1.1

N. L. P. Formulation examples & techniques
Katta G. Murty, IOE 611 Lecture slides

Rectangular Rail Car Design: Top & bottom sheet $325/m^2$. Siding sheet $175/m^2$. Volume should \( \geq 1000 \ m^3 \). Height should \( \leq 3 \ m \). Total area of sides should \( \leq 75 \ m^2 \). Find min cost design.

Simple Portfolio Model: Data

\[
\begin{align*}
\mathbf{b} &= \text{budget ($ available to invest)} \\
\mathbf{n} &= \text{number investment opportu. available} \\
\mu_j &= \text{expected fraction return in } j\text{th opportu., } j = 1 \text{ to } n \\
\mathbf{D} &= n \times n \text{ variance-covariance matrix of fraction returns.}
\end{align*}
\]

\((\mu_j), \mathbf{D}\) usually estimated from past data.

Actual returns random variables. Two important objective functions. One to max expected return. 2nd to min risk, measured by variance of total fraction return.
Banks, mutual fund operators use models like this (mostly based on QP models) to plan investments.
Curve Fitting, Parameter estimation in non-linear modeling

In above, objective and constraint functions came with problem. Not common in NLP modeling.

In most, functional forms of objective and constraint functions to be determined empirically. This work called Curve Fitting. Two phases in it.

Phase I: Select Model Function: Either based on theoretical considerations, or by guessing suitable one looking at plots of data.

Phase II: Parameter Estimation: Usually model function has parameters whose values to be determined to give best fit to data.

Input for curve fitting:

INDEPENDENT VARIABLES: $x = (x_1, \ldots, x_n)^T$

DEPENDENT VARIABLE: $y = \text{Characteristic we trying to}$
determine as func of $x$.

**DATA FOR PARAMETER ESTIMATION**: Observations on the values of $y$ at $m$ points in $x$-space, i.e.,

$$y = y^r \quad \text{when} \quad x = x^r, \quad r = 1 \text{ to } m$$

**MODEL FUNCTION SELECTED**: $f(a, x)$ where $a = (a_1, \ldots, a_s)^T$ is vector of parameters to be estimated.

**PARAMETER ESTIMATION**: **Case 1**: $m = s$: In this case try to determine the parameter vector $a$ from the equations

$$y^1 = f(a, x^1)$$

$$\vdots$$

$$y^m = f(a, x^m)$$

**Case 2**: $m > s$: Most commonly used method is Least squares method. It determines $a$ to min the least squares measure of deviation,
\[ L_2(a) = \sum_{r=1}^{m} (y^r - f(a, x^r))^2 \]

Least squares solution \( \bar{a} \) minimizes \( L_2(a) \). Minimum value \( L_2(\bar{a}) \) called Residue.

Residue measures error in fit. If residue small, \( f(\bar{a}, x) \) accepted as function representing \( y \) as a function of \( x \).

If residue large, two remedies:

- Model function selected inappropriate. Select a better one and try again.

- Least squares problem may be nonconvex, algorithm used may not have obtained good solution. Run algorithm with a different initial point, or use another algo.

If \( f(a, x) \) is linear in \( a \), measures of deviation \( L_1(a) \), \( L_\infty(a) \) are used, and parameter estimation carried out using LP.
Optimizing Nitric Acid Production Costs:

Important process variables:  \( P_1 = \) air compressor discharge pressure in psi

\( Q_1 = \) flow rate thro’ compressor in cfm at 70F & 14.7 psi

\( P_2 = \) CC outlet pressure in psi

\( Q_2 = \) portion of \( Q_1 \) going thro’ PRT in cfm

\( T_2 = \) inlet temp. at PRT in F

Curve fitting jobs for constructing model:

1. \( Q_1, P_1 \) related through an equation. With a linear model function, a good fit found:

\[
P_1 + 0.002032Q_1 = 186.916
\]

2. \( P_2, Q_2, T_2 \) related through an eq. Best linear fit was

\[
P_2 - 0.100027T_2 + 0.0234267Q_2 = 683.81
\]
3. $P_1 - P_2 = \text{pressure drop across process train, is function of } Q_1, T_2$. Best linear fit was:

$$P_1 - P_2 = 0.001104Q_1 + 5.4469$$

4. $C_{hp}$ is a function of $P_1, Q_1$. $PRT_{hp}$ is function of $Q_2, T_2, P_2$. Best fits are:

$$C_{hp} = 0.07706Q_1 - P_1^{0.3548} - 0.2Q_1$$

$$PRT_{hp} = 0.0003452Q_2T_2[1 - \frac{1.7835}{p_2^{0.2217}}]$$

Also, $ST_{hp} = C_{hp} - PRT_{hp}$ has an upper bound of 2000.

5. Costs estimated in terms of $$/\text{CF} \text{ of air processed through compressor, are}$

$$ST \text{ steam cost} = \frac{1.29066 \times 10^{-4} \times ST_{hp}}{Q_1}$$

$$CC \text{ fuel cost} = 4.31402 \times 10^{-9} T_2 - 5.86707 \times 10^{-6}$$
Labor cost  \[= 0.1025 \times 10^{-4} - 0.1970 \times 10^{-9} Q_1\]

Other costs  \[= 0.4333 \times 10^{-4} - 0.8055 \times 10^{-9} Q_1\]

So overall model is:
Minimize $1.290066 \times 10^{-4}(C_{hp} - PRT_{hp})/Q_1$

\[ + 4.31402 \times 10^{-9}T_2 - 1.0025 \times 10^{-9}Q_1 \]

subject to $100 \leq P_1 \leq 140$

\[ 85 \leq P_2 \leq 100 \]

$24,000 \leq Q_1 \leq 28,000$

$900 \leq T_2 \leq 1250$

$Q_1 - Q_2 \geq 0$

$Q_2 \geq 0$

$C_{hp} = 0.07706Q_1P_1^{0.3548} - 0.2Q_1$

$PRT_{hp} = 0.0003452Q_2T_2[1 - (1.7835/P_2^{0.2217})]$

$0 \leq C_{hp} - PRT_{hp} \leq 2000$

$P_1 + 0.002032Q_1 = 186.916$

$P_2 - 0.100027T_2 + 0.0234267Q_2 = 683.81$

$P_1 - P_2 = 0.001104Q_1 + 5.4469$

When model solved optimum solution reduced pro-
duction cost in $/ton of acid to 14.6 from 15.3 by operating conditions at time of study.

Boiler Shop Optimization:

Company has 5 boilers of different load ranges, working in parallel for generating steam.

<table>
<thead>
<tr>
<th>Boiler $i$</th>
<th>Boiler load range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower $\ell_i$</td>
</tr>
<tr>
<td>1</td>
<td>10 units</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
</tr>
</tbody>
</table>

Each boiler’s energy efficiency depends on load. $f_i(x_i)$ is efficiency for $i$th boiler when load is $x_i$ units of steam. We fit a cubic polynomial $a_{0i} + a_{1i}x_i + a_{2i}x_i^2 + a_{3i}x_i^3$ for $f_i(x_i)$. Best values of parameters are:
Model problem of allocating total load of 350 units of steam among 5 boilers to minimize total fuel costs.