CS/ECE/ISYE524: Introduction to Optimization – Convex Optimization Models

Jeff Linderoth

Department of Industrial and Systems Engineering
University of Wisconsin-Madison

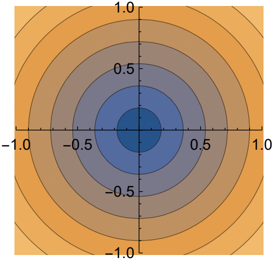
March 13, 2024

Ellipsoids

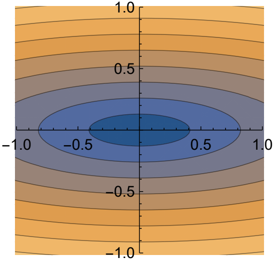
- For linear constraints, the set of x satisfying $c^{\mathsf{T}}x = b$ is a *hyperplane* and the set $c^{\mathsf{T}}x \leq b$ is a *halfspace*.
- For quadratic constraints, the set of x satisfying $x^{\mathsf{T}}Qx \leq b$ is an *ellipsoid* if $Q \succ 0$.
- If $Q \succ 0$, then $x^{\mathsf{T}}Qx \leq b \iff \|Q^{1/2}x\|^2 \leq b$.
- (Recall that if we write the eigenvalue decomposition $Q = U\Lambda U^T$, then $Q^{1/2} = U\Lambda^{1/2}U^T$, where $\Lambda^{1/2}$ is the diagonal matrix whose diagonal entries are the square roots of the diagonals of Λ.)

Degenerate Ellipsoids

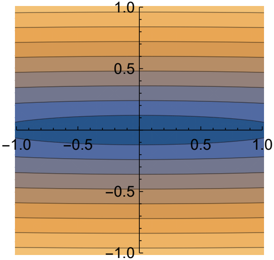
Ellipsoid axes have length $\frac{1}{\sqrt{\lambda_i}}$. When an eigenvalue is close to zero, contours are stretched in that direction.



$x^2 + y^2$



$\frac{1}{10}x^2 + y^2$



$\frac{1}{100}x^2 + y^2$

- Warmer colors = larger values
- If $\lambda_i = 0$, then $Q \succeq 0$. The ellipsoid $x^\top Q x \leq 1$ is *degenerate* (stretches out to infinity (is constant) in direction u_i).

Ellipsoids with linear terms

If $Q \succ 0$, then the quadratic form with extra linear term:

$$x^{\mathsf{T}} Q x + r^{\mathsf{T}} x + s$$

defines an *shifted* ellipsoid, whose center is not at 0. To see why, complete the square!
For scalars (ellipsoids in $\mathbb{R}^1$ are not very interesting), we have:

$$qx^2 + rx + s = q\left(x + \frac{r}{2q}\right)^2 + \left(s - \frac{r^2}{4q}\right)$$

In the matrix case, we have:

$$x^{\mathsf{T}} Q x + r^{\mathsf{T}} x + s = \left(x + \tfrac{1}{2} Q^{-1} r\right)^{\mathsf{T}} Q \left(x + \tfrac{1}{2} Q^{-1} r\right) + \left(s - \tfrac{1}{4} r^{\mathsf{T}} Q^{-1} r\right)$$

Ellipsoids with linear terms

Therefore, the inequality $x^\mathsf{T} Q x + r^\mathsf{T} x + s \le b$ is equivalent to:

$$\left(x + \tfrac{1}{2} Q^{-1} r\right)^\mathsf{T} Q \left(x + \tfrac{1}{2} Q^{-1} r\right) \le \left(b - s + \tfrac{1}{4} r^\mathsf{T} Q^{-1} r\right)$$

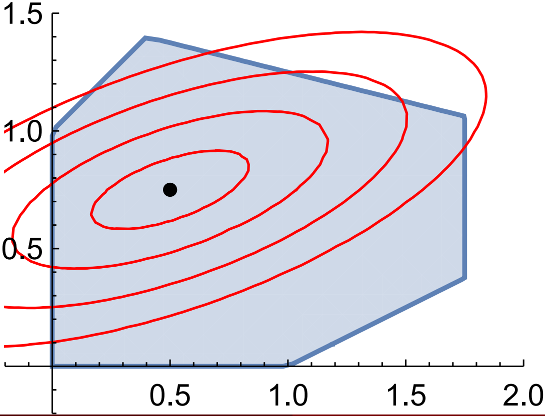
This is an ellipse centered at $-\frac{1}{2} Q^{-1} r$ — but its shape is still defined by the matrix Q.

Writing this using the matrix square root, we have:

$$\left\| Q^{1/2} x + \tfrac{1}{2} Q^{-1/2} r \right\|^2 \le \left(b - s + \tfrac{1}{4} r^\mathsf{T} Q^{-1} r\right)$$

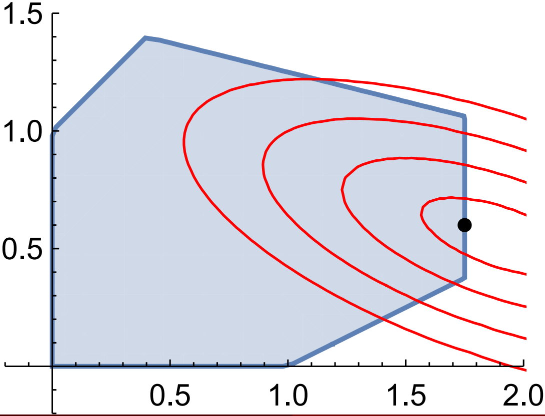
Quadratic programs

$$\begin{array}{ll} \underset{x}{\text{minimize}} & x^\top P x + q^\top x + r \\ \text{subject to:} & Ax \leq b \end{array}$$



Quadratic programs

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & x^\top P x + q^\top x + r \\ \text{subject to:} \quad & Ax \leq b \end{aligned}$$



QCQPs

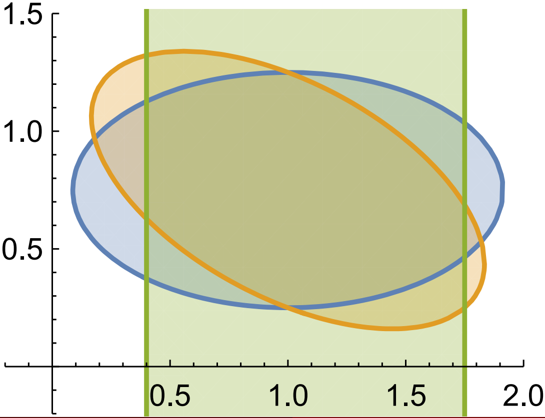
Quadratically constrained quadratic program (QCQP) has both a quadratic cost and quadratic constraints:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & x^\mathsf{T} P_0 x + q_0^\mathsf{T} x + r_0 \\ \text{subject to:} \quad & x^\mathsf{T} P_i x + q_i^\mathsf{T} x + r_i \leq 0 \quad \text{for } i = 1, \dots, m \end{aligned}$$

- If $P_i \succeq 0$ for $i = 0, 1, \dots, m$, it is a *convex QCQP*
  - feasible set is convex
  - solution can be on boundary or in the interior
  - relatively easy to solve
- If any $P_i \not\succeq 0$, the QCQP becomes **very hard** to solve.

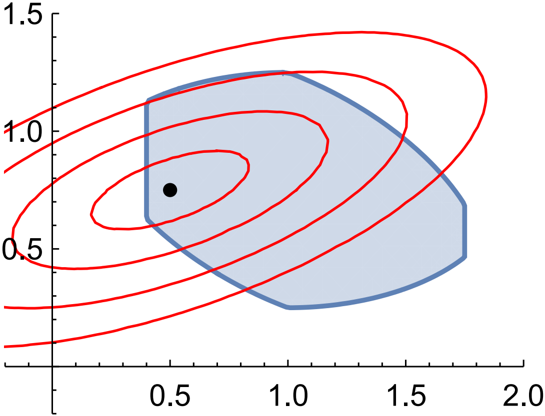
QCQPs

$$\begin{array}{ll} \underset{x}{\text{minimize}} & x^\top P_0 x + q_0^\top x + r_0 \\ \text{subject to:} & x^\top P_i x + q_i^\top x + r_i \le 0 \quad \text{for } i = 1, \ldots, m \end{array}$$



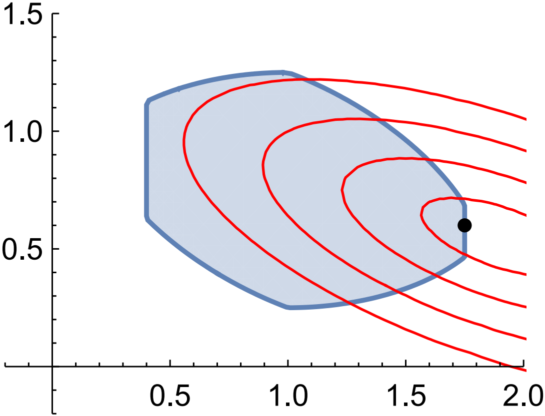
QCQPs

$$\begin{array}{ll} \underset{x}{\text{minimize}} & x^\top P_0 x + q_0^\top x + r_0 \\ \text{subject to:} & x^\top P_i x + q_i^\top x + r_i \le 0 \quad \text{for } i = 1, \dots, m \end{array}$$



QCQPs

$$\begin{array}{ll} \underset{x}{\text{minimize}} & x^\mathsf{T} P_0 x + q_0^\mathsf{T} x + r_0 \\ \text{subject to:} & x^\mathsf{T} P_i x + q_i^\mathsf{T} x + r_i \leq 0 \quad \text{for } i = 1, \ldots, m \end{array}$$

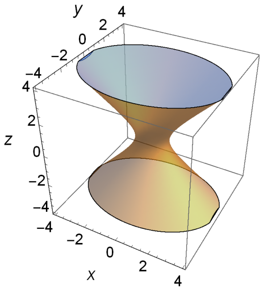


Difficult quadratic constraints

The following types of quadratic constraints make a problem nonconvex and generally difficult to solve (but not always).

Indefinite quadratic constraints.

- Example: $x^2 + 2y^2 - z^2 \leq 1$ corresponds to the nonconvex region on the right.
- **Note:** Be mindful of $\leq$ vs $\geq$! e.g. $x^2 + y^2 \geq 1$ is nonconvex.



Quadratic equalities.

- Using quadratic equalities, you can encode Boolean constraints. Example: $x^2 = 1$ is equivalent to $x \in \{-1, 1\}$. (There are many interesting problems with these kinds of variables!)

Where do quadratics commonly occur?

1. As a regularization or penalty term
   - (cost) $+ \lambda \|x\|^2$: standard L_2 regularizer
   - (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.

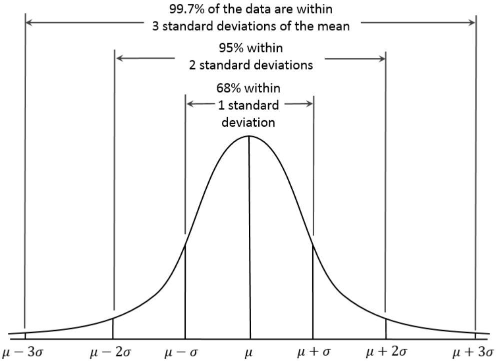
4. Energy quantities (physics/mechanics)
   - examples: $\underset{\text{(kinetic)}}{\tfrac{1}{2}mv^2}$, $\underset{\text{(spring)}}{\tfrac{1}{2}kx^2}$, $\underset{\text{(capacitor)}}{\tfrac{1}{2}CV^2}$, $\underset{\text{(rotational)}}{\tfrac{1}{2}I\omega^2}$, $\underset{\text{(strain)}}{\tfrac{1}{2}VE\varepsilon^2}$.

5. Covariance constraints (statistics)

Example: portfolio optimization

We must decide how to invest our money, and we can choose between $i = 1, 2, \ldots, N$ different assets.

- Each asset can be modeled as a random variable (RV) with an expected return μ_i and a standard deviation σ_i.



- Standard deviation is a measure of uncertainty.

Example: portfolio optimization

If Z is the RV representing an asset:

- The expected return is $\mu = \mathbf{E}(Z)$ (expected value)
- The variance is $\mathbf{var}(Z) = \sigma^2 = \mathbf{E}\left((Z-\mu)^2\right)$
- The standard deviation is the square root of the variance.
- Sometimes use the notation $Z \sim (\mu, \sigma^2)$.

If $Z_1 \sim (\mu_1, \sigma_1^2)$ and $Z_2 \sim (\mu_2, \sigma_2^2)$ are two RVs

- The covariance is $\mathbf{cov}(Z_1, Z_2) = \mathbf{E}\left((Z_1 - \mu_1)(Z_2 - \mu_2)\right)$.
- Note that: $\mathbf{var}(Z) = \mathbf{cov}(Z, Z)$
- covariance measures tendency of RVs to move together.

Example: portfolio optimization

If $Z_1 \sim (\mu_1, \sigma_1^2)$ and $Z_2 \sim (\mu_2, \sigma_2^2)$, what is $x_1 Z_1 + x_2 Z_2$?

Calculating the mean:

$$\begin{aligned}\mathbf{E}(x_1 Z_1 + x_2 Z_2) &= x_1 \mu_1 + x_2 \mu_2 \\ &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\end{aligned}$$

Calculating the variance:

$$\begin{aligned}\mathbf{var}(x_1 Z_1 + x_2 Z_2) &= \mathbf{E}\left(x_1(Z_1 - \mu_1) + x_2(Z_2 - \mu_2)\right)^2 \\ &= x_1^2\,\mathbf{var}(Z_1) + 2x_1 x_2\,\mathbf{cov}(Z_1, Z_2) + x_2^2\,\mathbf{var}(Z_2) \\ &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \mathbf{cov}(Z_1, Z_1) & \mathbf{cov}(Z_1, Z_2) \\ \mathbf{cov}(Z_2, Z_1) & \mathbf{cov}(Z_2, Z_2) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\end{aligned}$$

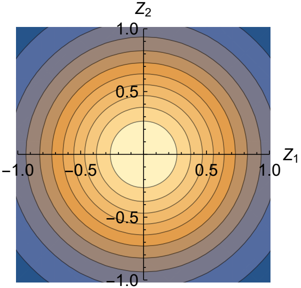
Example: portfolio optimization

If $Z_1, \dots, Z_n$ are **jointly distributed** with:

- mean $\mu = \begin{bmatrix} \mathbf{E}(Z_1) \\ \vdots \\ \mathbf{E}(Z_n) \end{bmatrix}$
- covariance matrix $\Sigma = \begin{bmatrix} \mathbf{cov}(Z_1, Z_1) & \dots & \mathbf{cov}(Z_1, Z_n) \\ \vdots & \ddots & \vdots \\ \mathbf{cov}(Z_n, Z_1) & \dots & \mathbf{cov}(Z_n, Z_n) \end{bmatrix}$
- short form: $Z \sim (\mu, \Sigma)$.

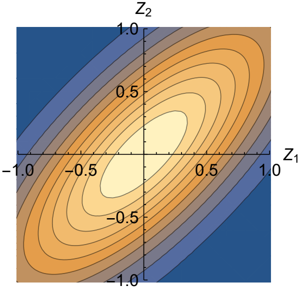
$$\sum_{i=1}^{n} x_i Z_i \sim \left(x^\mathsf{T}\mu \,,\, x^\mathsf{T}\Sigma x\right)$$

Example: portfolio optimization



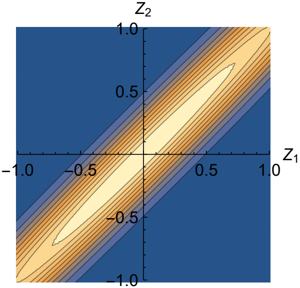
uncorrelated

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



somewhat correlated

$$\Sigma = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$$



highly correlated

$$\Sigma = \begin{bmatrix} 1 & .99 \\ .99 & 1 \end{bmatrix}$$

Correlation is modeled by a **confidence ellipsoid**

Example: portfolio optimization

Example:

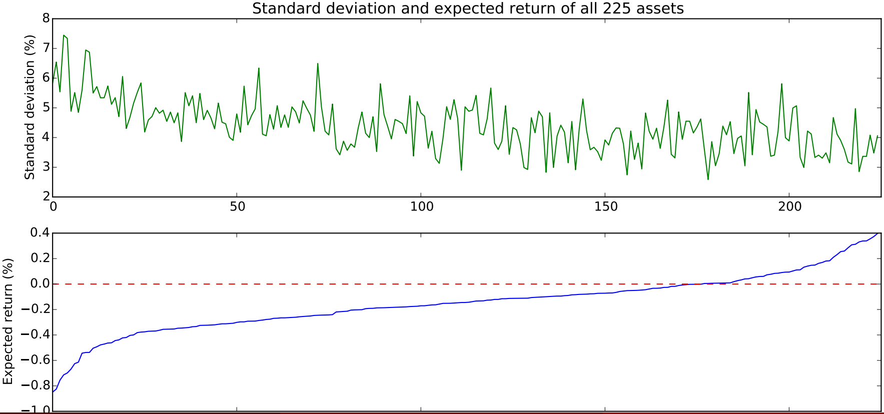
- There are 16 different stocks: $Z_1, \ldots, Z_{16}$. Each has expected return of 2% with standard deviation of 5%.
  You have \$100 in total to invest.
- If you invest in just one of them, you will earn $\$102 \pm \$5$.
- If the stocks are all correlated (e.g. all the same industry) and you invest evenly in all stocks, you still earn: $\$102 \pm \$5$.
- If the stocks are **uncorrelated** (e.g. very diverse) and you invest evenly in all stocks, the new variance is $16 \times (\frac{5}{16})^2$. Therefore, you will earn $\$102 \pm \$1.25$.

Julia code: Portfolio.ipynb

Example: portfolio optimization

Dataset containing 225 assets. How should we invest?

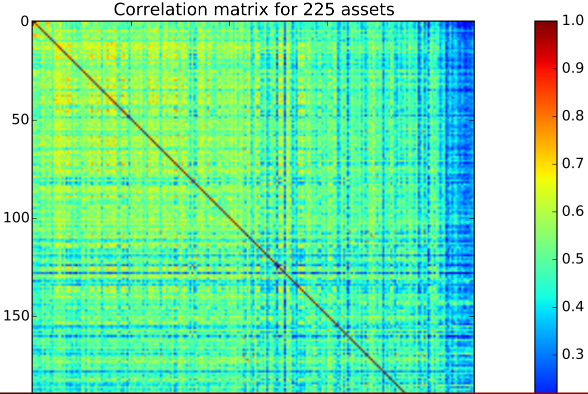
- We know the expected return μ_i for each asset
- We know the covariance Σ_{ij} for each pair of assets



Example: portfolio optimization

Dataset containing 225 assets. How should we invest?

- We know the expected return μ_i for each asset
- We know the covariance Σ_{ij} for each pair of assets



Example: portfolio optimization

Suppose we buy x_i of asset Z_i. We want:

- A high total return. Maximize $x^\mathsf{T}\mu$.
- Low variance (risk). Minimize $x^\mathsf{T}\Sigma x$.

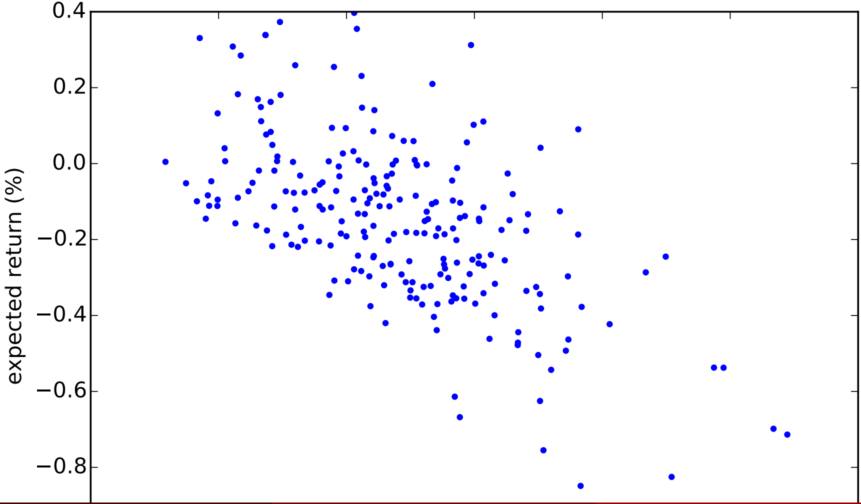
Pose the optimization problem as a tradeoff:

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & -x^\mathsf{T}\mu + \lambda x^\mathsf{T}\Sigma x \\
\text{subject to:} \quad & x_1 + \cdots + x_{225} = 1 \\
& x_i \ge 0
\end{aligned}
$$

Fun fact: This is the basic idea behind "Modern portfolio theory". Introduced by economist Harry Markowitz in 1952, for which he was awarded the Nobel Memorial Prize in Economics in 1990.

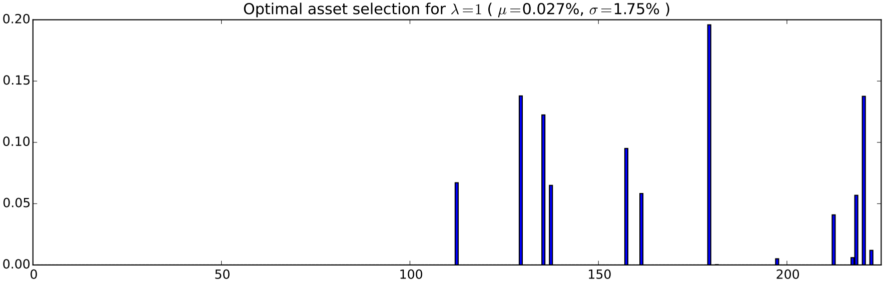
Example: portfolio optimization

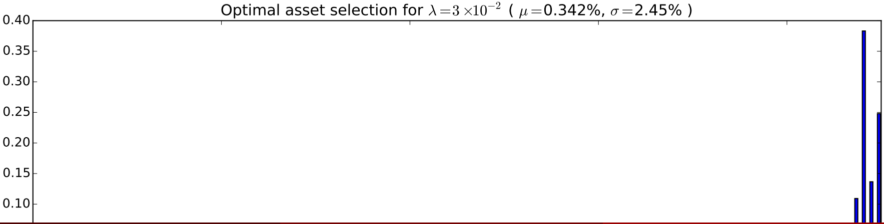
Quality of each individual asset:



Example: portfolio optimization

Some solutions:





Example: portfolio optimization

Pareto curve ("efficient frontier"). Note that most "random" portfolios are far away from the frontier!