ENE 2XX: Renewable Energy Systems and Control

LEC 02 : Convex Programs

Professor Scott Moura
University of California, Berkeley

Summer 2017

What is an Optimization Program?

- We seek “the best” values for design variables $x \in \mathbb{R}^n$
- Must respect certain constraints / limitations

$$\begin{aligned} \text{minimize} \quad & f(x) && \text{[Objective Function]} \\ \text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \cdots, m && \text{[Inequality constraints]} \\ & h_j(x) = 0, \quad j = 1, \cdots l && \text{[Equality constraints]} \end{aligned}$$

What is an Optimization Program?

- We seek "the best" values for design variables $x \in \mathbb{R}^n$
- Must respect certain constraints / limitations

$$\begin{aligned} \text{minimize} \quad & f(x) & & \text{[Objective Function]} \\ \text{subject to} \quad & g(x) \leq 0 & & \text{[Inequality constraints]} \\ & h(x) = 0 & & \text{[Equality constraints]} \end{aligned}$$

Vector notation

What is an Optimization Program?

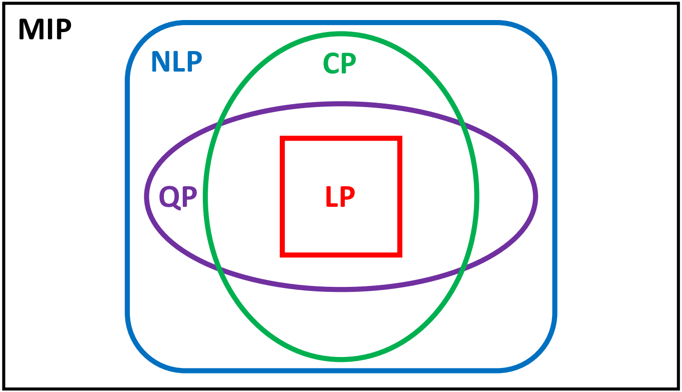
- We seek "the best" values for design variables $x \in \mathbb{R}^n$
- Must respect certain constraints / limitations

$$
\begin{aligned}
\text{minimize} \quad & f(x) && \text{[Objective Function]} \\
\text{subject to} \quad & g(x) \leq 0 && \text{[Inequality constraints]} \\
& h(x) = 0 && \text{[Equality constraints]}
\end{aligned}
$$

Vector notation

A value $x^\star$ that solves this optimization program is called a "minimizer".

Classes of Optimization Programs



LP = Linear Program; QP = Quadratic Program; CP = Convex Program;
NLP = Nonlinear Program; MIP = Mixed Integer Program

Outline

Convex Programs

A *convex optimization problem* has the form

$$\begin{aligned} \text{minimize} \quad & f(x) & (1)\\ \text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \cdots, m & (2)\\ & a_j^T x = b_j, \quad j = 1, \cdots, l. & (3) \end{aligned}$$

Comparing this problem with the abstract optimization problem defined before, the *convex optimization problem* has three additional requirements:

- objective function $f(x)$ must be convex,
- inequality constraint functions $g_i(x)$ must be convex for all $i = 1, \cdots, m$,
- the equality constraint functions $h_j(x)$ must be affine for all $j = 1, \cdots, l$.

Note that in the convex optimization problem, we can only tolerate affine equality constraints, meaning (3) takes the matrix-vector form of $A_{eq}x = b_{eq}$.

Why care?

- No general analytic solutions, however VERY powerful methods exist to solve CPs numerically
- Ex: Easily solve CPs with 100's or 1000's of variables in just a few seconds
- Ex: Easily solve CPs with 1M's of variables in tens of seconds
- CP solvers are off-the-shelf technology
- YOUR focus: Find ways to convert your problem into a CP
- If you formulate your problem into a CP, then you have essentially solved it
- Converting your problem into a CP requires both art & technical skill

Key CP Properties

- If a local minimum exists, then it is the global minimum.
- If the objective function is strictly convex, and a local minimum exists, then it is a unique minimum.

Sub-classes of Convex Programs

- Linear Programs (LPs)
- <u>Some</u> Quadratic Programs (QPs)
- Second Order Cone Programs (SOCPs)
- Maximum Likelihood Estimation (MLE)
- Geometric Programs (GPs)
- Semidefinite Programs (SDPs)

Exercise

Is this a convex program?

$$\begin{aligned} \text{minimize} \quad & f(x) = x_1^2 + x_2^2 && (4) \\ \text{subject to} \quad & g_1(x) = x_1/(1 + x_2^2) \leq 0 && (5) \\ & h_1(x) = (x_1 + x_2)^2 = 0 && (6) \end{aligned}$$

Exercise

Is this a convex program?

$$\begin{aligned} \text{minimize} \quad & f(x) = x_1^2 + x_2^2 && (4) \\ \text{subject to} \quad & g_1(x) = x_1/(1 + x_2^2) \leq 0 && (5) \\ & h_1(x) = (x_1 + x_2)^2 = 0 && (6) \end{aligned}$$

NOT a convex program.

- Inequality constraint function $g_1(x)$ is not convex in (x_1, x_2)
- Equality constraint function $h_1(x)$ is not affine in (x_1, x_2)

Exercise

Is this a convex program?

$$
\begin{aligned}
\text{minimize} \quad & f(x) = x_1^2 + x_2^2 && (4)\\
\text{subject to} \quad & g_1(x) = x_1/(1+x_2^2) \leq 0 && (5)\\
& h_1(x) = (x_1+x_2)^2 = 0 && (6)
\end{aligned}
$$

NOT a convex program.

- Inequality constraint function $g_1(x)$ is not convex in (x_1, x_2)
- Equality constraint function $h_1(x)$ is not affine in (x_1, x_2)

Now, an astute observer might comment that both sides of (5) can be multiplied by $(1 + x_2^2)$ and (6) can be represented simply by $x_1 + x_2 = 0$, without loss of generality.

Exercise

Is this a convex program?

$$\begin{aligned} \text{minimize} \quad & f(x) = x_1^2 + x_2^2 && (4) \\ \text{subject to} \quad & g_1(x) = x_1 \leq 0 && (5) \\ & h_1(x) = x_1 + x_2 = 0 && (6) \end{aligned}$$

YES. This is a convex program.

- Objective function $f(x)$ is convex in (x_1, x_2)
- Inequality constraint function $g_1(x)$ is convex in (x_1, x_2)
- Equality constraint function $h_1(x)$ is affine in (x_1, x_2)

Outline

Linear Programs

Linear program (LP) is defined as the following special case of a CP:

$$\begin{aligned} \text{minimize} \quad & c^T x && (7) \\ \text{subject to} \quad & Ax \leq b && (8) \\ & A_{eq} x = b_{eq} && (9) \end{aligned}$$

- $f(x)$ must be linear (or affine, before dropping the additive constant)
- $g_i(x)$ and $h_j(x)$ must be affine for all i and j, respectively.

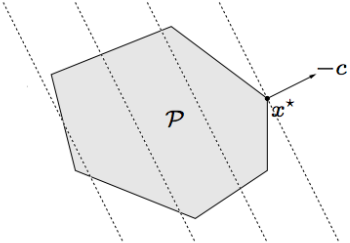


Figure: Feasible set of LPs always forms a polyhedron $\mathcal{P}$. Objective function visualized as isolines of constant cost (dotted lines). The optimal solution is at the boundary point that touches the isoline of least cost.

Nature of LP Solutions

Proposition (Nature of LP Solutions)

The solution to any linear program is characterized by one of the following three categories:

- **[No Solution]** Occurs when feasible set is empty, or objective function is unbounded.
- **[One Unique Solution]** There exists a single unique solution at the vertex of the feasible set. That is, at least two constraints are active and their intersection gives the optimal solution. (see previous slide)
- **[A Non-Unique Solution]** There exists an infinite number of solutions, given by one edge of the feasible set. That is, one or more constraints are active and all solutions along the intersection of these constraints are equally optimal. This can only occur when the objective function gradient is orthogonal to one or multiple constraint.

LP Examples

Diet Problem: choose quantities $x_1, \cdots, x_n$ of n foods

- one unit of food j costs c_j, contains amount a_{ij} of nutrient i
- healthy diet requires nutrient i in quantity at least b_i

to find cheapest health diet

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b, \quad x \geq 0 \end{array}$$

LP Examples

Diet Problem: choose quantities $x_1, \cdots, x_n$ of n foods

- one unit of food j costs c_j, contains amount a_{ij} of nutrient i
- healthy diet requires nutrient i in quantity at least b_i

to find cheapest health diet

$$\begin{array}{rl} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b, \quad x \geq 0 \end{array}$$

Minimize a piecewise affine (PWA) function:

$$\text{minimize} \qquad \max_{i=1,\cdots,m} \left\{ a_i^T x + b_i \right\}$$

is equivalent to the LP

$$\begin{array}{rl} \text{minimize} & t \\ \text{subject to} & a_i^T x + b_i \leq t, \quad \forall\, i = 1, \cdots, m \end{array}$$

Optimal Economic Dispatch

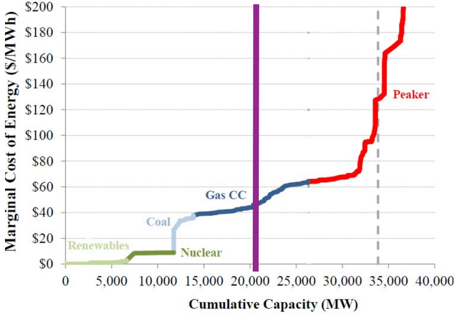
You are the California Independent System Operator (CAISO). You must schedule power generators for tomorrow (24 one-hour segments) to satisfy electricity demand. Given data:

- Generator i provides "marginal cost" c_i (units of USD/MW). Quantity c_i is financial compensation each generator requests for providing 1 MW.
- Generator i has maximum power capacity of $x_{i,\max}$ (units of MW).
- California electricity demand is $D(k)$, where k indexes each hour, i.e. $k = 0, 1, \cdots, 23$.

$$\begin{aligned} \text{minimize} \quad & \sum_{k=0}^{23} \sum_{i=1}^{n} c_i x_i(k) && (10) \\ \text{subject to} \quad & 0 \leq x_i(k) \leq x_{i,\max}, \quad \forall\, i = 1, \cdots, n, \quad k = 0, \cdots, 23 && (11) \\ & \sum_{i=1}^{n} x_i(k) = D(k), \quad k = 0, \cdots, 23 && (12) \end{aligned}$$

optimization var $x_i(k)$ is power produced by generator i during hour k.

Optimal Economic Dispatch



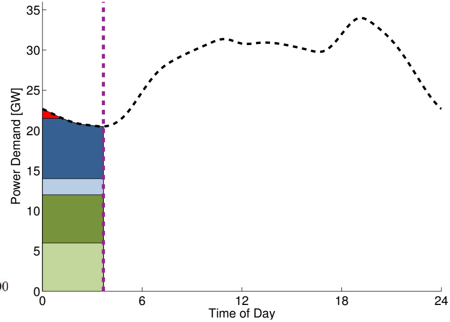


Figure: [LEFT] Marginal cost of electricity for various generators, as a function of cumulative capacity. The purple line indicates the total demand $D(k)$. All generators left of the purple line are dispatched. [RIGHT] Optimal supply mix and demand for 03:00.

Optimal Economic Dispatch

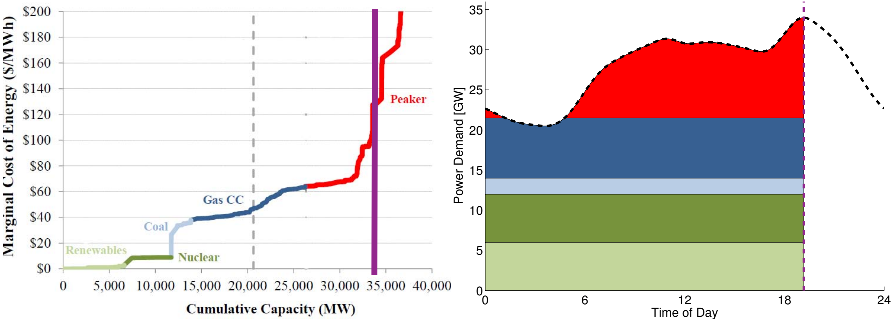


Figure: [LEFT] Marginal cost of electricity for various generators, as a function of cumulative capacity. The purple line indicates the total demand $D(k)$. All generators left of the purple line are dispatched. [RIGHT] Optimal supply mix and demand for 19:00.

Outline

Quadratic Programs

Quadratic program (QP) is defined as:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}x^TQx + R^Tx + S && (10) \\ \text{subject to} \quad & Ax \leq b && (11) \\ & A_{eq}x = b_{eq} && (12) \end{aligned}$$

- $f(x)$ must be quadratic in x
- $g_i(x)$ and $h_j(x)$ must be affine for all i and j, respectively.

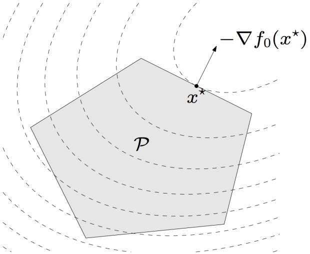


Figure: Feasible set of QPs always forms a polyhedron $\mathcal{P}$. Objective function visualized as convex quadratic iso-countours of constant cost (dotted lines).

Quadratic Programs

Quadratic program (QP) is defined as:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}x^T Q x + R^T x + S && (10) \\ \text{subject to} \quad & Ax \leq b && (11) \\ & A_{eq} x = b_{eq} && (12) \end{aligned}$$

- $f(x)$ must be quadratic in x
- $g_i(x)$ and $h_j(x)$ must be affine for all i and j, respectively.

Remark

Not all QPs are convex programs! A QP is a convex program only if $Q \succeq 0$, *i.e.* Q *is positive semi-definite. QPs where* $Q \not\succeq 0$ *are called non-convex QPs and are generally very hard to solve.*

Linear Regression Models

more specifically, linear-in-the-parameters models

Suppose you have data comprised of n-data pairs (x_i, y_i), where $i = 1, \cdots, n$. You seek to fit a mathematical model to this data, of the form:

$$y = \theta_1 x + \theta_0 \tag{13}$$

How do we determine θ_1, θ_0?

Regression Analysis

Establish a mathematical relationship between variables, given data.

Quoted Text Message from Tech IP Attorney to Me

Attorney: One of our outside consultant firms billed us 170,000 USD to do an SQL regression model

Me: My undergrads would do that for 25 USD and pizza.

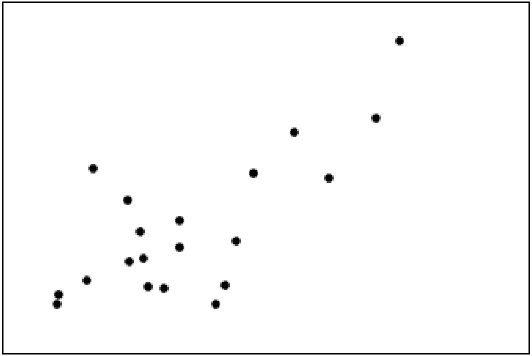
Attorney: yeah, next time we should go that route

Graphical Version

Determine a “best fit” for m, b in the linear model

$$y = mx + b \tag{14}$$

given n-data pairs (x_i, y_i), where $i = 1, \cdots, n$.
In other words, find the line that best fits data points:



Least Squares

a.k.a. Ordinary Least Squares (OLS) or Linear Least Squares

Let us define best fit as follows. Define the "residual" r_i for m, b and data pair (x_i, y_i) as follows:

$$r_i = mx_i + b - y_i \tag{15}$$

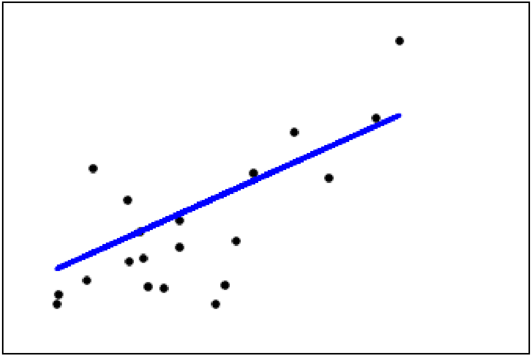
Obviously, when r_i then m, b fit that data pair perfectly. We would like to select m, b such that the sum of all residuals squared are minimized:

$$\begin{aligned} \min_{m,b} \sum_{i=1}^{i=n} r_i^2 &= \min_{m,b} \sum_{i=1}^{i=n} (mx_i + b - y_i)^2 & (16) \\ &= \min_{\theta=[m,b]} \|X\theta - Y\|_2^2 & (17) \end{aligned}$$

where

$$\theta = \begin{bmatrix} m \\ b \end{bmatrix}, \quad X = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{18}$$

Graphical Result



Other Linear-in-the-Parameter Models

Polynomial: $y = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_p x^p$. Residual $r = X\theta - Y$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_p \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ 1 & x_2 & x_2^2 & \cdots & x_2^p \\ \vdots & \vdots & & \vdots & \\ 1 & x_n & x_n^2 & \cdots & x_n^p \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{19}$$

Harmonic: $y = \theta_1 \sin(x) + \theta_2 \cos(x) + \theta_3 \sin(2x) + \theta_4 \cos(2x)$.
Residual $r = X\theta - Y$

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}, \quad X = \begin{bmatrix} \sin(x_1) & \cos(x_1) & \sin(2x_1) & \cos(2x_1) \\ \sin(x_2) & \cos(x_2) & \sin(2x_2) & \cos(2x_2) \\ \vdots & \vdots & & \vdots \\ \sin(x_n) & \cos(x_n) & \sin(2x_n) & \cos(2x_n) \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{20}$$

Other Linear-in-the-Parameter Models

Radial Basis Function: $y = \theta_1 e^{-(x+0.5)^2} + \theta_2 e^{-(x)^2} + \theta_3 e^{-(x-0.5)^2}$.
Residual $r = X\theta - Y$

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}, \quad X = \begin{bmatrix} e^{-(x_1+0.5)^2} & e^{-(x_1)^2} & e^{-(x_1-0.5)^2} \\ e^{-(x_2+0.5)^2} & e^{-(x_2)^2} & e^{-(x_2-0.5)^2} \\ \vdots & \vdots & \vdots \\ e^{-(x_n+0.5)^2} & e^{-(x_n)^2} & e^{-(x_n-0.5)^2} \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{21}$$

limited only by your imagination

Optimization Perspective

All regression problems for linear-in-the-parameters models can be written:

$$\text{minimize}_\theta \|X\theta - Y\|_2^2, \qquad X \in \mathbb{R}^{n\times p}, \theta \in \mathbb{R}^p, Y \in \mathbb{R}^n \tag{22}$$

n : number of data pairs (x_i, y_i)
p : number of coefficients $\theta_1, \cdots, \theta_p$.
We assume $n > p$.

Recall First Order Necessary Condition (FONC): If θ^* minimizes (22), then $\frac{d}{d\theta}\|X\theta - Y\|_2^2 = 0$. Let's expand this condition!

$$\begin{aligned}
0 &= \frac{d}{d\theta}\|X\theta - Y\|_2^2 \\
&= \frac{d}{d\theta}(X\theta - Y)^T(X\theta - Y) \\
&= \frac{d}{d\theta}\left(\theta^T X^T X\theta - 2Y^T X\theta + Y^T Y\right) \\
&= 2X^T X\theta - 2X^T Y \\
&\Rightarrow \boxed{\theta^* = \left(X^T X\right)^{-1} X^T Y}
\end{aligned}$$

Least Squares with L_2 Regularization

a.k.a. Ridge Regression

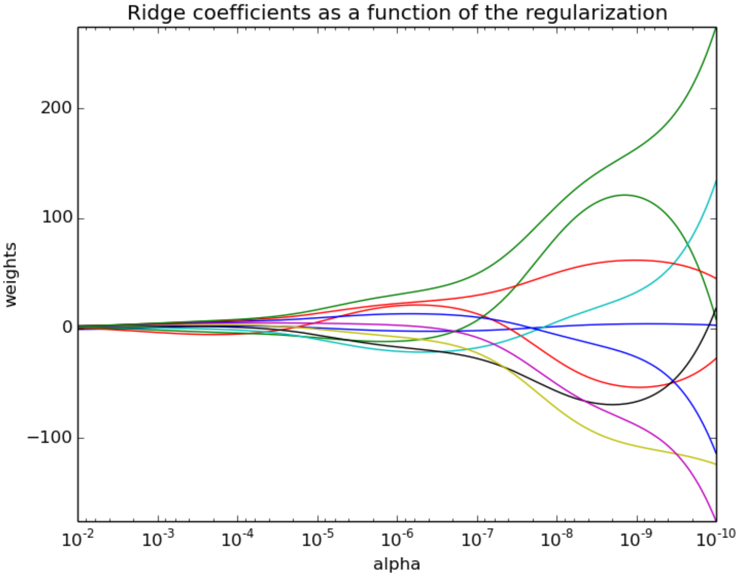
What if we define "best fit" by a different criterion? For example, we minimize the sum of residuals squared, but penalize the coefficients from getting "too big". Consider

$$\text{minimize}_\theta \quad \|X\theta - Y\|_2^2 + \alpha\|\theta\|_2^2 \tag{23}$$

Apply FONC:

$$\begin{aligned}
0 &= \frac{d}{d\theta}\|X\theta - Y\|_2^2 + \alpha\|\theta\|_2^2 \\
&= \frac{d}{d\theta}(X\theta - Y)^T(X\theta - Y) + \theta^T\theta \\
&= \frac{d}{d\theta}\left(\theta^T(X^TX + \alpha I)\theta - 2Y^TX\theta + Y^TY\right) \\
&= 2(X^TX + \alpha I)\theta - 2X^TY \\
\Rightarrow &\quad \boxed{\theta^* = \left(X^TX + \alpha I\right)^{-1} X^TY}
\end{aligned}$$

Ridge Coefficients as you vary α



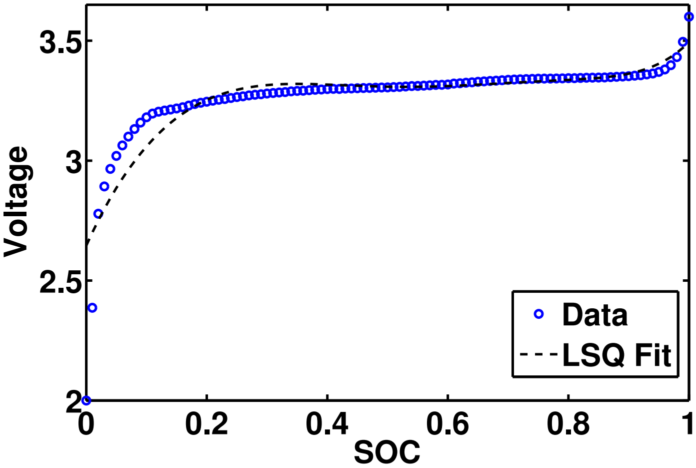
What if we define "best fit" by a different criterion? For example, suppose our data occasionally contains outliers that can bias our fitted linear model undesirably. Is there a "robust regression" method? Yes.

$$\min_{\theta} \|X\theta - Y\|_2^2 + \alpha\|\theta\|_1 \tag{24}$$

- L_2 penalties place small weight on small coefficients
- θ_i^2 is very small when θ_i is small
- Little incentive to drive θ_i to zero, unless you consider $|\theta_i|$ instead.

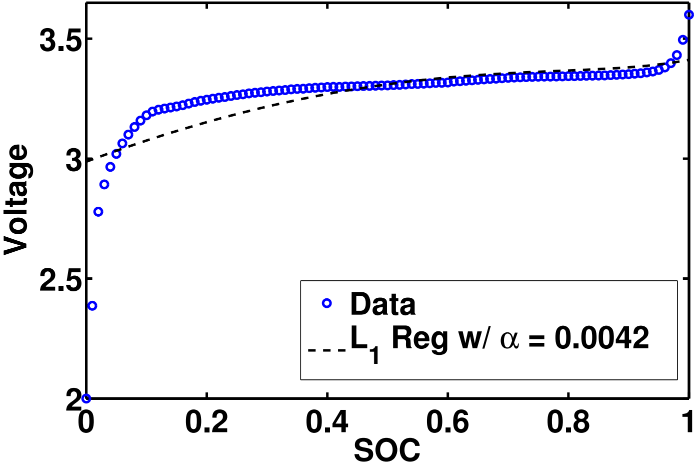
Note: Due to the 1-norm, this is no longer a QP! It is, however, a CP.

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_8 x^8$$



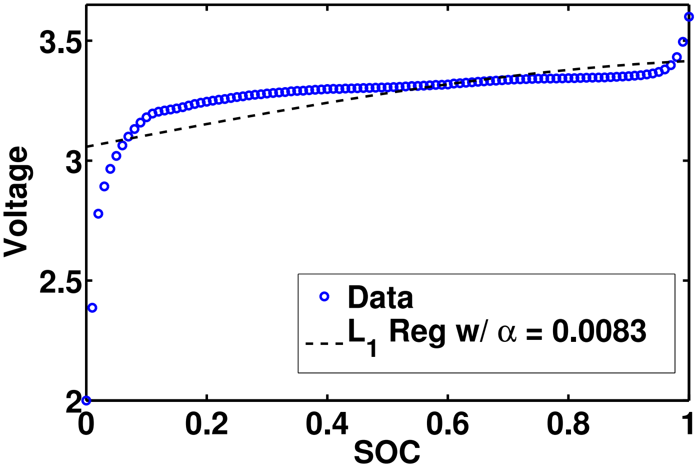
$\theta_0 = 2.6460; \theta_1 = 5.5442; \theta_2 = -15.7690; \theta_3 = 16.4894; \theta_4 = -0.9965;$
$\theta_5 = -4.2202; \theta_6 = -2.8927; \theta_7 = 0.0602; \theta_8 = 2.6326;$

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_8 x^8$$



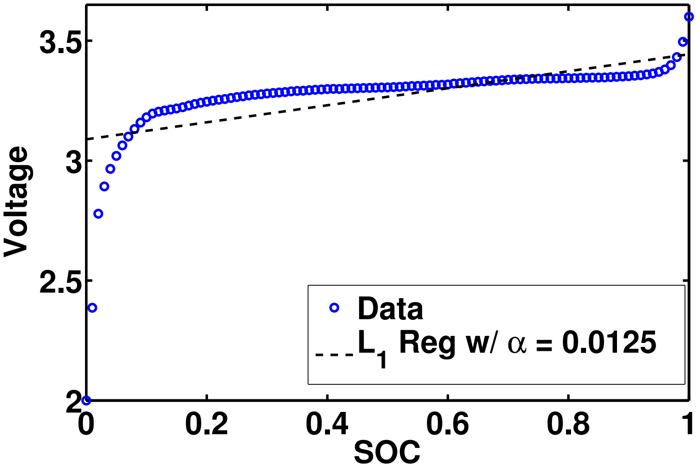
$\theta_0 = 2.9873; \theta_1 = 0.9366; \theta_2 = -0.5531; \theta_3 = -0.0641; \theta_4 = 0;$
$\theta_5 = 0; \theta_6 = 0; \theta_7 = 0; \theta_8 = 0.1052$

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_8 x^8$$



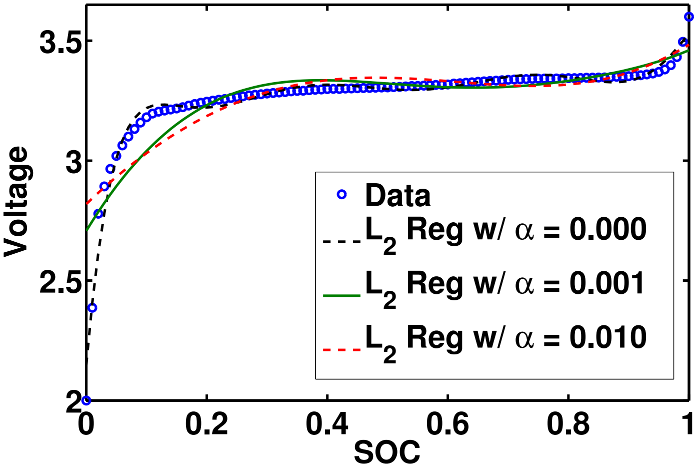
$\theta_0 = 3.0579; \theta_1 = 0.4773; \theta_2 = 0; \theta_3 = -0.1202; \theta_4 = 0;$
$\theta_5 = 0; \theta_6 = 0; \theta_7 = 0; \theta_8 = 0$

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_8 x^8$$



$\theta_0 = 3.0889; \theta_1 = 0.3547; \theta_2 = 0; \theta_3 = 0; \theta_4 = 0;$
$\theta_5 = 0; \theta_6 = 0; \theta_7 = 0; \theta_8 = 0$

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_8 x^8$$



$\theta_0 = 2.1; \theta_1 = 28.8; \theta_2 = -301; \theta_3 = 1595; \theta_4 = -4743; \theta_5 = 8222; \theta_6 = -8238; \theta_7 = 4414; \theta_8 = -977$

$\theta_0 = 2.71; \theta_1 = 4.13; \theta_2 = -8.51; \theta_3 = 4.10; \theta_4 = 3.42; \theta_5 = -0.34; \theta_6 = -2.31; \theta_7 = -1.49; \theta_8 = 1.75$

$\theta_0 = 2.82; \theta_1 = 2.40; \theta_2 = -2.81; \theta_3 = -0.33; \theta_4 = 0.71; \theta_5 = 0.69; \theta_6 = 0.32; \theta_7 = -0.04; \theta_8 = -0.27$

A Generalized Linear Model

The generalized linear model is given by:

$$y = \sum_{i=1}^{p} \theta_i \phi_i(x) = \theta^T \phi(x) \qquad (25)$$

- $y \in \mathbb{R}$ is the output of interest
- $\theta \in \mathbb{R}^p$ are the coefficients or parameters to fit
- $\phi(x)$ are "regressors" or "predictors", which can involve dependent data in a nonlinear way

Summary of Regression Procedures

- Least Squares (LSQ), a.k.a. linear least squares, ordinary least squares (convex QP)
- LSQ w/ L_2 Regularization, a.k.a. ridge regression (convex QP)
- LSQ w/ L_1 Regularization, a.k.a. lasso regression (CP)
- LSQ w/ L_1 and L_2 Regularization, a.k.a. elastic net (CP)
- LSQ w/ Huber Regularization (Hybridized L_1+L_2), a.k.a. Robust LSQ (CP)

Markowitz Portfolio Optimization - I

Problem Statement: Imagine you are an investment portfolio manager. You control a large sum of money, and can invest in n different assets. At the end of some time period, your investment produces a financial return. The key challenge, here, is the return is not easily predictable. It is random.

Notation:

- x_i denotes the percentage of fund to invest in asset i. Note that $\sum_{i=1}^{n} x_i = 1$, and $x_i \geq 0$
- Return is well characterized by Gaussian distribution $\mathcal{N}(\mu, \Sigma)$, where $\mu \in \mathbb{R}^n$ is expected return and $\Sigma \in \mathbb{R}^{n \times n}$ is the covariance

Examples:

- Asset i has expected return $\mu_i =$2%, with std dev of $\sqrt{\Sigma_{ii}} =$5%
- Asset j has expected return $\mu_i =$5%, with std dev of $\sqrt{\Sigma_{ii}} =$50%

Markowitz Portfolio Optimization - II

Suppose we seek to

- maximize expected return, AND
- minimize risk

These two objectives cannot be achieved w/o tradeoffs. Therefore, one often "scalarizes" this bi-criterion problem to explore the tradeoff:

$$\begin{aligned} \text{minimize} \quad & -\mu^T x + \gamma \cdot x^T \Sigma x && (26) \\ \text{subject to} \quad & \mathbb{1}^T x = 1, \quad x \succeq 0 && (27) \end{aligned}$$

where the parameter $\gamma \geq 0$ is called the "risk aversion" parameter.

- Increasing γ increases your sensitivity to risk
- $\gamma = 0$ means you are risk neutral
- $\gamma < 0$ means are are a risk seeker. Note this is NOT a convex QP.

Markowitz Portfolio Optimization - III

Consider this expected return & covariance data for a portfolio of 3 assets:

$$\mu = [1.02,\ 1.05,\ 1.04]^T, \quad \Sigma = \begin{bmatrix} (0.05)^2 & 0 & 0 \\ 0 & (0.5)^2 & 0 \\ 0 & 0 & (0.1)^2 \end{bmatrix} \tag{28}$$

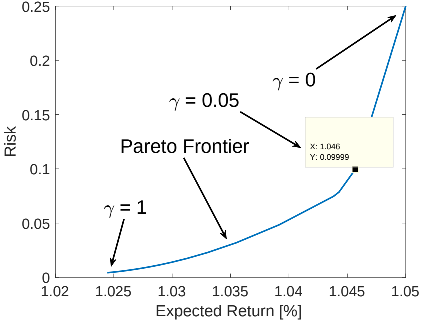


Figure: Trade off between maximizing expected return and minimizing risk. This trade off curve is called a “Pareto Frontier”

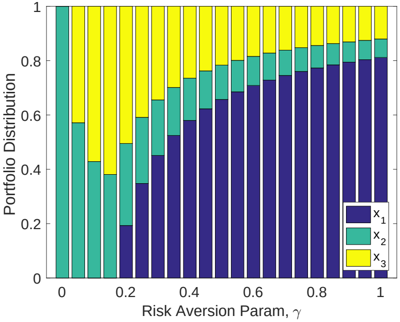


Figure: Optimal portfolio investment strategy, as risk aversion parameter γ increases.

Quadratically Constrained QPs

A generalization of the convex QP problem is the quadratically constrained QP (QCQP):

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}x^T Q x + R^T x + S && (29) \\ \text{subject to} \quad & \frac{1}{2}x^T Q_i x + R_i^T x + S_i \leq 0, \quad \forall\, i = 1, \cdots, m && (30) \\ & A_{eq} x = b_{eq} && (31) \end{aligned}$$

where $Q, Q_i \succeq 0$ for the program to be convex.

Outline

Second Order Cone Programs

Second Order Cone Program (SOCP) is defined as:

$$\text{minimize} \quad f^T x \tag{32}$$

$$\text{subject to} \quad \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \cdots, m \tag{33}$$

$$A_{eq} x = b_{eq} \tag{34}$$

- Inequalities form a “second order cone” constraint

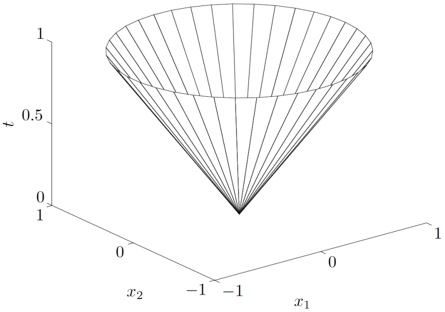


Figure: Boundary of second-order cone in $\mathbb{R}^3$, $\{(x_1, x_2, t) | (x_1^2 + x_2^2)^{1/2} \leq t\}$.