

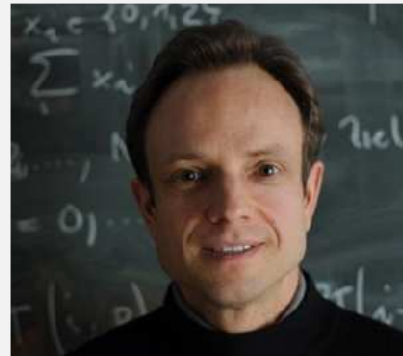
Course Video

Section 3.6: Constrained Least Squares

Prof. R. Hiptmair, SAM, ETH Zurich

Date: October 1, 2020

(C) Seminar für Angewandte Mathematik, ETH Zürich



Prerequisites.

- Gradient of a function $\mathbb{R}^n \rightarrow \mathbb{R}$

Dependencies. Section 3.1, Section 3.4.1

Duration: 45 minutes

III. Direct Methods for Linear Least Squares Problems

3.6. Constrained Least Squares

$m \times n$ OD-LS E, but $p < n$ eqns. have to be satisfied exactly.

$$\begin{matrix} m \\ p \end{matrix} \begin{matrix} \updownarrow \\ \updownarrow \end{matrix} \begin{bmatrix} A \\ C \end{bmatrix} x = \begin{bmatrix} b \\ d \end{bmatrix}, \quad m \geq n$$

Linear least squares problem with linear constraint:

Given: $A \in \mathbb{R}^{m,n}$, $m \geq n$, $\text{rank}(A) = n$, $b \in \mathbb{R}^m$,
 $C \in \mathbb{R}^{p,n}$, $p < n$, $\text{rank}(C) = p$, $d \in \mathbb{R}^p$

Find: $x \in \mathbb{R}^n$ such that $\|Ax - b\|_2 \rightarrow \min$ and $Cx = d$.

(3.6.0.1)

Linear constraint

2

3.6.1.: Solution via Lagrange Multipliers

→ essential tool in
constrained optimization



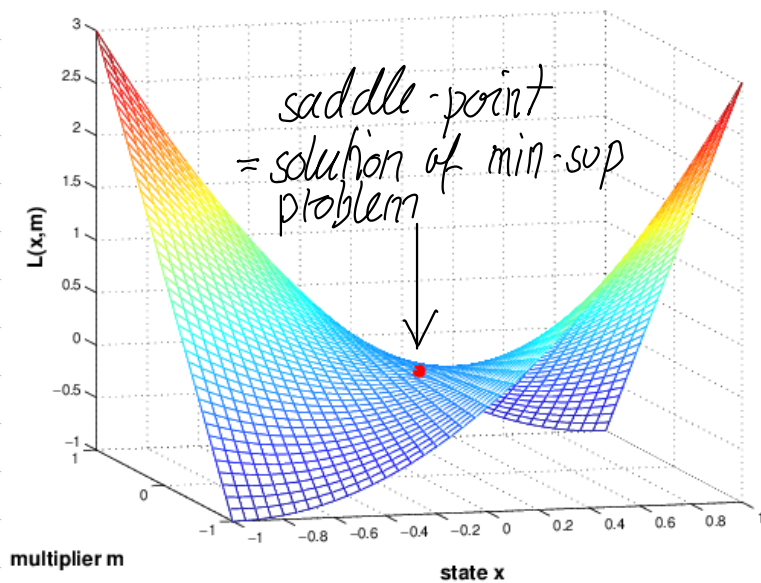
Idea: coupling the constraint using the **Lagrange multiplier** $\mathbf{m} \in \mathbb{R}^p$

$$\mathbf{x} = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \sup_{\mathbf{m} \in \mathbb{R}^p} L(\mathbf{y}, \mathbf{m}), \quad (3.6.1.2)$$

$$L(\mathbf{y}, \mathbf{m}) := \frac{1}{2} \|\mathbf{A}\mathbf{y} - \mathbf{b}\|_2^2 + \mathbf{m}^\top (\mathbf{C}\mathbf{y} - \mathbf{d}). \quad (3.6.1.3)$$

Lagrange functional = ∞ , unless $\mathbf{C}\mathbf{y} - \mathbf{d} = 0$
Constraint satisfied by any minimizer

(3.6.1.2) is called a **saddle-point problem**



$$L(x, m) = x^2 - 2xm$$

[$n = p = 1$]

△ Graph of L is flat*
in the saddle point

* all partial derivatives vanish
in a saddle point

If $(\mathbf{x}, \mathbf{q})$ is a saddle point of L
 $\Rightarrow \frac{\partial L}{\partial \mathbf{y}}(\mathbf{x}, \mathbf{q}) = \frac{\partial L}{\partial \mathbf{m}}(\mathbf{x}, \mathbf{q}) = 0$

Computing partial derivatives

$$L(\mathbf{y}, \mathbf{m}) = \frac{1}{2} \sum_{i=1}^n \left(\sum_{j=1}^n (A)_{ji} y_j - b_i \right)^2 + \sum_{k=1}^p m_k \left(\sum_{i=1}^n (C)_{ki} y_i - d_k \right)$$

$$\mathbf{y} = [y_1, \dots, y_n]^\top, \quad \mathbf{m} = [m_1, \dots, m_p]^\top$$

A quadratic polynomial in y_i, m_k

Eqs.: $\frac{\partial L}{\partial y_1}(\mathbf{x}, \mathbf{q}) = \dots = \frac{\partial L}{\partial y_n}(\mathbf{x}, \mathbf{q}) = \frac{\partial L}{\partial m_1}(\mathbf{x}, \mathbf{q}) = \dots = \frac{\partial L}{\partial m_p}(\mathbf{x}, \mathbf{q}) = 0$

$$\frac{\partial L}{\partial \mathbf{y}}(\mathbf{x}, \mathbf{q}) = \mathbf{A}^\top (\mathbf{A}\mathbf{x} - \mathbf{b}) + \mathbf{C}^\top \mathbf{q} \stackrel{!}{=} 0, \quad (3.6.1.5a)$$

$$\frac{\partial L}{\partial \mathbf{m}}(\mathbf{x}, \mathbf{q}) = \mathbf{C}\mathbf{x} - \mathbf{d} \stackrel{!}{=} 0. \quad (3.6.1.5b)$$



$(n+p) \times (n+p)$ LSE

$$\begin{bmatrix} \mathbf{A}^\top \mathbf{A} & \mathbf{C}^\top \\ \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{q} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^\top \mathbf{b} \\ \mathbf{d} \end{bmatrix} \quad \text{Augmented normal equations} \\ \text{(matrix saddle point problem)} \quad (3.6.1.6)$$

↑
symmetric with zero right-lower block

③ Unique solution, if

- $\text{rank}(A) = n$
- $\text{rank}(C) = p$

Algorithm based on block-LU decomposition

$$\begin{bmatrix} A^T A & C^T \\ C & 0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ Y & I \end{bmatrix} \begin{bmatrix} U_1 & X \\ 0 & U_2 \end{bmatrix}$$

- $U_1 = A^T A$

- $X = C^T$

- $Y U_1 = C \Leftrightarrow U_1^T Y^T = C^T$ [p nx n-LSE]

- $Y X + U_2 = 0$

Rem: If A large, sparse $\Rightarrow$ do not compute $A^T A$, but solve extended equation with auxiliary unknown $\underline{r} := Ax - b$ (residual)

$$\begin{bmatrix} -I & A & 0 \\ A^T & 0 & C^T \\ 0 & C & 0 \end{bmatrix} \begin{bmatrix} \underline{r} \\ \underline{x} \\ \underline{m} \end{bmatrix} = \begin{bmatrix} \underline{b} \\ 0 \\ \underline{d} \end{bmatrix} \triangleq \text{Extended augmented normal equations} \quad (3.6.1.8)$$

3.6.2. Solution via SVD

Constraint: $C \underline{x} = \underline{d}$ with $C \in \mathbb{R}^{p,n}$ ($p < n$), of full rank, $\text{rank}(C) = p$

$$\underline{x} \in \underline{x}_0 + \mathcal{N}(C)$$

$\uparrow$ particular solution: $C \underline{x}_0 = \underline{d}$
 $\uparrow$ accessible via SVD

Lemma 3.4.1.13. SVD and rank of a matrix $\rightarrow$

If, for some $1 \leq r \leq p := \min\{m, n\}$, the singular values of $A \in \mathbb{K}^{m,n}$ satisfy $\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_p = 0$, then

$$A = U \Sigma V:$$

- $\text{rank}(A) = r$ (no. of non-zero singular values),
- $\mathcal{N}(A) = \text{Span}\{(V)_{:,r+1}, \dots, (V)_{:,n}\}$,
- $\mathcal{R}(A) = \text{Span}\{(U)_{:,1}, \dots, (U)_{:,r}\}$.

Full SVD: $C = U \begin{bmatrix} \Sigma_p & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$

$U \in \mathbb{R}^{p,p}$ orth.
 $\Sigma \in \mathbb{R}^{p,p}$ diag.
 $V \in \mathbb{R}^{n,n}$ orth.

$\mathcal{N}(C) = \mathcal{R}(V_2)$ $\quad \quad \quad \uparrow = V$

$$\underline{x} = \underline{x}_0 + V_2 \underline{y}, \quad \underline{y} \in \mathbb{R}^{n-p}$$

$$\|A \underline{x} - \underline{b}\|_2 = \|A(\underline{x}_0 + V_2 \underline{y}) - \underline{b}\|_2 = \underbrace{\|A V_2 \underline{y} - (\underline{b} - A \underline{x}_0)\|_2}_{m \times (n-p)} \rightarrow \min$$

④ $\hat{=}$ Std. linear LSQ in $y \in \mathbb{R}^{n-p}$

Rem: Finding $x_0 \in \mathbb{R}^n$

Idea: $x_0 \in \mathcal{N}(C)^\perp = \mathcal{R}(V_1)$

$$x_0 = V_1 \underline{z}, \quad \underline{z} \in \mathbb{R}^p$$

$$[Cx = d] \quad C V_1 \underline{z} = \underline{d}$$

$$\Leftrightarrow M [\Sigma_p \ 0] \underbrace{\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}}_{= \begin{bmatrix} I_p \\ 0 \end{bmatrix}} V_1 \underline{z} = \underline{d}$$

$\Sigma_p \in \mathbb{R}^{p,p}$ is regular, since
 $\text{rank}(C) = p$

$$\Leftrightarrow M \Sigma_p \underline{z} = \underline{d}$$

$$\Leftrightarrow \underline{z} = \Sigma_p^{-1} M^T \underline{d}$$

↑ just scaling

$$\underline{x}_0 = V_1 \Sigma_p^{-1} M^T \underline{d}$$

Review questions

Study the review questions 3.6.2.1

5

7