Constrained Least Squares

Stephen Boyd

EE103
Stanford University

November 16, 2015
Outline

Linearly constrained least squares

Least-norm problem

Solving the constrained least squares problem
Least squares with equality constraints

- The (linearly) constrained least squares problem (CLS) is

\[
\text{minimize} \quad \| Ax - b \|^2 \\
\text{subject to} \quad Cx = d
\]

- Variable (to be chosen/found) is \( n \)-vector \( x \)
- \( m \times n \) matrix \( A \), \( m \)-vector \( b \), \( p \times n \) matrix \( C \), and \( p \)-vector \( d \) are problem data (i.e., they are given)
- \( \| Ax - b \|^2 \) is the objective function
- \( Cx = d \) are the equality constraints
- \( x \) is feasible if \( Cx = d \)
- \( \hat{x} \) is a solution of CLS if \( C\hat{x} = d \) and \( \| A\hat{x} - b \|^2 \leq \| Ax - b \|^2 \) holds for any \( n \)-vector \( x \) that satisfies \( Cx = d \)
Least squares with equality constraints

- CLS combines solving linear equations with least squares problem
- like a bi-objective least squares problem, with infinite weight on second objective $||C'x - d||^2$
Piecewise-polynomial fitting

▶ *piecewise-polynomial* \(\hat{f}\) has form

\[
\hat{f}(x) = \begin{cases} 
  p(x) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3 & x \leq a \\
  q(x) = \theta_5 + \theta_6 x + \theta_7 x^2 + \theta_8 x^3 & x > a
\end{cases}
\]

\((a \text{ is given})\)

▶ we require \(p(a) = q(a), \; p'(a) = q'(a)\)

▶ fit \(\hat{f}\) to data \((x_i, y_i), \; i = 1, \ldots, N\) by minimizing sum square error

\[
\sum_{i=1}^{N} (\hat{f}(x_i) - y_i)^2
\]

▶ can express as a constrained least squares problem
Example

\[ \hat{f}(x) \]

\[ p(x) \]

\[ q(x) \]

Linearly constrained least squares
constraints are (linear equations in $\theta$)

$$\begin{align*}
\theta_1 + \theta_2 a + \theta_3 a^2 + \theta_4 a^3 - \theta_5 - \theta_6 a - \theta_7 a^2 - \theta_8 a^3 &= 0 \\
\theta_2 + 2\theta_3 a + 3\theta_4 a^2 - \theta_6 - 2\theta_7 a - 3\theta_8 a^2 &= 0
\end{align*}$$

prediction error on $(x_i, y_i)$ is $a_i^T \theta - y_i$, with

$$(a_i)_j = \begin{cases} 
(1, x_i, x_i^2, x_i^3, 0, 0, 0, 0) & x_i \leq a \\
(0, 0, 0, 0, 1, x_i, x_i^2, x_i^3) & x_i > a
\end{cases}$$

sum square error is $\|A\theta - y\|^2$, where $a_i^T$ are rows of $A$
Outline

Linearly constrained least squares

Least-norm problem

Solving the constrained least squares problem
Least-norm problem

- special case of constrained least squares problem, with $A = I$, $b = 0$
- least-norm problem:

\[
\begin{align*}
\text{minimize} & \quad \|x\|^2 \\
\text{subject to} & \quad Cx = d
\end{align*}
\]

\text{i.e., find the smallest vector that satisfies a set of linear equations}
Froce sequence

- unit mass on frictionless surface, initially at rest
- 10-vector $f$ gives forces applied for one second each
- final velocity and position are

$$v^{\text{fin}} = f_1 + f_2 + \cdots + f_{10}$$
$$p^{\text{fin}} = (19/2)f_1 + (17/2)f_2 + \cdots + (1/2)f_{10}$$

- let’s find $f$ for which $v^{\text{fin}} = 0$, $p^{\text{fin}} = 1$
- $f^{\text{bb}} = (1, -1, 0, \ldots, 0)$ works (called ‘bang-bang’)

Least-norm problem
Bang-bang force sequence

Least-norm problem
Least-norm force sequence

- let’s find least-norm $f$ that satisfies $p^{\text{fin}} = 1, v^{\text{fin}} = 0$
- least-norm problem:

$$\begin{align*}
\text{minimize} & \quad \|f\|^2 \\
\text{subject to} & \quad \begin{bmatrix}
1 & 1 & \cdots & 1 & 1 \\
19/2 & 17/2 & \cdots & 3/2 & 1/2
\end{bmatrix} f = \begin{bmatrix}
0 \\
1
\end{bmatrix}
\end{align*}$$

with variable $f$

- solution $f^{\ln}$ satisfies $\|f^{\ln}\|^2 = 0.0121$ (compare to $\|f^{\text{bb}}\|^2 = 2$)
Least-norm force sequence

Least-norm problem
Outline

Linearly constrained least squares

Least-norm problem

Solving the constrained least squares problem
Optimality conditions via calculus

to solve constrained optimization problem

\[
\begin{align*}
\text{minimize} & \quad f(x) = \|Ax - b\|^2 \\
\text{subject to} & \quad c_i^T x = d_i, \quad i = 1, \ldots, p
\end{align*}
\]

1. form *Lagrangian* function, with *Lagrange multipliers* \( z_1, \ldots, z_p \)

\[
L(x, z) = f(x) + z_1 (c_1^T x - d_1) + \cdots + z_p (c_p^T x - d_p)
\]

2. optimality conditions are

\[
\frac{\partial L}{\partial x_i}(\hat{x}, z) = 0, \quad i = 1, \ldots, n, \quad \frac{\partial L}{\partial z_i}(\hat{x}, z) = 0, \quad i = 1, \ldots, p
\]
Optimality conditions via calculus

\[ \frac{\partial L}{\partial z_i}(\hat{x}, z) = c_i^T \hat{x} - d_i = 0, \] which we already knew

\[ \text{first } n \text{ equations are more interesting:} \]

\[ \frac{\partial L}{\partial x_i}(\hat{x}, z) = 2 \sum_{j=1}^{n} (A^T A)_{ij} \hat{x}_j - 2(A^T b)_i + \sum_{j=1}^{p} z_j c_i = 0 \]

\[ \text{in matrix-vector form: } 2(A^T A)\hat{x} - 2A^T b + C^T z = 0 \]

\[ \text{put together with } C\hat{x} = d \text{ to get KKT conditions} \]

\[ \begin{bmatrix} 2A^T A & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ z \end{bmatrix} = \begin{bmatrix} 2A^T b \\ d \end{bmatrix} \]

a square set of \( n + p \) linear equations in variables \( \hat{x}, z \)

\[ \text{KKT equations are extension of normal equations to CLS} \]
Solution of constrained least squares problem

- assuming the KKT matrix is invertible, we have

\[
\begin{bmatrix}
\hat{x} \\
z
\end{bmatrix} = \begin{bmatrix}
2A^T A & C^T \\
C & 0
\end{bmatrix}^{-1} \begin{bmatrix}
2A^T b \\
d
\end{bmatrix}
\]

- KKT matrix is invertible if and only if

\[C \text{ has independent rows, and } \begin{bmatrix} A \\ C \end{bmatrix} \text{ has independent columns}\]

- implies \(m + p \geq n\), \(p \leq n\)

- can compute \(\hat{x}\) in \(2(n + p)^3\) flops; order is \(n^3\) flops
Direct verification of solution

- to show that \( \hat{x} \) is solution, suppose \( x \) satisfies \( Cx = d \)

- then

\[
\|Ax - b\|^2 = \|(Ax - A\hat{x}) + (A\hat{x} - b)\|^2 \\
= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 + 2(Ax - A\hat{x})^T(A\hat{x} - b)
\]

- expand last term, using \( 2A^T(A\hat{x} - b) = -C^Tz, Cx = C\hat{x} = d \):

\[
2(Ax - A\hat{x})^T(A\hat{x} - b) = 2(x - \hat{x})^T A^T (A\hat{x} - b) \\
= -(x - \hat{x})^T C^T z \\
= -(C(x - \hat{x}))^T z \\
= 0
\]

- so \( \|Ax - b\|^2 = \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 \geq \|A\hat{x} - b\|^2 \)

- and we conclude \( \hat{x} \) is solution
Solution of least-norm problem

- least-norm problem: minimize $\|x\|^2$ subject to $Cx = d$
- matrix $\begin{bmatrix} I & C \end{bmatrix}$ always has independent columns
- we assume that $C$ has independent rows
- optimality condition reduces to
  $$\begin{bmatrix} 2I & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ d \end{bmatrix}$$
- so $\hat{x} = -(1/2)C^Tz$; second equation is then $-(1/2)CC^Tz = d$
- plug $z = -2(CC^T)^{-1}d$ into first equation to get
  $$\hat{x} = C^T(CC^T)^{-1}d = C^\dagger d$$

where $C^\dagger$ is (our old friend) the pseudo-inverse
so when \( C \) has independent rows:

- \( C^\dagger \) is a right inverse of \( C \)
- so for any \( d \), \( \hat{x} = C^\dagger d \) satisfies \( C\hat{x} = d \)
- and we now know: \( \hat{x} \) is the \textit{smallest} solution of \( Cx = d \)