^2 \quad \text{subject to } \|x\|_1 \leq \tau$$

Recall ℓ_1 norm: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$

- v_i all 2^n vectors of $+1, -1$
  $v_0 = (1, \ldots, 1)$, $v_1 = (-1, 1, \ldots, 1)$, etc
- LASSO equivalent to exponential QP

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \|Ax - b\|_2^2 \\ \text{subject to} \quad & v_i^T x \leq \tau, \ \forall i \end{aligned}$$

... QP with 2^n constraints



Regularized Least-Squares as QP Problem

Regularized least-squares

$$\underset{x}{\text{minimize}} \;\; \|Ax - b\|_2^2 \; + \lambda \|x\|_1$$

- Introduce variables x_i^+, x_i^- for positive/negative part of x_i
- Then it follows that

$$x_i = x_i^+ - x_i^-, \quad |x_i| = x_i^+ + x_i^-, \quad x_i^+ \geq 0, x_i^- \geq 0$$

- Regularized least-squares equivalent to

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \|Ax - b\|_2^2 + \lambda \left(e^T x^+ + e^T x^- \right) \\ \text{subject to} \quad & x = x^+ - x^- \\ & x^+ \geq 0, \; x^- \geq 0 \end{aligned}$$

where $e = (1, \ldots, 1)$

Outline

Active-Set Method for Quadratic Programming

Quadratic Programming Problem (QP)

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & \tfrac{1}{2}x^T G x + g^T x \\
\text{subject to } & a_i^T x = b_i \quad i \in \mathcal{E} \\
& a_i^T x \geq b_i \quad i \in \mathcal{I},
\end{aligned}
$$

Active-Set Method for QPs

- Create sequence of (feasible) iterates $x^{(k)}$
- Fix active constraints, $\mathcal{W} \subset \mathcal{A}(x^{(k)})$
  - Solve equality-constrained QP
  - Either prove optimality, or find descend direction
- Update active set.

... first consider QPs with equality constraints only

Equality-Constrained QPs

Wlog assume solution, x^*, exists (other cases easily detected)

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{2}x^T G x + g^T x \\ \text{subject to} & A^T x = b, \end{array}$$

where

- Columns of matrix $A \in \mathbb{R}^{n \times m}$ are a_i for $i \in \mathcal{E}$
- Assume $m \leq n$ and A has full rank
  $\Rightarrow$ which implies that unique multipliers exist

QPs have meaningful solutions even for equality-constraints

- If $G \succeq 0$ positive semi-definite $\Rightarrow$ x^* global solution
- If $G \succ 0$ positive definite $\Rightarrow$ x^* is unique

Equality-Constrained QPs

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad \tfrac{1}{2}x^T G x + g^T x \\ &\text{subject to } A^T x = b, \end{aligned}$$

A full rank $\Rightarrow$ partition x and A:

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix},$$

where $x_1 \in \mathbb{R}^m$, $A_1 \in \mathbb{R}^{m \times m}$ nonsingular

$$\text{Then} \quad A^T x = b \quad \Leftrightarrow \quad A_1^T x_1 + A_2^T x_2 = b$$

A full rank $\Rightarrow$ A_1^{-T} exists ... eliminate x_1:

$$x_1 = A_1^{-T}\left(b - A_2^T x_2\right)$$

Equality-Constrained QPs

$$\underset{x}{\text{minimize}} \;\; \frac{1}{2}x^T G x + g^T x \quad \text{subject to } A^T x = b$$

Partition: $x = (x_1, x_2)$, similarly for A etc: A_1^{-1} exists

- In practice, factorize A_1 ... check rank!
- Check whether $Ax = b$ inconsistent $\Rightarrow$ QP no solution

Partitioning Hessian, G and gradient g

$$g = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \quad G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix},$$

$\Rightarrow$ eliminate $x_1 = A_1^{-T}\left(b - A_2^T x_2\right)$, get reduced unconstrained QP:

$$\underset{x_2}{\text{minimize}} \;\; \frac{1}{2}x_2^T \tilde{G} x_2 + \tilde{g}^T x_2,$$

For expressions for $\tilde{G}$ and $\tilde{g}$, see Exercises!

Equality-Constrained QPs

$$\text{Reduced QP} \qquad \underset{x_2}{\text{minimize}} \ \tfrac{1}{2} x_2^T \tilde{G} x_2 + \tilde{g}^T x_2,$$

has unique solution, if reduced Hessian, $\tilde{G} \succ 0$, is positive definite

Solve reduced QP by solving the linear system $\tilde{G} x_2 = -\tilde{g}$

- Apply Cholesky factors, or LDL^T factors
- Reduced Hessian factors can be updated in active-set scheme
- Factorization reveals whether problem unbounded:
  If $\tilde{G}$ has negative eigenvalues, then reduced QP unbounded.

Get x_1 and multipliers by substituting/solving

$$x_1 = A_1^{-T} \left(b - A_2^T x_2 \right) \quad \text{and} \quad A_1 y = g_1$$

Generalize elimination technique later!

General Quadratic Programs

General Quadratic Program (QP)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \tfrac{1}{2} x^T G x + g^T x \\ \text{subject to} \quad & a_i^T x = b_i \quad i \in \mathcal{E} \\ & a_i^T x \geq b_i \quad i \in \mathcal{I}, \end{aligned}$$

Active-Set Method for QPs

- Builds on solving equality-constrained QPs (EQPs)
- Start from initial feasible, $x^{(k)}$, with working set, $\mathcal{W}^{(k)}$
- Regard inequality constraints $\mathcal{W}^{(k)}$ temporarily as equations
- Solve corresponding EQP, one of two outcomes:
  1. Prove $x^{(k)}$ is optimal
  2. Find descend direction, and change active set

General Quadratic Programs

General Quadratic Program (QP)

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{2}x^T G x + g^T x \\ \text{subject to} & a_i^T x = b_i \quad i \in \mathcal{E} \\ & a_i^T x \geq b_i \quad i \in \mathcal{I}, \end{array}$$

Can have 0 to n active constraints in, $\mathcal{W}^{(k)}$: EQP($\mathcal{W}^{(k)}$):

$$\begin{array}{l} \underset{x}{\text{minimize}} \;\; \frac{1}{2}x^T G x + g^T x \\ \text{subject to } a_i^T x = b_i \quad i \in \mathcal{W}^{(k)}, \end{array}$$

... solve with any method available for EQPs

Two Key Questions

1. When is solution of EQP($\mathcal{W}^{(k)}$) optimal for general QP?
2. If EQP($\mathcal{W}^{(k)}$) not optimal, where's a descend direction?

Active-Set General QPs

Let solution EQP($\mathcal{W}^{(k)}$) by $\hat{x}^{(k)}$

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad \tfrac{1}{2}x^T G x + g^T x \\ &\text{subject to } a_i^T x = b_i \quad i \in \mathcal{W}^{(k)}, \end{aligned}$$

Solution of EQP($\mathcal{W}^{(k)}$) is optimal for general QP, iff

- If $\hat{x}^{(k)}$ satisfies inactive inequality constraints:

$$a_i^T \hat{x}^{(k)} \geq b_i \quad i \in \mathcal{I} \quad \text{feasibility test}$$

- Multipliers have “right” sign:

$$y_i^{(k)} \geq 0, \ \forall i \in \mathcal{I} \cap \mathcal{W}^{(k)} \quad \text{optimality test}$$

Active-Set General QPs

Let solution EQP($\mathcal{W}^{(k)}$) by $\hat{x}^{(k)}$

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad \tfrac{1}{2}x^T G x + g^T x \\ &\text{subject to } a_i^T x = b_i \quad i \in \mathcal{W}^{(k)}, \end{aligned}$$

If solution of EQP($\mathcal{W}^{(k)}$) is not optimal for general QP, then either

- $\exists q : \ y_q < 0$, e.g. $y_q := \min\{y_i \ : \ i \in \mathcal{I} \cap \mathcal{W}^{(k)}\}$
    - Can move away from constraint q, reducing objective
    - Get search direction, s, by solving new EQP for $\mathcal{W}^{(k+1)} := \mathcal{W}^{(k)} - \{q\}$.

  or ...
- Inactive constraint becomes feasible ... ratio test

Active-Set Method for Quadratic Programming

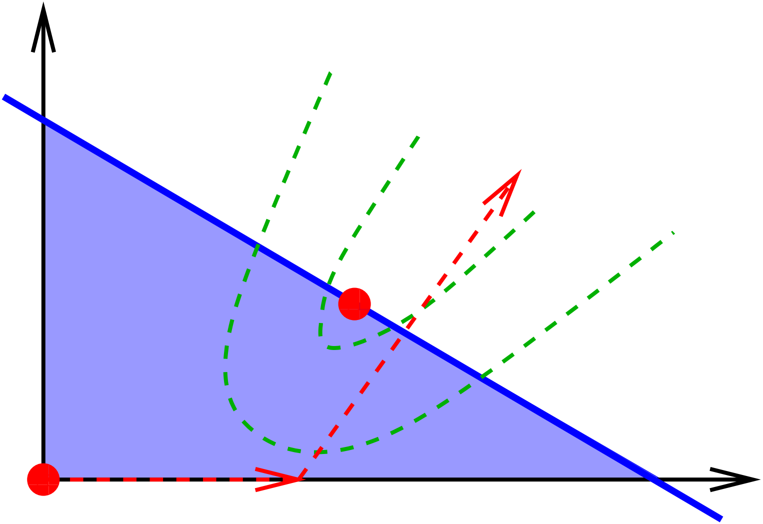
Given initial feasible, $x^{(0)}$, and working set, $\mathcal{W}^{(0)}$, set $k = 0$.

repeat

- **if** $x^{(k)}$ *does not solve the EQP for* $\mathcal{W}^{(k)}$ **then**
  - Solve the EQP($\mathcal{W}^{(k)}$), get $\hat{x}$ and set $s^{(k)} := \hat{x} - x^{(k)}$
  - Ratio Test: $$\alpha = \min_{i \in \mathcal{I} : i \notin \mathcal{W}^{(k)}, a_i^T s_q < 0} \left\{ 1, b_i - a_i^T x^{(k)} / (-a_i^T s_q) \right\}$$
  - **if** $\alpha < 1$ **then**
    - Update $\mathcal{W}$: Add p (min above) to $\mathcal{W}^{(k+1)} = \mathcal{W}^{(k)} \cup \{p\}$
  - Set $x^{(k+1)} = x^{(k)} + \alpha s^{(k)}$ and $k = k + 1$
- **else**
  - Optimality Test: Find $y_q := \min \left\{ y_i : i \in \mathcal{W}^{(k)} \cap \mathcal{I} \right\}$
  - **if** $y_q \geq 0$ **then** $x^{(k)}$ optimal solution ;
  - **else**
    - Update $\mathcal{W}$: Remove q from $\mathcal{W}^{(k+1)} = \mathcal{W}^{(k)} - \{q\}$
  - **end**
- **end**

until $x^{(k)}$ *is optimal or QP unbounded*;

Active-Set Method for Quadratic Programming



Iterates are solutions to EQPs, or ratio test.

Active-Set Method for Quadratic Programming

Can implement algorithm in stable/efficient way

- Update LU factors of A_1
- Update LDL^T factors of reduced Hessian
  ... can include term for one negative eigenvalue

Get initial feasible point using LP phase I approach

Algorithm is primal active-set method (iterates remain feasible)

Dual active-set method can be derived

- Maintains dual feasibility, i.e. multipliers satisfy $y_i^{(k)} \geq 0$
- Move toward primal feasibility
- Equivalent to applying primal active-set method to dual QP
  $\Rightarrow$ requires G^{-1} to exist!
- Fast re-optimization ... good for MIQP solvers

Outline

Generalized Elimination for EQPs

Consider general EQP

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \tfrac{1}{2}x^T G x + g^T x \\ \text{subject to } & A^T x = b \end{aligned}$$

Assumption

Assume $A \in \mathbb{R}^{n \times m}$, with $m \leq n$ has full rank

- *If $n = m$, then solution of EQP is $x = A^{-1}b$*
- *Interested in case $m < n$*

A full rank implies that

$$\exists \; [Y : Z] \quad \text{nonsingular} \quad Y^T A = I_m, \; Z^T A = 0$$

... Y^T is left generalized inverse of A, Z is basis of null-space

Generalized Elimination for EQPs

Consider general EQP

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad \tfrac{1}{2}x^T G x + g^T x \\ &\text{subject to } A^T x = b \end{aligned}$$

A full rank implies that

$$\exists \; [Y : Z] \quad \text{nonsingular} \quad Y^T A = I_m, \; Z^T A = 0$$

... Y^T is left generalized inverse of A, Z is basis of null-space

$\Rightarrow$ all solution of $A^T x = b$ are

$$x = Yb + Z\delta$$

... any point in feasible set can be expressed in this way.

Generalized Elimination for EQPs

Consider general EQP

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \tfrac{1}{2} x^T G x + g^T x \\ \text{subject to } & A^T x = b \end{aligned}$$

$\Rightarrow$ all solution of $A^T x = b$ are

$$x = Yb + Z\delta$$

Use equation to "eliminate" x, get reduced QP:

$$\underset{\delta}{\text{minimize}} \ \tfrac{1}{2} \delta^T \left(Z^T G Z \right) \delta + (g + GYb)^T Z\delta$$

If reduced Hessian $Z^T G Z \succ 0$ pos. def., then unique solution:

$$\nabla_\delta = 0 \ \Leftrightarrow \ \left(Z^T G Z \right) \delta = -Z^T (g + GYb)$$

Generalized Elimination for EQPs

Consider general EQP

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \tfrac{1}{2} x^T G x + g^T x \\ \text{subject to } & A^T x = b \end{aligned}$$

Once we have δ^*, get

$$x^* = Yb + Z\delta^*$$

Find multipliers from

$$Gx^* + g = Ay^* \quad \Leftrightarrow \quad y^* = Y^T (Gx^* + g)$$

because $Y^T A = I_m$, left generalized inverse

Generalized Elimination for EQPs

Consider general EQP

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{2}x^T G x + g^T x \\ \text{subject to } A^T x = b \end{array}$$

A full rank implies that

$$\exists \; [Y : Z] \quad \text{nonsingular} \quad Y^T A = I_m, \; Z^T A = 0$$

That's all very cute ...

... but how on Earth am I supposed to find , Z???

1. Orthonormal QR factors of A
2. General elimination: border $[A : V]$ invertible

Orthogonal Elimination for EQPs

$$(\text{EQP}) \quad \begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{2}x^T G x + g^T x \\ \text{subject to } A^T x = b \end{array}$$

Define QR factors of A (exist, because A has full rank)

$$A = Q \begin{bmatrix} R \\ 0 \end{bmatrix}, \quad Q = [Q_1 : Q_2]$$

where $Q_1 \in \mathbb{R}^{m \times m}$ and R upper triangular
Setting $Z = Q_2$, and $Y = Q_1 R^{-T}$, we observe

$$Y^T A = R^{-1} Q_1^T [Q_1 : Q_2] \begin{bmatrix} R \\ 0 \end{bmatrix} = R^{-1} I_m R = I_m$$

because Q_1 orthonomal, and

$$Z^T A = Q_2^T [Q_1 : Q_2] \begin{bmatrix} R \\ 0 \end{bmatrix} = [0 : I] \begin{bmatrix} R \\ 0 \end{bmatrix} = 0$$

... so factors have desired format, and are numerically stable!

General Elimination for EQPs

$$\text{(EQP)} \quad \begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{2}x^T G x + g^T x \\ \text{subject to } A^T x = b \end{array}$$

Border A by matrix V such that $[A : V]$ nonsingular (exists!)

Define Y, Z as

$$\left[A \; : \; V\right]^{-1} = \begin{bmatrix} Y^T \\ Z^T \end{bmatrix}$$

Then, it follows that

$$I_n = \begin{bmatrix} Y^T \\ Z^T \end{bmatrix} \left[A \; : \; V\right] = \left[\begin{array}{c|c} Y^T A & Y^T V \\ \hline Z^T A & Z^T V \end{array}\right]$$

$\Rightarrow$ $Y^T A = I$ and $Z^T A = 0$ as desired.

In practice, use "previously" active columns to form V
$\Rightarrow$ using LU factors, sparse updates, efficient

General Elimination for EQPs

$$\text{(EQP)} \quad \begin{array}{l} \underset{x}{\text{minimize}} \quad \frac{1}{2}x^T G x + g^T x \\ \text{subject to } A^T x = b \end{array}$$

Define Y, Z as

$$\left[A \; : \; V\right]^{-1} = \begin{bmatrix} Y^T \\ Z^T \end{bmatrix}$$

Border A with special matrix V ... to get first approach

$$\left[\begin{array}{c|c} A_1 & 0 \\ \hline A_2 & I \end{array}\right]^{-1} = \left[\begin{array}{c|c} A_1^{-1} & 0 \\ \hline -A_2 A_1^{-1} & I \end{array}\right] = \begin{bmatrix} Y^T \\ Z^T \end{bmatrix}$$

Then $x = Yb + Z\delta$ becomes

$$x = \begin{bmatrix} A_1^{-T} \\ 0 \end{bmatrix} b + \begin{bmatrix} -A_1^{-T} A_2^{-T} \\ I \end{bmatrix} \delta$$

... and $\delta = x_2$ ... from our original partition method!

Lagrangian Method for EQPs

$$\text{(EQP)} \quad \begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{2}x^T Gx + g^T x \\ \text{subject to} & A^T x = b \end{array}$$

Lagrangian: $\mathcal{L}(x,y) = \frac{1}{2}x^T Gx + g^T x - y^T \left(A^T x - b\right)$

First-order optimality gives: $\nabla_x \mathcal{L} = 0$ and $\nabla_y \mathcal{L} = 0$:

$$\begin{bmatrix} G & -A \\ -A^T & 0 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = - \begin{pmatrix} g \\ b \end{pmatrix}$$

... symmetric system, use factorization that reveals inertia

Summary and Teaching Points

Quadratic Programs

- Many applications in finance, data analysis
- Building block for algorithms for nonlinear optimization

Active-Set Method for QPs

- Generalizes active-set methods for LPs
- Moves from EQP to another ... exploring active sets
- Method of choice for MIQPs (next week)