ISE 789 / OR 791: LECTURE 4 - QUADRATIC OPTIMIZATION MODEL

1. Functions – convex and nonconvex quadratic functions
2. Graphs, contours, governing matrix
3. Systems – Method of least squares, k-means clustering,

   Markowitz mean-variance model
4. Programs - linearly constrained QP (QP) and

   quadratically constrained QP (QCQP)

Optimization models to be covered

1. Linear Optimization
2. Integer Linear Optimization
3. Quadratic Optimization
4. Linear Conic Optimization
5. Nonlinear Optimization

Structure of optimization models

1. Functions
   - form, graph, contour, first/second order information

2. Systems
   - system of equations
   - system of inequalities

3. Programs - Duality and optimality
   - primal problem
   - dual problem
   - complementarity

Quadratic functions

- Quadratic function: a function $f: \mathbb{R}^n \to \mathbb{R}$ is quadratic if

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T M \mathbf{x} + \boldsymbol{b}^T \mathbf{x} + \mathrm{c}$$

where $M \in \mathbb{R}^{n \times n}, \boldsymbol{b} \in \mathbb{R}^n, \mathbf{x} \in \mathbb{R}^n, c \in \mathbb{R}$.

General form:

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q \mathbf{x} + \boldsymbol{b}^T \mathbf{x} + \mathrm{c}$$

where $Q \in \mathbb{S}^n$ (a symmetric matrix).

- Why? $Q = \frac{M+M^T}{2}$

Quadratic functions - I

- Properties:

(i) If f and g are quadratic and $\alpha \in \mathbb{R}$, then $f+g$, $f-g$, αf are quadratic.

- why?

(ii) First order information

- $$\nabla f(\mathbf{x}) = Q\mathbf{x} + \boldsymbol{b}$$

- implications?

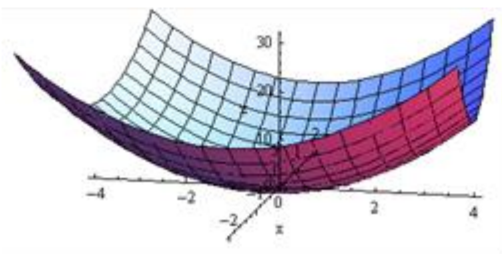
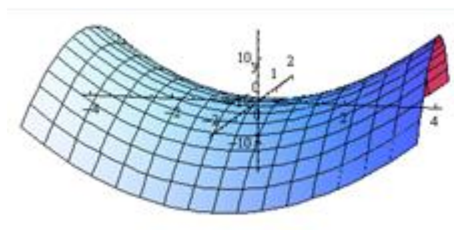
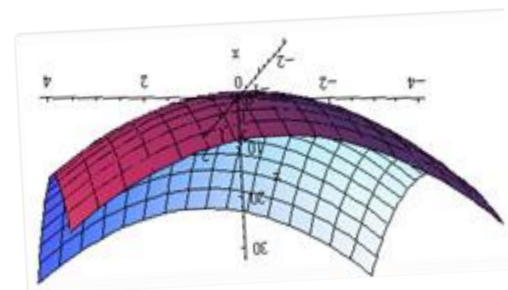
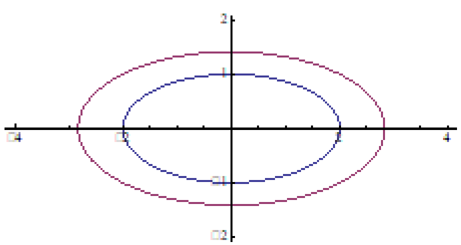
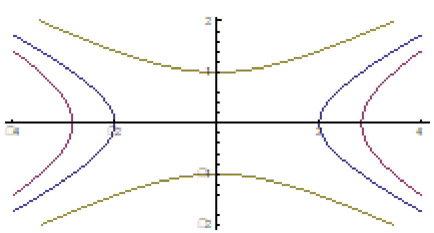
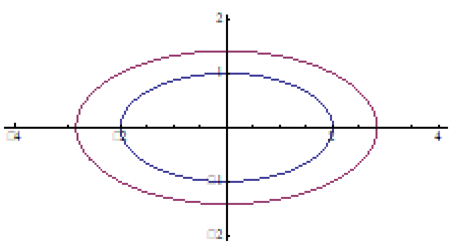
(iii) Second order information

- $$F(\mathbf{x}) = Q$$

- implications?

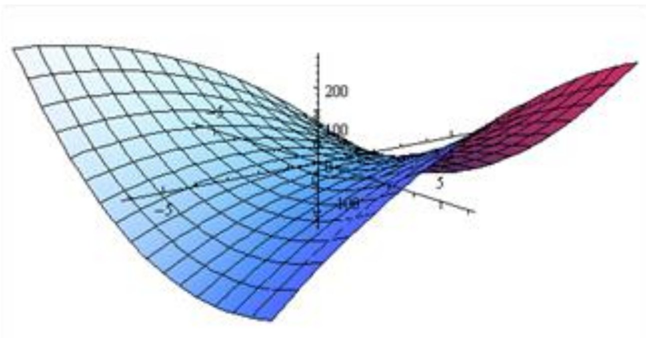
Graphs & contours of quadratic functions

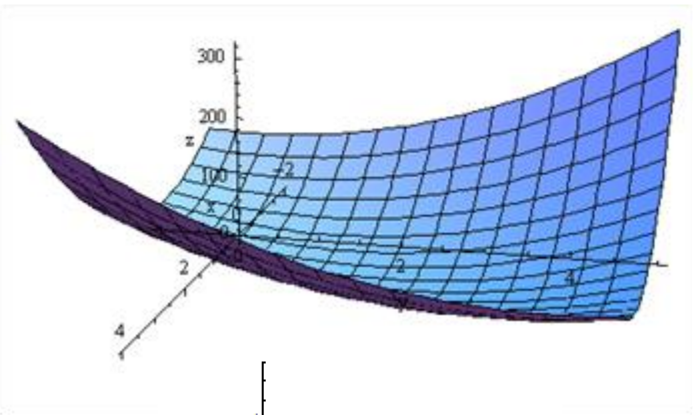
- Quadratic functions of two variables (from umich.edu Math217 Notes 8.3)

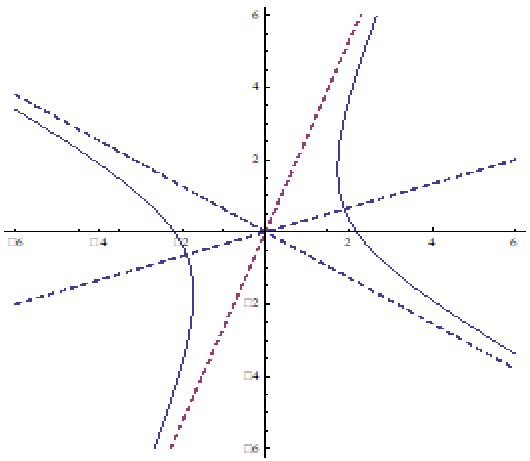
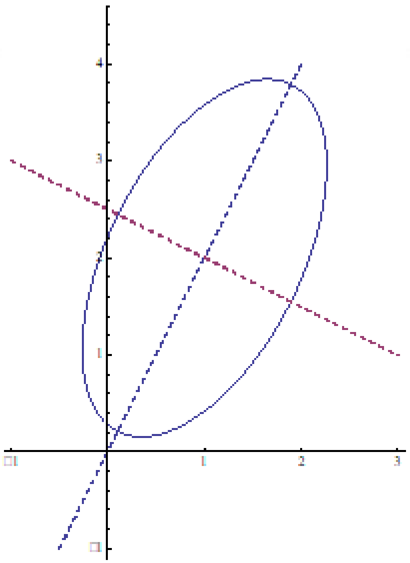







Graphs and contours – cont'

- Quadratic functions with two variables





Special symmetric matrices

- Definition: A real symmetric matrix M of degree n ($M \in \mathbb{S}^n$) is

  (i) positive semidefinite (psd), i.e., $M \in \mathbb{S}^n_+$, if

  $$\mathbf{x}^{\mathrm{T}} M \mathbf{x} \geq 0, \forall \mathbf{x} \in \mathbb{R}^n;$$

  (ii) negative semidefinite (nsd), if

  $$\mathbf{x}^{\mathrm{T}} M \mathbf{x} \leq 0, \forall \mathbf{x} \in \mathbb{R}^n;$$

  (iii) indefinite, if M is neither psd nor nsd.

  Properties: M is psd $\Leftrightarrow -M$ is nsd;

  αM is psd, for $\alpha \geq 0$; nsd for $\alpha \leq 0$;

  M, N are psd (nsd) $\Rightarrow$ $M + N$ is psd (nsd).

Special symmetric matrices

- Definition: A real symmetric matrix M of degree n ($M \in \mathbb{S}^n$) is

(i) positive definite (pd), i.e., i.e., $M \in \mathbb{S}^n_{++}$, if

$$\mathbf{x}^{\mathrm{T}} M \mathbf{x} > 0, \forall \mathbf{x} \in \mathbb{R}^n \backslash \{0\};$$

(ii) negative definite (nd), if

$$\mathbf{x}^{\mathrm{T}} M \mathbf{x} < 0, \forall \mathbf{x} \in \mathbb{R}^n \backslash \{0\};$$

Properties: M is pd $\Leftrightarrow -M$ is nd;

αM is pd, for $\alpha > 0$; nd for $\alpha < 0$;

M, N are pd (nd) $\Rightarrow$ $M + N$ is pd (nd).

Special symmetric matrices

- Theorem: A symmetric matrix M of degree n is

  (i) psd (pd) iff all eigenvalues are non-negative (positive);

  (ii) nsd (nd) iff all eigenvalues are non-positive (negative).

Why? Spectral decomposition

For a non-defective (diagonalizable) matrix M and $\mathbf{x} \in \mathbb{R}^n$, we have

$$\mathbf{x}^{\mathrm{T}} M \mathbf{x} = \mathbf{x}^{\mathrm{T}} U \Lambda U^{T} \mathbf{x} = \sum_{i=1}^{n} \lambda_i \left(\left(\mathbf{x}^{\mathrm{T}} U \right)_i \right)^2$$

where $\Lambda = diag(\lambda_1, \dots, \lambda_n)$ and matrix U contains the n orthogonal eigenvectors of M.

Special symmetric matrices

Properties:

- M is psd (pd) ⇒ all diagonal elements are nonnegative (positive);
- M is psd ⇒ $M = M^{\frac{1}{2}} M^{\frac{1}{2}}$ ($M^{\frac{1}{2}}$ is a psd square root matrix);
- M is pd ⇒ M is nonsingular and M^{-1} is pd.

Special insight: a quadratic function defined by a non-defective matrix can be transformed to an angle (in view of the coordinates referring to the n orthogonal eigenvectors) such that either you see a convex quadratic curve, a flat line, or a concave quadratic curve along each coordinate.

Quadratic function - II

- Properties:
- (iv) A quadratic function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \boldsymbol{b}^T\mathbf{x} + \mathrm{c}$
  with $Q \in \mathbb{S}^n$ is a convex function
  iff Q is positive semidefinite (i.e., $Q \in \mathbb{S}_+^n$).

- (v) $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \boldsymbol{b}^T\mathbf{x} + \mathrm{c}$ is strictly convex
  iff Q is positive definite (i.e., $Q \in \mathbb{S}_{++}^n$).

- (vi) $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \boldsymbol{b}^T\mathbf{x} + \mathrm{c}$ with $Q \in \mathbb{S}^n$ is concave
  iff Q is negative semidefinite (i.e., $-Q \in \mathbb{S}_+^n$).

- (vii) $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \boldsymbol{b}^T\mathbf{x} + \mathrm{c}$ is strictly concave
  iff Q is negative definite (i.e., $-Q \in \mathbb{S}_{++}^n$).

How quadratic functions come into picture?

- Approximation – Taylor's approximation

  - For a function $f \in C^2(\mathbb{R}^n)$ around a point $\mathbf{x}'$, we have

$$f(\mathbf{x}) \approx f(\mathbf{x}') + \nabla f(\mathbf{x}')^T(\mathbf{x} - \mathbf{x}') + \frac{1}{2}(\mathbf{x} - \mathbf{x}')^T F(\mathbf{x}')\,(\mathbf{x} - \mathbf{x}')$$

How quadratic functions come into picture?

- Distance – measurement of error or discrepancy

  - For a point $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ and point $\mathbf{y} = (y_1, y_2, \dots, y_n)^{\mathrm{T}}$, the Minkowski distance of order $p \geq 1$ (***p*-norm distance**) is defined as:
    - (1-norm distance) $= \sum_{i=1}^{n} |x_i - y_i|$
    - (2-norm distance) $= (\sum_{i=1}^{n} |x_i - y_i|^2)^{\frac{1}{2}}$
    - (p-norm distance) $= (\sum_{i=1}^{n} |x_i - y_i|^p)^{\frac{1}{p}}$
    - (infinity norm distance) $= \lim_{p \to \infty} (\sum_{i=1}^{n} |x_i - y_i|^p)^{\frac{1}{p}}$

      $= \max\{|x_1 - y_1|, |x_2 - y_2|, \dots, |x_n - y_n|\}$

- $||\mathbf{x} - \mathbf{y}||_2^2$ is a convex quadratic function.
- $||\mathbf{x}||_2^2 = \mathbf{x}^T\mathbf{x}$ is a strictly convex quadratic function whose

  gradient $= 2\mathbf{x}$ and Hessian $= 2\,I$.

Logistics system – location problem

- Given the locations $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$ of n stores, find a location $\mathbf{x}$ in a feasible area C for hosting a warehouse that serves the stores in a most economic manner.

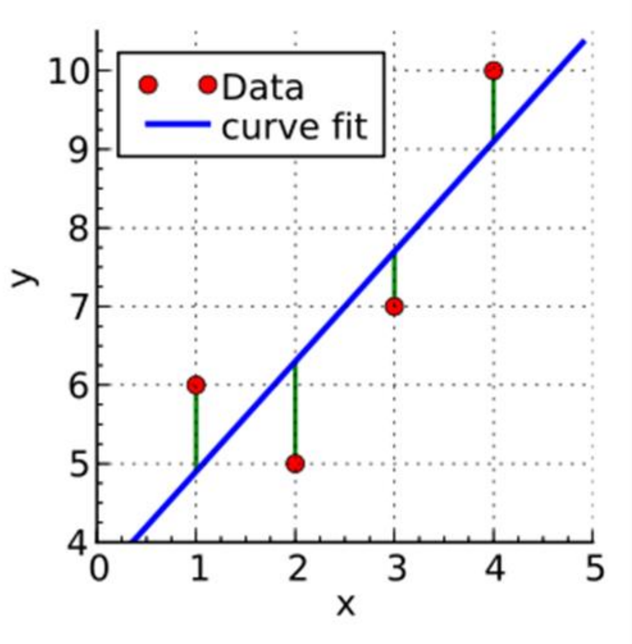
$$\begin{aligned} \min \quad & \sum_{i=1}^{n} \|\mathbf{x}^i - \mathbf{x}\|_2^2 \\ s.t. \quad & \mathbf{x} \in C. \end{aligned}$$

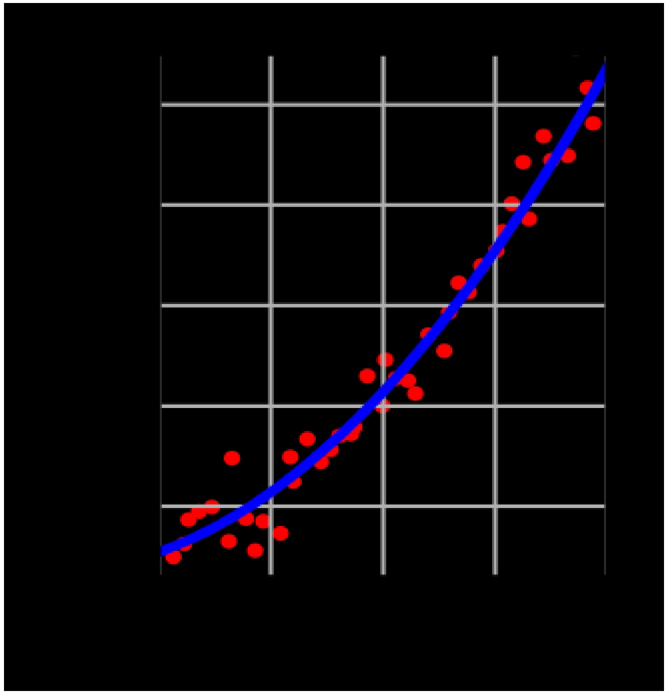
or,

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} t_i \\ s.t. \quad & \|\mathbf{x}^i - \mathbf{x}\|_2^2 \leqslant t_i, \quad i = 1, \ldots, n \\ & \mathbf{x} \in C. \end{aligned}$$

Method of least squares for data fitting

- Fit the data to unveil hidden relations





Method of least squares for data fitting

- Legendre's method (1805)

- Given n data points $\{\, (\mathbf{x}^i, y^i) \mid i = 1, .., n \,\}$ with y depending on $\mathbf{x}$ in the form of $f\,(\mathbf{x}, \boldsymbol{\beta})$ (called a model) where $\boldsymbol{\beta}$ is a vector of m parameters, our goal is to find the parameter values for the model that "best fit" the data.

- The method of least squares is often used in regression analysis to generate estimators and other statistics.

Methods of least squares for data fitting

- Legendre's method (1805)

  - residual

$$r_i = y^i - f(\mathbf{x}^i, \boldsymbol{\beta}).$$

  - sum of squared residuals

$$S = \sum_{i=1}^{n} r_i^2.$$

  - minimize S by setting the gradient to zero

$$\frac{\partial S}{\partial \beta_j} = 2\sum_{i=1}^{n} r_i \frac{\partial r_i}{\partial \beta_j} = 0, \quad j = 1, \ldots, m$$

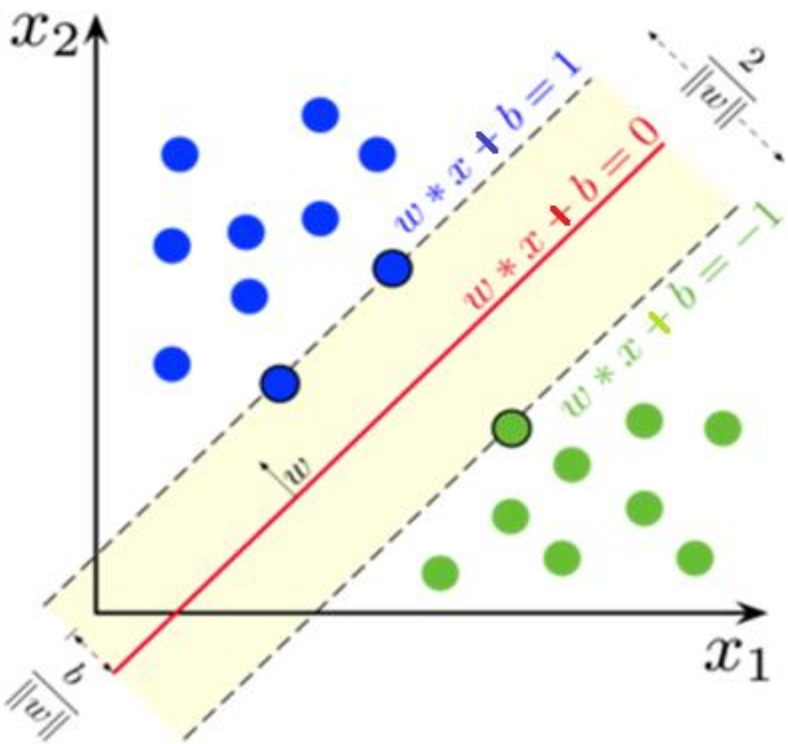
and since $r_i = y^i - f(\mathbf{x}^i, \boldsymbol{\beta})$, the gradient equations become

$$-2\sum_{i=1}^{n} r_i \frac{\partial f(\mathbf{x}^i, \boldsymbol{\beta})}{\partial \beta_j} = 0, \quad j = 1, \ldots, m$$

  - solve a system of m equations in m variables for $\boldsymbol{\beta}$.

Supervised learning – basic SVM model

- Linear separation with maximum margin (distance)



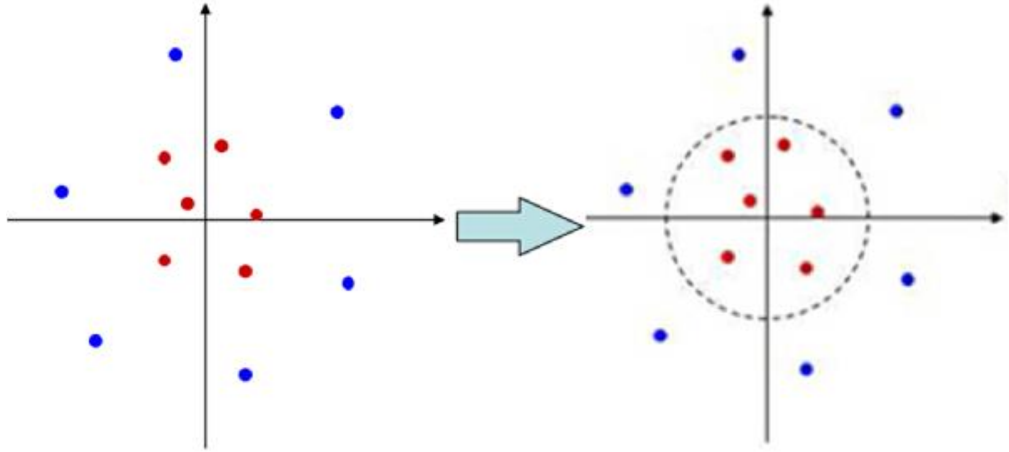
$$\max \quad \frac{2}{\|\boldsymbol{w}\|_2}$$
$$s.t. \quad y_i(\boldsymbol{w}^T\mathbf{x}^i + b) \geqslant 1$$
$$\forall i = 1, \ldots, N.$$
$$\boldsymbol{w} \in \mathbb{R}^n, b \in \mathbb{R}.$$

equivalently,

$$\min \quad \frac{\|\boldsymbol{w}\|_2^2}{2}$$
$$s.t. \quad y_i\left(\boldsymbol{w}^T\mathbf{x}^i + b\right) \geqslant 1,$$
$$\boldsymbol{w} \in \mathbb{R}^n, b \in \mathbb{R}.$$

Quadratic support vector machine (QSVM)

- Motivation:
  - Not linearly separable datasets may be nonlinearly separable



Quadratic support vector machine (QSVM)

- Basic idea: (Jian Luo 2016)
  - supervised learning using a quadratic surface to separate two classes of data points

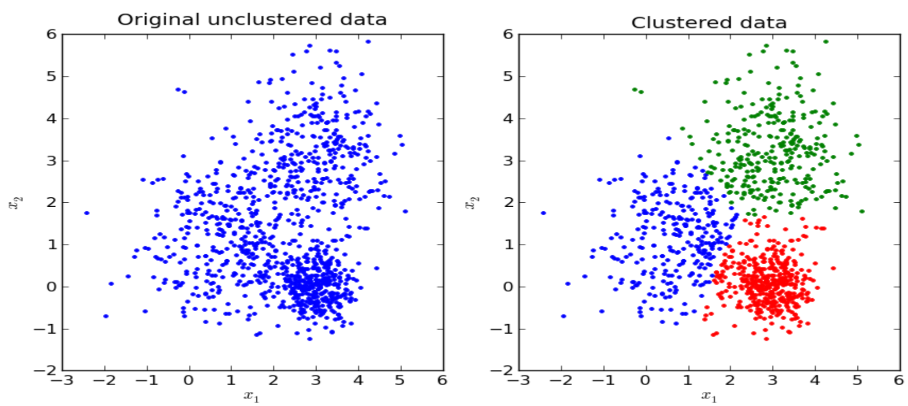
$$
\begin{aligned}
\max \quad & \sum_{i=1}^{n} \frac{1}{\|W\mathbf{x}^i + \boldsymbol{b}\|_2} \\
s.t. \quad & y_i \left(\frac{1}{2} \left(\mathbf{x}^i\right)^T W \mathbf{x}^i + \boldsymbol{b}^T \mathbf{x}^i + c \right) \geqslant 1, \quad \forall i = 1, \ldots, n. \\
& W \in \mathbb{S}^m, b \in \mathbb{R}^m, c \in \mathbb{R}.
\end{aligned}
$$

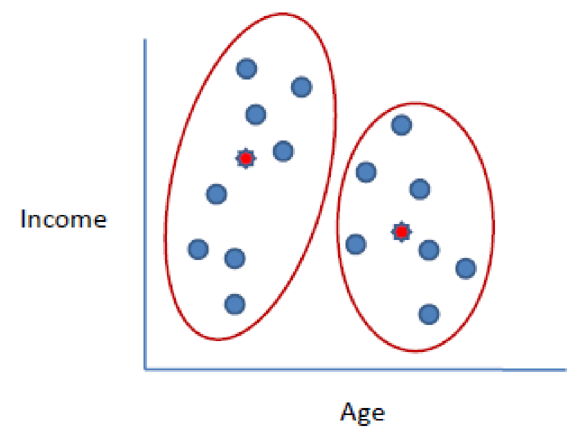
Equivalently,

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n} \|W\mathbf{x}^i + \boldsymbol{b}\|_2^2 \\
s.t. \quad & y_i \left(\frac{1}{2} \left(\mathbf{x}^i\right)^T W \mathbf{x}^i + \boldsymbol{b}^T \mathbf{x}^i + c \right) \geqslant 1, \quad \forall i = 1, \ldots, n. \\
& W \in \mathbb{S}^m, b \in \mathbb{R}^m, c \in \mathbb{R}.
\end{aligned}
$$

Unsupervised learning - clustering

- Uncover some hidden structure of datasets by clusters
- Clustering based on similarity
- Similarity based on distance





- <From https://mubaris.com/posts/kmeans-clustering/>

K-means clustering

- Centroid-based k-means clustering

  (Stuart Lloyd of Bell Labs, 1957)

Given a set of observations $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$, where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k $(\leq n)$ sets $\boldsymbol{S} = \{S_1, S_2, \ldots, S_k\}$ so as to minimize the within-cluster sum of squares (WCSS) (i.e. variance). Formally, the objective is to find:

$$\underset{\mathbf{S}}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}^i\|_2^2 = \underset{\mathbf{S}}{\operatorname{argmin}} \sum_{i=1}^{k} |S_i| \operatorname{Var} S_i.$$

where $\boldsymbol{\mu}_i$ is the mean of data points in S_i.

K-means clustering

- The problem is proven to be NP-hard.
  - *Why ?*

How about an IP formulation?

- *what do we have in hand?*
- *what do we want to achieve?*
- *what are variables? constraints? objective?*

IP model for k-means clustering

Define: $$\rho_{j,i} = \begin{cases} 1, & \text{if } \mathbf{x}^j \in S_i \\ 0, & \text{if } \mathbf{x}^j \notin S_i \end{cases}$$

1. $$\sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}^i\|_2^2 \quad \Rightarrow \quad \sum_{i=1}^{k} \sum_{j=1}^{n} \rho_{j,i} \|\mathbf{x}^j - \boldsymbol{\mu}^i\|_2^2$$

2. $$\boldsymbol{\mu}^i = \frac{1}{|S_i|} \sum_{\mathbf{x} \in S_i} \mathbf{x} \quad = \quad \frac{1}{\sum_{j=1}^{n} \rho_{j,i}} \sum_{j=1}^{n} \rho_{j,i} \mathbf{x}^j .$$

K-means (IP model)

$$\begin{aligned} \min \quad & \sum_{i=1}^{k} \sum_{j=1}^{n} \rho_{j,i} \|\mathbf{x}^j - \boldsymbol{\mu}^i\|_2^2 \\ s.t. \quad & \sum_{j=1}^{n} \rho_{j,i} \mathbf{x}^j = \boldsymbol{\mu}^i \sum_{j=1}^{n} \rho_{j,i}, \quad i = 1, \ldots, k \\ & \rho_{j,i} \in \{0, 1\}, \quad j = 1, \ldots, n, \ i = 1, \ldots, k \\ & \boldsymbol{\mu}^i \in \mathbb{R}^n, \quad i = 1, \ldots, k. \end{aligned}$$

Method for k-means clustering

- B&B? B&C? Heuristics?

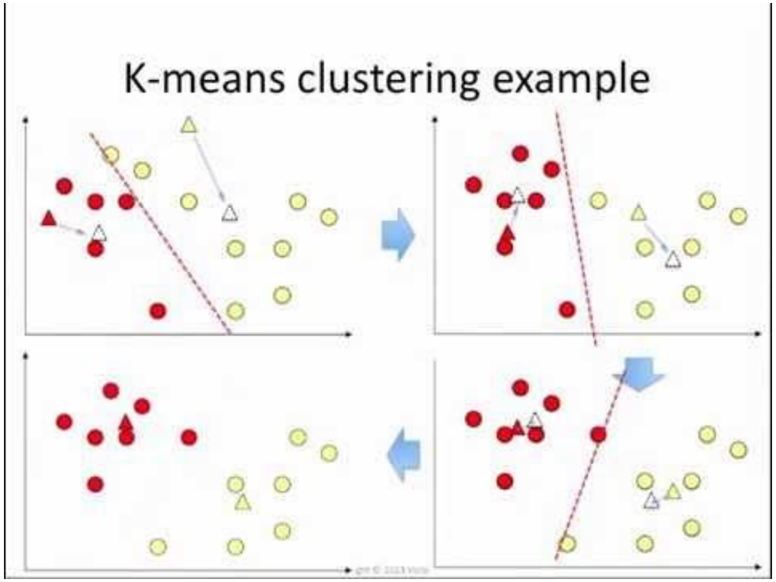
Assume K is given and $\mathbf{x}^i \in \mathbb{R}^d$

Step 1: assign any K points $\{\boldsymbol{\mu}_k \in \mathbb{R}^d, k = 1, \dots, K\}$

Step 2: Find $k^* = \underset{k}{\operatorname{argmin}} ||\mathbf{x}^i - \boldsymbol{\mu}_k||, \ i = 1, \dots, n.$ Assign point $\mathbf{x}^i$ to class S_{k^*} .

Step 3: update $\boldsymbol{\mu}_k = \sum_{i \in S_k} \frac{\mathbf{x}^i}{\mathrm{card}(S_k)}, k = 1, \dots, K.$

Step 4: Go to Step 2 until all $\boldsymbol{\mu}_k$ remain unchanged.



<https://www.youtube.com/watch?v=_aWzGGNrcic>

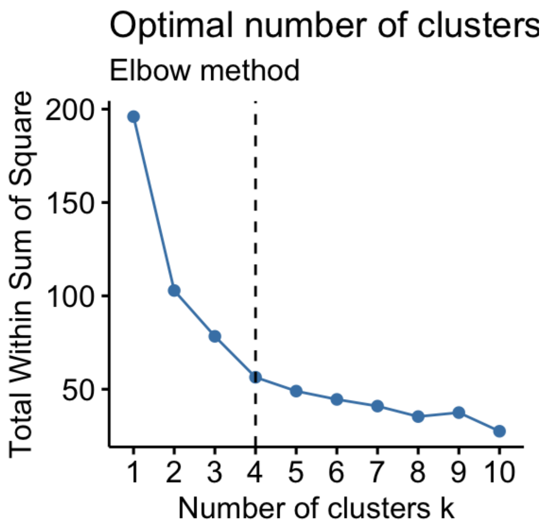
Questions

- How good is the heuristic algorithm?

  - will it converge?
  - converge to an optimal solution?
  - how fast?
  - can you develop a better algorithm?

How many clusters should there be?

- Can we learn form the dataset to decide an optimal K?
- On which basis?
  - minimize the “intra-cluster distance”?
  - maximize the “inter-cluster distance”?
- What’s your ideas?



How quadratic functions come into picture

- Covariance – measurement of risk

  - A random variable has its mean (average) and variance (spread).

  - Intuitively, the covariance matrix of a random vector generalizes the notion of variance to multiple dimensions.

  - Covariance matrix (also called dispersion matrix or variance–covariance matrix) is a square matrix giving the covariance between each pair of elements of a given random vector.

  - Each diagonal element of a covariance matrix reports the variance of a corresponding variable (covariance of a random variable with itself).

  - In statistics, the covariance matrix of a multivariate probability distribution is always positive semi-definite; and it is positive definite unless one variable is an exact linear function of the others.

How quadratic functions come into picture

- Covariance – measurement of risk
  - If the entries in the column vector $\mathbf{X} = [X_1,\ X_2, \dots, X_n]^T$ are random variables, each with finite variance and expected value, then the covariance matrix $K_{\mathbf{XX}}$ is the matrix whose (i, j) entry is the variance

$$K_{X_iX_j} = \operatorname{cov}[X_i, X_j] = E[(X_i - E[X_i])(X_j - E[X_j])]$$

where the operator E denotes the expected value (mean).

$$K_{\mathbf{XX}} = \begin{bmatrix} E[(X_1 - E[X_1])(X_1 - E[X_1])] & E[(X_1 - E[X_1])(X_2 - E[X_2])] & \cdots & E[(X_1 - E[X_1])(X_n - E[X_n])] \\ E[(X_2 - E[X_2])(X_1 - E[X_1])] & E[(X_2 - E[X_2])(X_2 - E[X_2])] & \cdots & E[(X_2 - E[X_2])(X_n - E[X_n])] \\ \vdots & \vdots & \ddots & \vdots \\ E[(X_n - E[X_n])(X_1 - E[X_1])] & E[(X_n - E[X_n])(X_2 - E[X_2])] & \cdots & E[(X_n - E[X_n])(X_n - E[X_n])] \end{bmatrix}$$

How quadratic functions come into picture

- Correlation matrix
  - Equivalently, the correlation matrix can be seen as the covariance matrix of the standardized random variables $\frac{X_i}{\sigma(X_i)}$ for $i = 1, \ldots, n$.

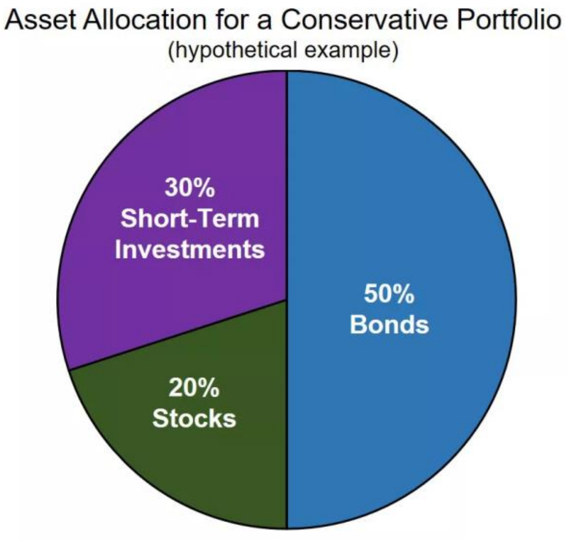
$$\operatorname{corr}(\mathbf{X}) = \begin{bmatrix} 1 & \frac{E[(X_1-\mu_1)(X_2-\mu_2)]}{\sigma(X_1)\sigma(X_2)} & \cdots & \frac{E[(X_1-\mu_1)(X_n-\mu_n)]}{\sigma(X_1)\sigma(X_n)} \\ \frac{E[(X_2-\mu_2)(X_2-\mu_2)]}{\sigma(X_2)\sigma(X_1)} & 1 & \cdots & \frac{E[(X_2-\mu_2)(X_n-\mu_n)]}{\sigma(X_2)\sigma(X_n)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{E[(X_n-\mu_n)(X_1-\mu_1)]}{\sigma(X_n)\sigma(X_1)} & \frac{E[(X_n-\mu_n)(X_2-\mu_2)]}{\sigma(X_n)\sigma(X_2)} & \cdots & 1 \end{bmatrix}$$

  - Each element on the principal diagonal of a correlation matrix is the correlation of a random variable with itself, which always equals 1. Each off-diagonal element is between -1 and +1 inclusively.

How quadratic functions come into picture

- Risk of investment

  - Asset is something valuable;
  - Portfolio is a combination of assets; $\sum_{i=1}^{n} w_i = 1$;
  - Return of each asset is a random variable r_i;
  - Expected return of asset $i = E[r_i] \quad (i.e.\ \bar{r}_i)$;
  - Expected return of investment = $\sum_{i=1}^{n} w_i \bar{r}_i$;
  - Variance of the return vector is the covariance matrix $K_{\boldsymbol{rr}}$;
  - Risk of investment = $\boldsymbol{w}^T K_{\boldsymbol{rr}} \boldsymbol{w}$, a convex quadratic function of $\boldsymbol{w}$.



Financial system – portfolio selection

- Harry Max Markowitz (1927 -)

  (PhD of Economics at U Chicago, 1954,

  Nobel Prize 1990)

- Mean-variance model (1952, Portfolio Selection)
  - Assuming that the returns of assets follow a multivariate normal distribution, the return of a portfolio can be completely described by the expected return and variance of the forming assets.

Mean-variance model

- Markowitz mean-variance model (QP model)
  - For a given set of n assets, let $\boldsymbol{r} = [r_1, \dots, r_n]^T$ and $D = \{D_{ij}\}_{i,j=1}^{n}$ denote the vector of expected returns and matrix of covariance, respectively. To find an investment portfolio with the minimum risk subject to receiving an expected return, we may consider the following Markowitz mean-variance model (Markowitz, 1952):

$$
\begin{aligned}
\min \quad & \frac{1}{2}\mathbf{x}^T D\mathbf{x} \\
s.t. \quad & \boldsymbol{r}^T\mathbf{x} = r_0 \\
& \boldsymbol{e}^T\mathbf{x} = 1 \\
& x_i \geqslant 0, \quad i = 1, \dots, n.
\end{aligned}
\qquad \text{(MV)}
$$

  where D is pd, $\boldsymbol{r} \in \mathbb{R}^n, \boldsymbol{e} = [1, \dots, 1]^T \in \mathbb{R}^n$ and r_0 is given.

- This is a convex quadratic program with linear constraints.

Variations of mean-variance model

- (V1)

$$\begin{aligned} \max \quad & \boldsymbol{r}^T \mathbf{x} \\ s.t. \quad & \frac{1}{2}\mathbf{x}^T D \mathbf{x} \leqslant \delta \\ & \boldsymbol{e}^T \mathbf{x} = 1 \\ & \mathbf{x} \in \mathbb{R}^n_+. \end{aligned}$$

where $\delta > 0$ is a given tolerance.

This is a convex program with quadratic constraints.

- (V2)

$$\begin{aligned} \max \quad & \boldsymbol{r}^T \mathbf{x} \\ \min \quad & \frac{1}{2}\mathbf{x}^T D \mathbf{x} \\ s.t. \quad & \boldsymbol{e}^T \mathbf{x} = 1 \\ & \mathbf{x} \in \mathbb{R}^n_+. \end{aligned}$$

This is a “multi-criteria decision making (MCDM)” problem.

Cardinality constrained mean-variance model

- CCMV model:

  - For a common practice, investors may confine themselves to selecting a small number of s $(s << n)$ assets in forming a portfolio for effective management.

$$
\begin{aligned}
\min \quad & \frac{1}{2}\mathbf{x}^T D\mathbf{x} \\
s.t. \quad & \boldsymbol{r}^T\mathbf{x} = r_0 \\
& \boldsymbol{e}^T\mathbf{x} = 1 \\
& \|\mathbf{x}\|_0 \leqslant s \\
& x_i \geqslant 0, \quad i = 1, \ldots, n.
\end{aligned}
\qquad \text{(CCMV)}
$$

  where $||\mathbf{x}||_0$ denotes the # of nonzero entries in $\mathbf{x}$.

- This is a quadratic program with non-differentiable constraints.

Cardinality constrained mean-variance model

- CCMV - MIQP model:
  - Problem (CCMV) can be reformulated as the following mixed integer quadratic program (MIQP)

$$
\begin{aligned}
\min \quad & \frac{1}{2}\mathbf{x}^T D\mathbf{x} \\
s.t. \quad & \boldsymbol{r}^T\mathbf{x} = r_0 \\
& \boldsymbol{e}^T\mathbf{x} = 1 \\
& \boldsymbol{e}^T\boldsymbol{z} \leqslant s \\
& 0 \leqslant x_i \leqslant z_i, \quad i = 1, \ldots, n \\
& z_i \in \{0, 1\}, \quad i = 1, \ldots, n.
\end{aligned}
\qquad \text{(CCMV\_MIQP)}
$$

where $\boldsymbol{z} \in \{0,1\}^n$ is a binary auxiliary variable.

- This is a mixed-integer quadratic program.

Questions

- Is the (MV) model an easy problem to solve?

    Yes, (MV) is (polynomial-time) solvable!

- Is the (CCMV) model a hard problem to solve?

    Yes, (CCMV) is NP-hard!

    How to find an efficient solution method?

Structure of optimization models

1. Functions
   - form, graph, contour, first/second order information

2. Systems
   - system of equations
   - system of inequalities

3. Programs - duality and optimality
   - primal problem
   - dual problem
   - complementarity

Linearly constrained quadratic program (LCQP)

- Given $Q \in \mathbb{S}^n, A \in \mathbb{R}^{m\times n}, \boldsymbol{b} \in \mathbb{R}^m, \boldsymbol{c} \in \mathbb{R}^n$, a linearly constrained quadratic program is given by
- (LCQP)

$$\begin{aligned} \min \quad & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \boldsymbol{c}^T\mathbf{x} \\ s.t. \quad & A\mathbf{x} = \boldsymbol{b} \\ & \mathbf{x} \in \mathbb{R}^n_+ \end{aligned}$$

- When $Q \in \mathbb{S}^n_+$, (LCQP) is a convex optimization problem, which is commonly called a "quadratic program."

Quadratic program (QP)

- Given $(Q, A, \boldsymbol{b}, \boldsymbol{c})$ in proper dimensions, we have a pair of (linearly constrained) convex quadratic programs:

$$\begin{aligned} \min \quad & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \boldsymbol{c}^T\mathbf{x} \\ s.t. \quad & A\mathbf{x} = \boldsymbol{b} \\ & \mathbf{x} \in \mathbb{R}^n_+ \end{aligned} \qquad \text{(P)}$$

$$\begin{aligned} \max \quad & \boldsymbol{b}^T\boldsymbol{w} - \frac{1}{2}\boldsymbol{z}^T Q\boldsymbol{z} \\ s.t. \quad & A^T\boldsymbol{w} \leqslant Q\boldsymbol{z} + \boldsymbol{c} \\ & \boldsymbol{w} \in R^m, \ \boldsymbol{z} \in \mathbb{R}^n. \end{aligned} \qquad \text{(D)}$$

where $Q \in \mathbb{S}^n_+, \ A \in \mathbb{R}^{m\times n}, \ \boldsymbol{b} \in \mathbb{R}^m, \ \boldsymbol{c} \in \mathbb{R}^n.$

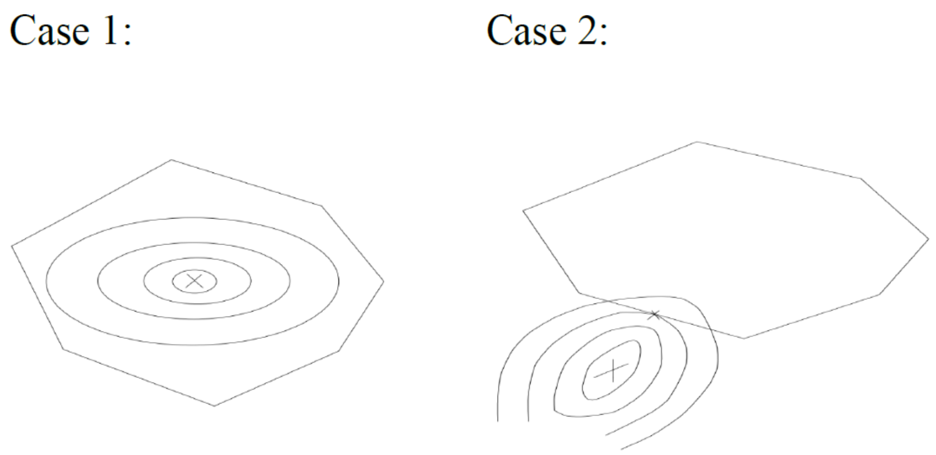
Duality theory

- Strong duality holds for convex (QP).

Implications: - no duality gap

- primal approach, dual approach, primal-dual approach

Solution methods



- Optimality may happen at any feasible point.
- Many solution methods
    - conjugate gradient method
    - preconditioned conjugate gradient method
    - active set algorithm
    - interior-point methods (polynomial-time solvable)

Quadratically constrained quadratic program (QCQP)

- QCQP:

$$
\begin{aligned}
\min \quad & \frac{1}{2}\mathbf{x}^T Q_0 \mathbf{x} + \boldsymbol{b}_0^T \mathbf{x} + c_0 \\
s.t. \quad & \frac{1}{2}\mathbf{x}^T Q_i \mathbf{x} + \boldsymbol{b}_i^T \mathbf{x} + c_i \leqslant 0 \\
& i = 1, \ldots, m. \\
& \mathbf{x} \in \mathbb{R}^n.
\end{aligned}
$$

where $Q_j \in \mathbb{S}^n,\ \boldsymbol{b}_j \in \mathbb{R}^n,\ c_j \in \mathbb{R}$, for $j = 0, \ldots, m.$

Convex QCQP (CQCQP)

- When the governing matrices are positive semidefinite, i.e.,

$$Q_j \in \mathbb{S}^n_+, \text{ for } j = 0,1,\ldots,m,$$

(QCQP) becomes a convex optimization problem.

- (CQCQP) is equivalent to a "second order cone program" (SOCP), which is a special class of the linear conic optimization problems (LCoP).

SOCP

- Second order cone
- Definition:

$$\mathcal{L}^n \triangleq \left\{ \mathbf{x} \in \mathbb{R}^n \,\middle|\, \sqrt{x_1^2 + \cdots + x_{n-1}^2} \leqslant x_n \right\}$$

- Second order cone programming

$$\begin{aligned} \min \quad & \boldsymbol{c}^T \mathbf{x} \\ s.t. \quad & A\mathbf{x} = \boldsymbol{b} \qquad\qquad \text{(SOCP)} \\ & \mathbf{x} \in \mathcal{L}^n. \end{aligned}$$

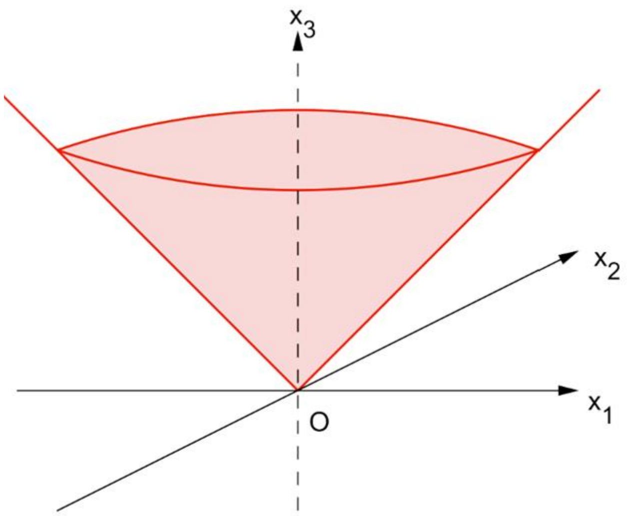


Figure : $\mathcal{L}^3$

Convex QCQP ⇒ SOCP

- CQCQP can be equivalently formulated as an SOCP

$$
\begin{aligned}
\min \quad & u \\
s.t. \quad & \begin{bmatrix} Q_0^{\frac{1}{2}} \mathbf{x} \\ -\boldsymbol{b}_0^T - \frac{c_0}{2} - \frac{1}{2} + \frac{u}{2} \\ -\boldsymbol{b}_0^T - \frac{c_0}{2} + \frac{1}{2} + \frac{u}{2} \end{bmatrix} \in \mathcal{L}^{n+2} \\
& \begin{bmatrix} Q_i^{\frac{1}{2}} \mathbf{x} \\ -\boldsymbol{b}_i^T - \frac{c_i}{2} - \frac{1}{2} \\ -\boldsymbol{b}_i^T - \frac{c_i}{2} + \frac{1}{2} \end{bmatrix} \in \mathcal{L}^{n+2}, \quad i = 1, \ldots, m. \\
& u \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n.
\end{aligned}
$$

Duality theory

- As a second order cone programming problem, CQCQP has a dual problem that follows the strong duality theorem.

- Theorem (SOCP duality theorem)
  i. If either SOCP or SOCD is unbounded, then the other one is infeasible.
  ii. If there exists a feasible solution $\bar{\mathbf{x}}$ such that $\bar{\mathbf{x}} \in \mathrm{int}(K)$, and $v(\mathrm{SOCP})$ is finite, then there exists $(\boldsymbol{y}^*, \boldsymbol{s}^*) \in \mathrm{feas}(\mathrm{SOCD})$ such that $v(\mathrm{SOCP}) = \boldsymbol{b}^T \boldsymbol{y}^* = v(\mathrm{SOCD})$.
  iii. If there exists a feasible solution $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{s}})$ such that $\bar{\boldsymbol{s}} \in \mathrm{int}(K)$, and $v(\mathrm{SOCD})$ is finite, then there exists $\mathbf{x}^* \in \mathrm{feas}(\mathrm{SOCP})$ such that $v(\mathrm{SOCP}) = \boldsymbol{c}^T \mathbf{x}^* = v(\mathrm{SOCD})$.

Solution methods

- As a second order cone program, CQCQP is polynomial-time solvable.

- Interior-point methods are available
-     - primal approach
-     - dual approach
-     - primal-dual approach