Positive definite matrices, Schur complements

ROB 501

Necmiye Ozay

- Positive definite matrices
  - Weighted least squares
- Schur complements
- Intro to Recursive Least Squares

↳ Def: A real, symmetric matrix P is positive-definite if $\forall x \in \mathbb{R}^n$, $x \neq 0$, $\boxed{x^T P x > 0}$

(we denote p.d. matrices by $P > 0$ or $P \succ 0$)

Thm: A symmetric matrix P is pos. def. iff all its e-values are positive.

Proof left as an exercise.

Def: N is a square root of a symmetric matrix P iff $N^T N = P$.

Assume $P \in \mathbb{R}^{n \times n}$

Thm: • $\underbrace{P \geq 0}_{\text{pos. semi-def.}} \iff \exists N$ s.t. $N^T N = P$.

• $\underbrace{P > 0}_{\text{pos. def.}} \iff \exists N$ $\mathrm{rank}(N) = n$ s.t. $N^T N = P$.

Note that if $P > 0$, * $f(x) = \sqrt{x^T P x}$ defines a norm.

* $g(x, y) = x^T P y$ defines an inner product.

Using the fact that $x^T P y$ is an inner product, we can generalize least squares to the weighted case.

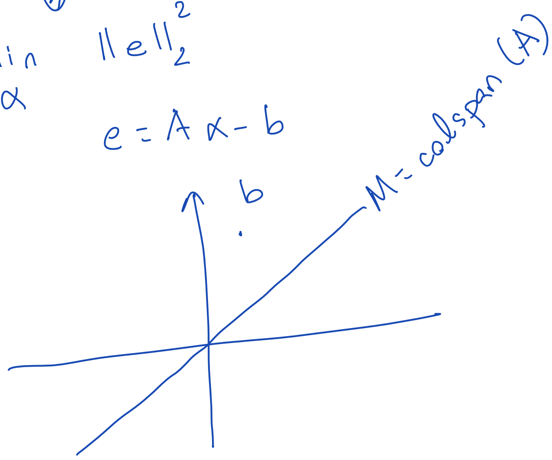
Standard LS

$$\min_{\alpha} \; \underbrace{\|A\alpha - b\|_2^2}_{e}$$

$$\Updownarrow$$

$$\min_{\alpha} \; \|e\|_2^2$$

$$e = A\alpha - b$$



Weighted LS (WLS)

Given $Q > 0$

$$\min_{\alpha} \; e^T Q e$$

$$e = A\alpha - b$$

$$\Updownarrow$$

$$\min_{\alpha} \; (A\alpha - b)^T Q (A\alpha - b)$$

$$\Updownarrow \quad \text{note} \quad Q = N^T N$$

$$\min_{\alpha} \; \|Ne\|_2^2$$

$$e = A\alpha - b$$

WLS:

(assume $\text{rank}(A) = m$)

$$A = [A_1 | \dots | A_m]$$

$$\hat{\alpha} = \text{argmin} \; (A\alpha - b)^T Q (A\alpha - b)$$

$$[G^T]_{ij} = \langle A_i, A_j \rangle = A_i^T Q A_j$$

$$\beta_i = \langle b, A_i \rangle = b^T Q A_i$$

$G^T \hat{\alpha} = \beta$ where G and β are as defined above.

Rob 501 Handout: Grizzle
Weighted Least Squares

Let M be an $n \times n$ positive definite matrix ($M \succ 0$) We revisit the over determined system of equations,

$$A\alpha = b,$$

where $A = n \times m, n \geq m, \text{rank}(A) = m, \alpha \in \mathbb{R}^m$, and $b \in \mathbb{R}^n$.

We seek $\hat{\alpha}$ such that

$$\|A\hat{\alpha} - b\| = \min_{\alpha \in \mathbb{R}^m} \|A\alpha - b\|$$

where $\|x\| := \left(x^\top M x\right)^{1/2}$ and $M > 0$.

Solution: Define an appropriate inner product space $\mathcal{X} = \mathbb{R}^n, \mathcal{F} = \mathbb{R}, \langle x, y \rangle := x^\top M y$ and decompose A into its columns

$$A = \left[\ A_1 \,\middle|\, A_2 \,\middle|\, \cdots \,\middle|\, A_m\ \right]$$

We seek

$$\hat{x} := \underset{x \in \text{span}\{A_1, \ldots, A_m\}}{\text{argmin}} ||x - b||^2$$

Normal Equations:

$$\hat{x} = \hat{\alpha}_1 A_1 + \hat{\alpha}_2 A_2 + \cdots + \hat{\alpha}_m A_m$$

$$G^\top \hat{\alpha} = \beta, \text{ with } G = G^\top$$

$A_i^T Q A_j = \quad \underline{[G^\top]_{ij}} = [G]_{ij} = \langle A_i, A_j \rangle = A_i^\top M A_j = [A^\top M A]_{ij}$

$b^T Q A_i = \quad \beta_i = \langle b, A_i \rangle = b^\top M A_i = A_i^\top M b = [A^\top M b]_i.$

$//$ $\longrightarrow$ since $b^T Q A_i$ is a scalar, it is equal to its transpose.

$(b^T Q A_i)^T$ $\quad \beta = A^T Q b$

$A_i^T Q b$

$G^T \hat{\alpha} = \beta$

$A^T Q A \hat{\alpha} = A^T Q b$

$\hat{\alpha} = (A^T Q A)^{-1} A^T Q b$

$$\boxed{\therefore \quad A^\top M A \hat{\alpha} = A^\top M b}.$$

$y = C \underline{x} + \eta$

stochastic noise model with mean & covariance → weight matrix will be related to cov. matrix.

Because $\operatorname{rank}(A) = m$, its columns are linearly independent and thus the Gram matrix is invertible. Hence, we conclude that

$$\boxed{\hat{\alpha} = (A^\top M A)^{-1} A^\top M b}.$$

Schur Complement Thm.

- Means to check whether a matrix is pos. def by studying smaller submatrices
- It shows up when we study conditional Gaussian random variables.

Thm: Suppose $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{m \times m}$, and

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

symmetric, (i.e., $A = A^T$, $C = C^T$)

Then, TFAE:

a) $M > 0$ (pos. def.)

b) $A > 0$ and $\underbrace{C - B^T A^{-1} B}_{\text{Schur complement of } A \text{ in } M} > 0$

c) $C > 0$ and $\underbrace{A - B C^{-1} B^T}_{\text{Schur complement of } C \text{ in } M} > 0$

<u>Corollary:</u> $A > 0$, $C > 0$ are necessary (but not sufficient) conditions for $M > 0$.

<u>Proof:</u> Will prove $(a \Leftrightarrow b)$ [$a \Leftrightarrow c$ is identical]

$a) \Rightarrow b)$

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} > 0 \quad \text{that is} \quad \forall \begin{bmatrix} x \\ y \end{bmatrix} \neq 0$$

$$(*) \quad \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} > 0$$

<u>Claim 1:</u> $(A > 0)$. Let $x \neq 0$ and arbitrary

and $y = 0$.

$$0 < \begin{bmatrix} x \\ 0 \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} x \\ 0 \end{bmatrix} = x^T A x$$

$$\Rightarrow x^T A x > 0 \quad \forall x \neq 0.$$

$$\therefore A > 0.$$

<u>Claim 2:</u> $(C - B^T A^{-1} B > 0)$. Let $y \neq 0$ and arbitrary and make a smart choice of x: $\underline{Ax + By} = 0 \Rightarrow x = -A^{-1}By$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -A^{-1}By \\ y \end{bmatrix} \neq 0 \quad (\text{because } y \neq 0)$$

$$0 < \begin{bmatrix} -A^{-1}By \\ y \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} -A^{-1}By \\ y \end{bmatrix}$$

$$= \begin{bmatrix} (-A^{-1}By)^T & y^T \end{bmatrix} \begin{bmatrix} 0 \\ -B^T A^{-1} By + Cy \end{bmatrix}$$

$$= -y^T B^T A^{-1} B y + y^T C y > 0$$

$$y^T (C - B^T A^{-1} B) y > 0$$

since y is arbitrary $\Rightarrow C - B^T A^{-1} B > 0$.

b) $\Rightarrow$ a) $A \succ 0$ and $C - B^T A^{-1} B \succ 0$

$\Longrightarrow M \succ 0$

Claim 1: Define $\tilde{x} = x + A^{-1} B y$

Then, $\begin{bmatrix} \tilde{x} \\ y \end{bmatrix} \neq 0 \iff \begin{bmatrix} x \\ y \end{bmatrix} \neq 0$

Why? If $y \neq 0$, both vectors are trivially non-zero.

If $y = 0$, then $\tilde{x} = x$, then $\begin{bmatrix} \tilde{x} \\ 0 \end{bmatrix} \neq 0 \iff \begin{bmatrix} x \\ 0 \end{bmatrix} \neq 0$

Claim 2: Take arbitrary $\begin{bmatrix} x \\ y \end{bmatrix} \neq 0$

$$\begin{bmatrix} x \\ y \end{bmatrix}^T \underbrace{\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}}_{M} \begin{bmatrix} x \\ y \end{bmatrix} = [x^T \; y^T] \begin{bmatrix} Ax + By \\ B^T x + Cy \end{bmatrix}$$

$$= \overset{①}{x^T A x} + \overset{②}{x^T B y} + y^T B^T x + y^T C y$$

use $x = \tilde{x} - A^{-1} B y$

$$= \tilde{x}^T A \tilde{x} + y^T (C - B^T A^{-1} B) y$$

$(\tilde{x} - A^{-1} B y)^T A (\tilde{x} - A^{-1} B y)$

① $= \tilde{x}^T A \tilde{x} - y^T B^T \tilde{x} - \cancel{\tilde{x}^T B y} \cancel{+ y^T B^T A^{-1} B y}$

$(\tilde{x} - A^{-1} B y)^T B y$

② $\cancel{\tilde{x}^T B y} - \cancel{y^T B^T A^{-1} B y}$

> 0 (by $A \succ 0$, $C - B^T A^{-1} B \succ 0$ and claim 1 and $\begin{bmatrix} x \\ y \end{bmatrix} \neq 0$)

$\therefore M$ is p.d.

Ex 1:

$$M = \begin{bmatrix} a & b \\ b & c \end{bmatrix}_{2\times 2} > 0 \iff a>0 \text{ and } c - ba^{-1}b > 0$$

$$\iff a>0 \text{ and } ac - b^2 > 0$$

$$\det(M) > 0$$

$\begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix} \overset{?}{>} 0$ yes! p.d.

$\begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} \overset{?}{>} 0$ $2>0$ but $\det = -5$

no! not p.d.

Ex 2:

$$M = \left[\begin{array}{c|cc} 2 & 1 & 1 \\ \hline 1 & 2 & 1 \\ 1 & 1 & 3 \end{array}\right]$$

$A = 2$ $B = [1 \ \ 1]$

$$C = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

$M \overset{?}{>} 0$ • $\overset{=2}{A} > 0$ ✓

and • $C - B^T A^{-1} B = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1}{2} [1 \ \ 1]$

$$= \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 2.5 \end{bmatrix} > 0 \ ✓$$

$1.5 > 0$ $\det = \frac{14}{4} > 0$ $\implies M > 0$

Rob 501 Handout: Grizzle
Recursive Least Squares

Model:

$$y_i = C_i x + e_i, \;\; i = 1, 2, 3, \cdots$$

$$\begin{aligned} C_i &\in \mathbb{R}^{m \times n} \\ i &= \text{time index} \\ x &= \text{an unknown } \underline{\text{constant}} \text{ vector} \in \mathbb{R}^n \\ y_i &= \text{measurements} \in \mathbb{R}^m \\ e_i &= \text{model "mismatch"} \in \mathbb{R}^m \end{aligned}$$

Objective 1: Compute a least squared error estimate of x at time k, using all available data at time k, $(y_1, \cdots, y_k)$!

Objective 2: Discover a computationally attractive form for the answer.

Solution:

$$\begin{aligned} \hat{x}_k := & \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left(\sum_{i=1}^{k} (y_i - C_i x)^\top S_i (y_i - C_i x) \right) \\ = & \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left(\sum_{i=1}^{k} e_i^\top S_i e_i \right) \end{aligned}$$

where $S_i = m \times m$ positive definite matrix. ($S_i > 0$ for all time index i)

Batch Solution:

$$Y_k = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}, A_k = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_k \end{bmatrix}, E_k = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_k \end{bmatrix}$$

$$Y_k \doteq \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix} \qquad A_k \doteq \begin{bmatrix} C_1 \\ \vdots \\ C_k \end{bmatrix}$$

$$\hat{x}_k = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \|Y_k - A_k x\|^2_{R_k}$$

$$R_k = \begin{bmatrix} S_1 & & & \mathbf{0} \\ & S_2 & & \\ & & \ddots & \\ \mathbf{0} & & & S_k \end{bmatrix} = diag(S_1, S_2, \cdots, S_k) > 0$$

$Y_k = A_k x + E_k$, [model for $1 \le i \le k$]
$\|Y_k - A_k x\|^2 = \|E_k\|^2 := E_k^\top R_k E_k$

Since $\hat{x}_k$ is the value minimizing the error $\|E_k\|$, which is the unexplained part of the model,

$$\hat{x}_k = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \|E_k\| = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \|Y_k - A_k x\|,$$

which satisfies the Normal Equations $(A_k^\top R_k A_k)\hat{x}_k = A_k^\top R_k Y_k$.

$\therefore \boxed{\hat{x}_k = (A_k^\top R_k A_k)^{-1} A_k^\top R_k Y_k}$ which is called a <u>Batch Solution</u>.

Drawback: $A_k = km \times n$ matrix, and grows at each step!

Solution: Find a recursive means to compute $\hat{x}_{k+1}$ in terms of $\hat{x}_k$ and the new measurement y_{k+1}!

Normal equations at time k, $(A_k^\top R_k A_k)\hat{x}_k = A_k^\top R_k Y_k$, is equivalent to

$$\left(\sum_{i=1}^{k} C_i^\top S_i C_i \right) \hat{x}_k = \sum_{i=1}^{k} C_i^\top S_i y_i.$$

We define

$$M_k \hat{x}_k = \sum_{i=1}^{k} C_i^\top S_i y_i$$

$$M_k = \sum_{i=1}^{k} C_i^\top S_i C_i$$

so that

$$M_{k+1} = M_k + C_{k+1}^\top S_{k+1} C_{k+1}.$$

$$\left(M_k + C_{k+1}^\top S_{k+1} C_{k+1}\right) \hat{x}_{k+1} = M_k \hat{x}_{k+1} + C_{k+1}^\top S_{k+1} C_{k+1} \hat{x}_{k+1}$$

At time $k+1$,

$$\underbrace{\left(\sum_{i=1}^{k+1} C_i^\top S_i C_i\right)}_{M_{k+1}} \hat{x}_{k+1} = \sum_{i=1}^{k+1} C_i^\top S_i y_i$$

or

$$M_{k+1}\hat{x}_{k+1} = \underbrace{\sum_{i=1}^{k} C_i^\top S_i y_i}_{M_k \hat{x}_k} + C_{k+1}^\top S_{k+1} y_{k+1}.$$

$\therefore M_{k+1}\hat{x}_{k+1} = M_k \hat{x}_k + C_{k+1}^\top S_{k+1} y_{k+1}$

Good start on recursion! Estimate at time $k+1$ expressed as a linear combination of the estimate at time k and the latest measurement at time $k+1$.

Continuing,

$$\hat{x}_{k+1} = M_{k+1}^{-1}\left[M_k \hat{x}_k + C_{k+1}^\top S_{k+1} y_{k+1}\right].$$

Because

$$M_k = M_{k+1} - C_{k+1}^\top S_{k+1} C_{k+1},$$

we have

$$\hat{x}_{k+1} = \hat{x}_k + \underbrace{M_{k+1}^{-1} C_{k+1}^\top S_{k+1}}_{\text{Kalman gain}} \underbrace{\left(y_{k+1} - C_{k+1}\hat{x}_k\right)}_{\text{Innovations}}.$$

Innovations $y_{k+1} - C_{k+1}\hat{x}_k$ = measurement at time $k+1$ minus the "predicted" value of the measurement = "new information".

In a real-time implementation, computing the inverse of M_{k+1} can be time consuming. An attractive alternative can be obtained by applying the Matrix Inversion Lemma:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B\left(DA^{-1}B + C^{-1}\right)^{-1} DA^{-1}$$

Now, following the substitution rule as shown below,

$$A \leftrightarrow M_k \quad B \leftrightarrow C_{k+1}^\top \quad C \leftrightarrow S_{k+1} \quad D \leftrightarrow C_{k+1},$$

we can obtain that

$$
\begin{aligned}
M_{k+1}^{-1} &= \left(M_k + C_k^\top S_{k+1} C_{k+1}\right)^{-1} \\
&= M_k^{-1} - M_k^{-1} C_{k+1}^\top \left[C_{k+1} M_k^{-1} C_{k+1}^\top + S_{k+1}^{-1}\right]^{-1} C_{k+1} M_k^{-1},
\end{aligned}
$$

which is a recursion for M_k^{-1}!

Upon defining

$$P_k = M_k^{-1},$$

we have

$$P_{k+1} = P_k - P_k C_{k+1}^\top \left[C_{k+1} P_k C_{k+1}^\top + S_{k+1}^{-1}\right]^{-1} C_{k+1} P_k$$

We note that we are now inverting a matrix that is $m \times m$, instead of one that is $n \times n$. Typically, $n > m$, sometimes by a lot!

$$\underbrace{\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}}_{A_1}, \underbrace{\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}}_{A_2}$$

$A_1^\top v_3 = 0$

$A_2^\top v_3 = 0$

$\begin{bmatrix} A_1^\top \\ A_2^\top \end{bmatrix} v_3 = 0$

$M = \begin{bmatrix} A_1^\top \\ A_2^\top \end{bmatrix}$

$M v_3 = 0$

$A_{1,2} \in \mathbb{R}^3$

$\dim\{\operatorname{span}(A_1, A_2)\} = 2$

$\dim(\mathbb{R}^3) = 3$

$\exists\, v_3$ s.t.

$v_3 \perp \operatorname{span}\{A_1, A_2\}$

and $\{A_1, A_2, v_3\}$

forms a basis for $\mathbb{R}^3$

$$v_1 = \begin{bmatrix} 1 \\ -5 \\ 1 \end{bmatrix}$$

$A^2 = A \cdot A$

sqrt. of A

$$u = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$A = \underline{N^T N}$

$A = N^T N$

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$\cancel{B}$

$A = M \Sigma M^{-1}$

$A = A^T \qquad M^{-1} = M^T$

$A = O \underset{\sim}{\Sigma} O^T$

$\Rightarrow \qquad = O (\Sigma)^{1/2} \underline{(\Sigma)^{1/2}} O^T$

$$A = N^T \underline{N}$$

$$\tilde{N} = \underline{\underline{QN}}$$ where Q is an orthogonal matrix

$$\tilde{N}^T = N^T Q^T = N^T Q^{-1}$$

$$\tilde{N}^T \tilde{N} = N^T \cancel{Q^{-1}} \cancel{Q} N = A$$

---

if x is a sol.

$$Ax = b$$

and $\hat{x} \in N(A)$, then

$\tilde{x} = x + \hat{x}$ is also a sol'n to $Ax = b$.

Consider any $\tilde{x}$.

and consider $(\mathcal{N}(A))^\perp$

(which depends on the inner product)

$$\tilde{x} = \underbrace{x_{[\mathcal{N}(A)]^\perp}} + x_{\mathcal{N}(A)}$$

$$\|\tilde{x}\|_Q^2 = \|x_{[\mathcal{N}(A)]^\perp} + x_{\mathcal{N}(A)}\|_Q^2$$

$$= \|\underbrace{x_{[\mathcal{N}(A)]^\perp}}\|_Q^2 + \|x_{\mathcal{N}(A)}\|_Q^2$$

$$\geq 0 \qquad\qquad \geq 0$$

matrix of basis vectors for $[\mathcal{N}(A)]^\perp = B$

$[b_1 \ldots b_k]$

$$\tilde{x} = B\alpha$$

$$A\tilde{x} = b$$

$$AB\alpha = b$$