Bayesian Scientific Computing
Day 3

Daniela Calvetti, Erkki Somersalo

Case Western Reserve University
Department of Mathematics, Applied Mathematics and Statistics

Lyngby, December 2019

Freeing the Boundaries

In the smoothness priors discussed so far, we assumed that the boundary values vanished. Often, this is not in line with what we believe about the object.

Consider the 1D discretized problem.

$$x_j = u(t_j), \quad t_j = jh, \quad 0 \leq j \leq n,$$

and $h = 1/n$. Use the second order model (interior points),

$$x_{j-1} - 2x_j + x_{j+1} = h^2 \gamma w_j, \quad 1 \leq j \leq n-1.$$

$$x_0 = ?, \quad x_n = ?$$

Freeing the Boundaries

Bayesian philosophy: *If you don't know something, model it as a random variable.*

$$
\begin{aligned}
X_0 &= \alpha W_0, \\
X_0 - 2X_1 + X_2 &= h^2 \gamma W_1, \\
X_1 - 2X_2 + X_3 &= h^2 \gamma W_2, \\
\vdots &= \vdots \\
X_{n-2} - 2X_{n-1} + X_n &= h^2 \gamma W_{n-1}, \\
X_n &= \alpha W_n.
\end{aligned}
$$

Freeing the Boundaries

In matrix form,

$$\mathsf{L}_\alpha X = h^2\gamma W, \quad W \sim \mathcal{N}(0, \mathsf{I}_{n+1}),$$

where

$$\mathsf{L}_\alpha = \begin{bmatrix} h^2\gamma/\alpha & & & & & \\ 1 & -2 & 1 & & & \\ & 1 & -2 & & & \\ & & \ddots & & & \\ & & & 1 & -2 & 1 \\ & & & & & h^2\gamma/\alpha \end{bmatrix}.$$

Freeing the Boundaries

How do we choose α?

1. If you **believe** that x_0 and x_n has a certain range around zero, use the information to set α.
2. If you want x_0 and x_n have more or leas the same variance as the other degrees of freedom, we can use a sequential adjustment as explained below.

Freeing the Boundaries

Sequential adjustment:

1. Start with the zero Dirichlet data, assuming $x_0 = x_n = 0$. Corresponding prior:

$$\mathsf{L}_D \widetilde{X} = h^2 \gamma \widetilde{W},$$

where

$$\widetilde{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_{n-1} \end{bmatrix}, \quad \mathsf{L}_D = \begin{bmatrix} -2 & 1 & & \\ 1 & -2 & 1 & \\ & & \ddots & \\ & & 1 & -2 \end{bmatrix}.$$

Freeing the Boundaries

2. Calculate the variances of the single degrees of freedom:

$$\sigma_j^2 = \mathrm{var}(X_j) = \mathsf{E}\{X_j^2\} = e_j^{\mathsf{T}} E\{XX^{\mathsf{T}}\} e_j,$$

   where e_j is the canonical jth basis vector.

3. Adjust α so that

$$\mathrm{var}(X_0) = \alpha^2 = \max\{\sigma_j^2\}.$$

Freeing the Boundaries

To compute σ_j^2: We have

$$\widetilde{X} = h^2\gamma \mathsf{L}_D^{-1}\widetilde{W},$$

so

$$\begin{aligned} \mathrm{cov}(\widetilde{X}) &= \mathsf{E}\{\widetilde{X}\widetilde{X}^\mathsf{T}\} = h^4\gamma^2\mathsf{L}_D^{-1}\mathsf{E}\{\widetilde{W}\widetilde{W}^\mathsf{T}\}\mathsf{L}_D^{-\mathsf{T}} \\ &= h^4\gamma^2\mathsf{L}_D^{-1}\mathsf{L}_D^{-\mathsf{T}}, \end{aligned}$$

so

$$\sigma_j^2 = h^4\gamma^2 e_j^\mathsf{T}\mathsf{L}_D^{-1}\mathsf{L}_D^{-\mathsf{T}}e_j = h^4\gamma^2\|\mathsf{L}_D^{-\mathsf{T}}e_j\|^2.$$

The maximal variance is reached at the midpoint of the interval, so (assuming that n is even, we choose

$$\alpha^2 = h^4\gamma^2\|\mathsf{L}_D^{-\mathsf{T}}e_{n/2}\|^2.$$

It is left as an exercise to test the idea.

Linear Gaussian inverse Problems

In this section, we consider linear inverse problems with additive noise,

$$b = \mathsf{A}x + \varepsilon,$$

assuming that all underlying probability densities are Gaussian.

- Usually, the noise ε is assumed to be independent of x.
- This assumption is not necessary, and it is not always valid:
  - Modeling error
  - Boundary artifacts
  - Artificial truncation of the computational domain
  - Approximation of non-additive noise by additive noise

Schur Complements and Conditioning

Assume that $X \sim \mathcal{N}(0, \Gamma)$, where $\Gamma \in \mathbb{R}^{n \times n}$ is a given SPD matrix.
Partitioning of X,

$$X = \left[\begin{array}{c} X_1 \\ X_2 \end{array} \right] \begin{array}{l} \in \mathbb{R}^k \\ \in \mathbb{R}^{n-k} \end{array} .$$

Question: *Assume that $X_2 = x_2$ is observed. What is the conditional probability density of X_1,*

$$\pi_{X_1}(x_1 \mid x_2) = ?$$

Schur Complements and Conditioning

Write

$$\pi_X(x) = \pi_{X_1,X_2}(x_1, x_2).$$

Bayes' formula: The distribution of unknown part x_1 provided that x_2 is known, is

$$\pi_{X_1}(x_1 \mid x_2) \propto \pi_{X_1,X_2}(x_1, x_2), \quad x_2 = x_{2,\mathrm{observed}}.$$

In terms of the Gaussian density,

$$\pi_{X_1,X_2}(x_1, x_2) \propto \exp\left(-\frac{1}{2}x^\mathsf{T}\Gamma^{-1}x\right). \tag{1}$$

Schur Complements and Conditioning

Partitioning of the covariance matrix:

$$\Gamma = \left[\begin{array}{cc} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{array}\right] \in \mathbb{R}^{n\times n}, \tag{2}$$

where

$$\Gamma_{11} \in \mathbb{R}^{k\times k}, \quad \Gamma_{22} \in \mathbb{R}^{(n-k)\times(n-k)}, \quad k < n,$$

and

$$\Gamma_{12} = \Gamma_{21}^{\mathsf{T}} \in \mathbb{R}^{k\times(n-k)}.$$

Schur Complements and Conditioning

Precision matrix $\mathsf{B} = \Gamma^{-1}$.
Partition B:

$$\mathsf{B} = \begin{bmatrix} \mathsf{B}_{11} & \mathsf{B}_{12} \\ \mathsf{B}_{21} & \mathsf{B}_{22} \end{bmatrix} \in \mathbb{R}^{n\times n}. \tag{3}$$

Quadratic form $x^{\mathsf{T}}\mathsf{B}x$ appearing in the exponential:

$$\mathsf{B}x = \begin{bmatrix} \mathsf{B}_{11} & \mathsf{B}_{12} \\ \mathsf{B}_{21} & \mathsf{B}_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mathsf{B}_{11}x_1 + \mathsf{B}_{12}x_2 \\ \mathsf{B}_{21}x_1 + \mathsf{B}_{22}x_2 \end{bmatrix},$$

Schur Complements and Conditioning

Completing the square,

$$
\begin{aligned}
x^{\mathsf T}\mathsf{B}x &= \begin{bmatrix} x_1^{\mathsf T} & x_2^{\mathsf T} \end{bmatrix} \begin{bmatrix} \mathsf{B}_{11}x_1 + \mathsf{B}_{12}x_2 \\ \mathsf{B}_{21}x_1 + \mathsf{B}_{22}x_2 \end{bmatrix} \\
&= x_1^{\mathsf T}\left(\mathsf{B}_{11}x_1 + \mathsf{B}_{12}x_2\right) + x_2^{\mathsf T}\left(\mathsf{B}_{21}x_1 + \mathsf{B}_{22}x_2\right) \\
&= x_1^{\mathsf T}\mathsf{B}_{11}x_1 + 2x_1^{\mathsf T}\mathsf{B}_{12}x_2 + x_2^{\mathsf T}\mathsf{B}_{22}x_2 \\
&= \left(x_1 + \mathsf{B}_{11}^{-1}\mathsf{B}_{12}x_2\right)^{\mathsf T}\mathsf{B}_{11}\left(x_1 + \mathsf{B}_{11}^{-1}\mathsf{B}_{12}x_2\right) \\
&\qquad + \underbrace{x_2^{\mathsf T}\left(\mathsf{B}_{22} - \mathsf{B}_{21}\mathsf{B}_{11}^{-1}\mathsf{B}_{12}\right)x_2}_{\text{independent of } x_1}.
\end{aligned}
$$

Schur Complements and Conditioning

Hence,

$$
\begin{aligned}
\pi_{X_1,X_2}(x_1,x_2) &\propto \exp\left(-\frac{1}{2}x^{\mathsf{T}}\mathsf{B}x\right) \\
&= \exp\left(-\frac{1}{2}\left(x_1+\mathsf{B}_{11}^{-1}\mathsf{B}_{12}x_2\right)^{\mathsf{T}}\mathsf{B}_{11}\left(x_1+\mathsf{B}_{11}^{-1}\mathsf{B}_{12}x_2\right)\right) \\
&\qquad \times \underbrace{\exp\left(-\frac{1}{2}x_2^{\mathsf{T}}\left(\mathsf{B}_{22}-\mathsf{B}_{21}\mathsf{B}_{11}^{-1}\mathsf{B}_{12}\right)x_2\right)}_{\text{scalar depending on data}} \\
&= C\exp\left(-\frac{1}{2}\left(x_1+\mathsf{B}_{11}^{-1}\mathsf{B}_{12}x_2\right)^{\mathsf{T}}\mathsf{B}_{11}\left(x_1+\mathsf{B}_{11}^{-1}\mathsf{B}_{12}x_2\right)\right).
\end{aligned}
$$

Schur Complements and Conditioning

We conclude that

$$\pi_{X_1|X_2}(x_1 \mid x_2) \propto \exp\left(-\frac{1}{2}\left(x_1 + \mathsf{B}_{11}^{-1}\mathsf{B}_{12}x_2\right)^{\mathsf{T}}\mathsf{B}_{11}\left(x_1 + \mathsf{B}_{11}^{-1}\mathsf{B}_{12}x_2\right)\right).$$

Thus the conditional density is Gaussian, with mean

$$\boxed{\overline{x}_1 = -\mathsf{B}_{11}^{-1}\mathsf{B}_{12}x_2,}$$

and covariance matrix

$$\boxed{\mathsf{C} = \mathsf{B}_{11}^{-1}.}$$

Question: *How to express these formulas in terms of* Γ*?*

Schur Complements and Conditioning

Consider a partitioned SPD matrix $\Gamma \in \mathbb{R}^{n\times n}$.
For any $v \in \mathbb{R}^k$, $x \neq 0$

$$v^{\mathsf T}\Gamma_{11}v = \left[\begin{array}{cc} v^{\mathsf T} & 0 \end{array}\right]\left[\begin{array}{cc} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{array}\right]\left[\begin{array}{c} v^{\mathsf T} \\ 0 \end{array}\right] > 0.$$

Conclusions:

- Γ_{11} is SPD.
- With similar reasoning, Γ_{22} is SPD.
- In particular, both Γ_{11} and Γ_{22} are invertible.

Schur Complements and Conditioning

To calculate the inverse of Γ, we solve the equation

$$\Gamma x = y$$

in block form.
By partitioning,

$$x = \left[\begin{array}{c} x_1 \\ x_2 \end{array}\right] \begin{array}{l} \in \mathbb{R}^k \\ \in \mathbb{R}^{n-k} \end{array}, \quad y = \left[\begin{array}{c} y_1 \\ y_2 \end{array}\right] \begin{array}{l} \in \mathbb{R}^k \\ \in \mathbb{R}^{n-k} \end{array}.$$

we have

$$\begin{aligned} \Gamma_{11}x_1 + \Gamma_{12}x_2 &= y_1, \\ \Gamma_{21}x_1 + \Gamma_{22}x_2 &= y_2. \end{aligned}$$

Schur Complements and Conditioning

Eliminate x_2 from the second equation,

$$x_2 = \Gamma_{22}^{-1}\big(y_2 - \Gamma_{21}x_1\big),$$

substitute back into the first equation:

$$\Gamma_{11}x_1 + \Gamma_{12}\Gamma_{22}^{-1}\big(y_2 - \Gamma_{21}x_1\big) = y_1,$$

and by rearranging the terms,

$$\big(\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}\big)x_1 = y_1 - \Gamma_{12}\Gamma_{22}^{-1}y_2.$$

Define the **Schur complement** of Γ_{22}:

$$\widetilde{\Gamma}_{22} = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}.$$

Schur Complements and Conditioning

The fact that the original equation $\Gamma x = y$ has a unique solution implies that $\widetilde{\Gamma}_{22}$ is invertible, and

$$x_1 = \widetilde{\Gamma}_{22}^{-1} y_1 - \widetilde{\Gamma}_{22}^{-1} \Gamma_{12} \Gamma_{22}^{-1} y_2.$$

Similarly, interchanging the roles of x_1 and x_2,

$$x_2 = \widetilde{\Gamma}_{11}^{-1} y_2 - \widetilde{\Gamma}_{11}^{-1} \Gamma_{21} \Gamma_{11}^{-1} y_1,$$

where

$$\widetilde{\Gamma}_{11} = \Gamma_{22} - \Gamma_{21} \Gamma_{11}^{-1} \Gamma_{12}$$

is the Schur complement of Γ_{11}.

In matrix form:

$$\left[\begin{array}{c} x_1 \\ x_2 \end{array}\right] = \left[\begin{array}{cc} \widetilde{\Gamma}_{22}^{-1} & -\widetilde{\Gamma}_{22}^{-1} \Gamma_{12} \Gamma_{22}^{-1} \\ -\widetilde{\Gamma}_{11}^{-1} \Gamma_{21} \Gamma_{11}^{-1} & \widetilde{\Gamma}_{11}^{-1} \end{array}\right] \left[\begin{array}{c} y_1 \\ y_2 \end{array}\right]$$

Schur Complements and Conditioning

Conclusion:

$$\Gamma^{-1} = \left[\begin{array}{cc} \widetilde{\Gamma}_{22}^{-1} & -\widetilde{\Gamma}_{22}^{-1}\Gamma_{12}\Gamma_{22}^{-1} \\ -\widetilde{\Gamma}_{11}^{-1}\Gamma_{21}\Gamma_{11}^{-1} & \widetilde{\Gamma}_{11}^{-1} \end{array}\right] = \left[\begin{array}{cc} \mathsf{B}_{11} & \mathsf{B}_{12} \\ \mathsf{B}_{21} & \mathsf{B}_{22} \end{array}\right].$$

We therefore have

$$\begin{aligned} \overline{x}_1 &= -\mathsf{B}_{11}^{-1}\mathsf{B}_{12}x_2 = -\widetilde{\Gamma}_{22}\big(-\widetilde{\Gamma}_{22}^{-1}\Gamma_{12}\Gamma_{22}^{-1}\big) \\ &= \Gamma_{12}\Gamma_{22}^{-1}, \end{aligned}$$

and

$$B_{11}^{-1} = \widetilde{\Gamma}_{22}.$$

Schur Complements and Conditioning

Theorem

The conditional density $\pi_{X_1,X_2}(x_1 \mid x_2)$ *is a Gaussian,*

$$\pi_{X_1,X_2}(x_1 \mid x_2) \sim \mathcal{N}(\overline{x}_1, \mathsf{C}),$$

where the mean is given by

$$\boxed{\overline{x}_1 = \Gamma_{12}\Gamma_{22}^{-1}x_2,}$$

and the covariance is

$$\boxed{\mathsf{C} = \widetilde{\Gamma}_{22} = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}.}$$

Linear Inverse Problems

Apply to the problem of estimating x from

$$b = \mathsf{A}x + e, \quad \mathsf{A} \in \mathbb{R}^{m\times n}, \quad b = b_{\rm observed}.$$

Stochastic extension: Write a relation between random variables,

$$B = \mathsf{A}X + E,$$

and assume the noise and prior model

$$X \sim \mathcal{N}(0, \mathsf{D}), \quad E \sim \mathcal{N}(0, \Sigma).$$

Usually it is assumed that X and E are independent. In particular,

$$\mathsf{E}\{XE^{\mathsf{T}}\} = \mathsf{E}\{X\}\mathsf{E}\{E\}^{\mathsf{T}} = 0.$$

Linear Inverse Problems

However, this is not necessary, and we may have

$$\mathrm{E}\{XE^{\mathsf{T}}\} = \mathsf{R} \in \mathbb{R}^{n\times m}.$$

Define a new random variable

$$Z = \begin{bmatrix} X \\ B \end{bmatrix} \in \mathbb{R}^{n+m}.$$

Covariance matrix of Z:

$$\begin{aligned} ZZ^{\mathsf{T}} &= \begin{bmatrix} X \\ B \end{bmatrix} \begin{bmatrix} X^{\mathsf{T}} & B^{\mathsf{T}} \end{bmatrix} \\ &= \begin{bmatrix} XX^{\mathsf{T}} & XB^{\mathsf{T}} \\ BX^{\mathsf{T}} & BB^{\mathsf{T}} \end{bmatrix}. \end{aligned}$$

Linear Inverse Problems

Expectations:

$$\mathsf{E}\{XX^{\mathsf{T}}\} = \mathsf{D},$$

$$\begin{aligned}\mathsf{E}\{XB^{\mathsf{T}}\} &= \mathsf{E}\{X(\mathsf{A}X + E)^{\mathsf{T}}\} = \mathsf{E}\{XX^{\mathsf{T}}\mathsf{A}^{\mathsf{T}} + XE^{\mathsf{T}}\} \\ &= \mathsf{E}\{XX^{\mathsf{T}}\}\mathsf{A}^{\mathsf{T}} + \mathsf{E}\{XE^{\mathsf{T}}\} = \mathsf{D}\mathsf{A}^{\mathsf{T}} + \mathsf{R}.\end{aligned}$$

Furthermore,

$$\mathsf{E}\{BX^{\mathsf{T}}\} = \mathsf{E}\{XB^{\mathsf{T}}\}^{\mathsf{T}} = \mathsf{A}\mathsf{D} + \mathsf{R}^{\mathsf{T}},$$

Linear Inverse Problems

and, finally,

$$\begin{aligned} \mathsf{E}\left\{BB^\mathsf{T}\right\} &= \mathsf{E}\left\{(\mathsf{A}X+E)(\mathsf{A}X+E)^\mathsf{T}\right\} \\ &= \mathsf{E}\left\{\mathsf{A}XX^\mathsf{T}\mathsf{A}^\mathsf{T} + EX^\mathsf{T}\mathsf{A}^\mathsf{T} + \mathsf{A}XE^\mathsf{T} + EE^\mathsf{T}\right\} \\ &= \mathsf{ADA}^\mathsf{T} + \mathsf{R}^\mathsf{T}\mathsf{A}^\mathsf{T} + \mathsf{AR} + \Sigma. \end{aligned}$$

Conclusion:

$$\mathrm{Cov}(Z) = \left[\begin{array}{cc} \mathsf{D} & \mathsf{DA}^\mathsf{T} + \mathsf{R} \\ \mathsf{AD} + \mathsf{R}^\mathsf{T} & \mathsf{ADA}^\mathsf{T} + \mathsf{R}^\mathsf{T}\mathsf{A}^\mathsf{T} + \mathsf{AR} + \Sigma \end{array}\right] = \left[\begin{array}{cc} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{array}\right].$$

Linear Inverse Problems

For simplicity, let us assume that $\mathsf{R} = 0$.

Theorem

Given the linear observation model

$$B = \mathsf{A}X + E, \quad X \sim \mathcal{N}(0, \mathsf{D}), \quad \mathsf{E} \sim \mathcal{N}(0, \Sigma),$$

Posterior density $\pi_{X|B}(x \mid b)$ *is a Gaussian density, with mean*

$$\boxed{\overline{x} = \Gamma_{12}\Gamma_{22}^{-1} b = \mathsf{DA}^{\mathsf{T}}\big(\mathsf{ADA}^{\mathsf{T}} + \Sigma\big)^{-1} b,}$$

and covariance

$$\boxed{\mathsf{C} = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21} = \mathsf{D} - \mathsf{DA}^{\mathsf{T}}\big(\mathsf{ADA}^{\mathsf{T}} + \Sigma\big)^{-1}\mathsf{AD}.}$$

Gaussian distributions and information

Consider two Gaussian densities

$$\pi_1(x) \propto \exp\left(-\frac{1}{2}(x - x_1)^\mathsf{T} \mathsf{C}_1^{-1}(x - x_1)\right),$$

$$\pi_2(x) \propto \exp\left(-\frac{1}{2}(x - x_2)^\mathsf{T} \mathsf{C}_2^{-1}(x - x_2)\right).$$

where C_1, C_2 are symmetric positive definite. We say that π_1 is **more informative than** π_2 if

$$v^\mathsf{T} \mathsf{C}_1 v \leq v^\mathsf{T} \mathsf{C}_2 v$$

for all $v \neq 0$, that is, the variance of π_1 in any direction is always equal or smaller than that of π_2. We write

$$\mathsf{C}_1 \leq \mathsf{C}_2,$$

Gaussian distributions and information

A small exercise in matrix algebra: *Show that*

$$\mathsf{C}_1 \leq \mathsf{C}_2 \Rightarrow \mathsf{C}_2^{-1} \leq \mathsf{C}_1^{-1}.$$

This statement says that if the variance is smaller, the precision is larger. From the previous theorem, we conclude: If C is the posterior covariance and D is the prior covariance, then

$$\begin{aligned} v^{\mathsf{T}}\mathsf{C}v &= v^{\mathsf{T}}\mathsf{D}v - \underbrace{v^{\mathsf{T}}\mathsf{D}\mathsf{A}^{\mathsf{T}}\big(\mathsf{A}\mathsf{D}\mathsf{A}^{\mathsf{T}} + \Sigma\big)^{-1}\mathsf{A}\mathsf{D}v}_{\geq 0} \\ &\leq v^{\mathsf{T}}\mathsf{D}v. \end{aligned}$$

Conclusion: The posterior is always more informative than the prior.

Linear Inverse Problems

An alternative (but equivalent) formula: Write the prior and likelihood as

$$\begin{aligned} \pi_X(x) &\propto \exp\left(-\frac{1}{2}x^{\mathsf{T}}\mathsf{D}^{-1}x\right), \\ \pi_{B|X}(b \mid x) &\propto \exp\left(-\frac{1}{2}(b-\mathsf{A}x)^{\mathsf{T}}\Sigma^{-1}(b-\mathsf{A}x)\right), \end{aligned}$$

The posterior density is

$$\begin{aligned} \pi_{X|B}(x \mid b) &\propto \pi_X(x)\pi_{B|X}(b \mid x) \\ &\propto \exp\left(-\frac{1}{2}x^{\mathsf{T}}\mathsf{D}^{-1}x - \frac{1}{2}(b-\mathsf{A}x)^{\mathsf{T}}\Sigma^{-1}(b-\mathsf{A}x)\right). \end{aligned}$$

Linear Inverse Problems

Consider the exponent in the posterior density:

$$Q(x) = (b - \mathsf{A}x)^{\mathsf{T}}\Sigma^{-1}(b - \mathsf{A}x) + x^{\mathsf{T}}\mathsf{D}^{-1}x$$

Collect the terms of the same order in x together:

$$Q(x) = x^{\mathsf{T}}\underbrace{\left(\mathsf{A}^{\mathsf{T}}\Sigma^{-1}\mathsf{A} + \mathsf{D}^{-1}\right)}_{=\mathsf{M}}x - 2x^{\mathsf{T}}\mathsf{A}^{\mathsf{T}}\Sigma^{-1}b + b^{\mathsf{T}}\Sigma^{-1}b.$$

Complete the square:

$$\begin{aligned} Q(x) \quad &= \quad (x^{\mathsf{T}} - \mathsf{M}^{-1}\mathsf{A}^{\mathsf{T}}\Sigma^{-1}b)^{\mathsf{T}}\mathsf{M}(x^{\mathsf{T}} - \mathsf{M}^{-1}\mathsf{A}^{\mathsf{T}}\Sigma^{-1}b) \\ &\qquad + \text{terms independent of } x. \end{aligned}$$

Linear Inverse Problems

We conclude that the posterior density is of the form

$$\pi_{X|B}(x \mid b) \propto \exp\left(-\frac{1}{2}(x^{\mathsf{T}} - \mathsf{M}^{-1}\mathsf{A}^{\mathsf{T}}\Sigma^{-1}b)^{\mathsf{T}}\mathsf{M}(x^{\mathsf{T}} - \mathsf{M}^{-1}\mathsf{A}^{\mathsf{T}}\Sigma^{-1}b)\right),$$

where

$$\mathsf{M} = \mathsf{A}^{\mathsf{T}}\Sigma^{-1}\mathsf{A} + \mathsf{D}^{-1}.$$

This gives us an alternative form of the posterior density.

Linear Inverse Problems

Theorem

Given the linear observation model

$$B = \mathsf{A}X + E, \quad X \sim \mathcal{N}(0, \mathsf{D}), \quad \mathsf{E} \sim \mathcal{N}(0, \Sigma),$$

the posterior density $\pi_{X|B}(x \mid b)$ *is a Gaussian density, with mean*

$$\boxed{\overline{x} = \left(\mathsf{A}^\mathsf{T}\Sigma^{-1}\mathsf{A} + \mathsf{D}^{-1}\right)^{-1}\mathsf{A}^\mathsf{T}\Sigma^{-1}b}$$

and covariance

$$\boxed{\mathsf{C} = \left(\mathsf{A}^\mathsf{T}\Sigma^{-1}\mathsf{A} + \mathsf{D}^{-1}\right)^{-1}.}$$

- The formula for $\overline{x}$ is also known as *Wiener filtered solution*.
- The equivalence of the two different formulations can be shown by using the Sherman-Morrison-Woodbury identity.

Sherman-Morrison-Woodbury identity

Theorem

Let $\mathsf{M} \in \mathbb{R}^{n\times n}$ *be an invertible matrix, and* $\mathsf{U}, \mathsf{V} \in \mathbb{R}^{n\times k}$ *matrices, and let*

$$\mathsf{M}_+ = \mathsf{M} + \mathsf{U}\mathsf{V}^\mathsf{T}.$$

Then the inverse of M_+ *can be written as*

$$\mathsf{M}_+^{-1} = \mathsf{M}^{-1} - \mathsf{M}^{-1}\mathsf{U}\big(\mathsf{I}_k + \mathsf{V}^\mathsf{T}\mathsf{M}^{-1}\mathsf{U}\big)^{-1}\mathsf{V}^\mathsf{T}\mathsf{M}^{-1}.$$

Sherman-Morrison-Woodbury identity

Consider now the alternative formula for the posterior covariance:

$$C = \left(A^T \Sigma^{-1} A + D^{-1}\right)^{-1}.$$

In the SMW formula, identify

$$D^{-1} = M, \quad U = A^T \Sigma^{-1/2} = V,$$

Then,

$$\begin{aligned} I + V^T M^{-1} U &= I + \Sigma^{-1/2} A D A^T \Sigma^{-1/2} \\ &= \Sigma^{-1/2}\left(\Sigma + A D A^T\right)\Sigma^{-1/2}, \end{aligned}$$

and

$$C = M_+^{-1} = D - D A^T \left(\Sigma + A D A^T\right)^{-1} A D,$$

as claimed.