Computational inverse problems

Nuutti Hyvönen, Jenni Heino,
Juha-Pekka Puska

nuutti.hyvonen@aalto.fi, juha-pekka.puska@aalto.fi

Ninth lecture, March 29, 2021.

Discontinuities

Prior information: The unknown is a function of, say, time. It is known to be relatively stable for long periods of time, but contains now and then discontinuities. We may also have information on the size of the jumps or the rate of occurrence of the discontinuities.

A more concrete example: Unknown is a function $f : [0, 1] \to \mathbb{R}$. We know that $f(0) = 0$ and that the function may have large jumps at a few locations.

After discretizing f, impulse priors can be used to construct a prior on the finite difference approximation of the *derivative* of f.

Discretization of the interval $[0, 1]$: Choose grid points $t_j = j/N$, $j = 0, \ldots, N$, and set $x_j = f(t_j)$.

We write a Cauchy-type prior density

$$\pi_{\mathrm{pr}}(x) = \left(\frac{\alpha}{\pi}\right)^N \prod_{j=1}^{N} \frac{1}{1 + \alpha^2 (x_j - x_{j-1})^2}$$

that controls the jumps between the adjacent components of $x \in \mathbb{R}^{N+1}$. In particular, the components of X are not independent. (In addition to this prior, we know that $X_0 = x_0 = 0$.)

To make draws from the above density, we define new variables

$$\xi_j = x_j - x_{j-1}, \qquad 1 \leq j \leq N,$$

which are the changes in the function of interest between adjacent grid points.

Notice that $\tilde{x} = [x_1, \ldots, x_N]^{\mathrm{T}} \in \mathbb{R}^N$ satisfies

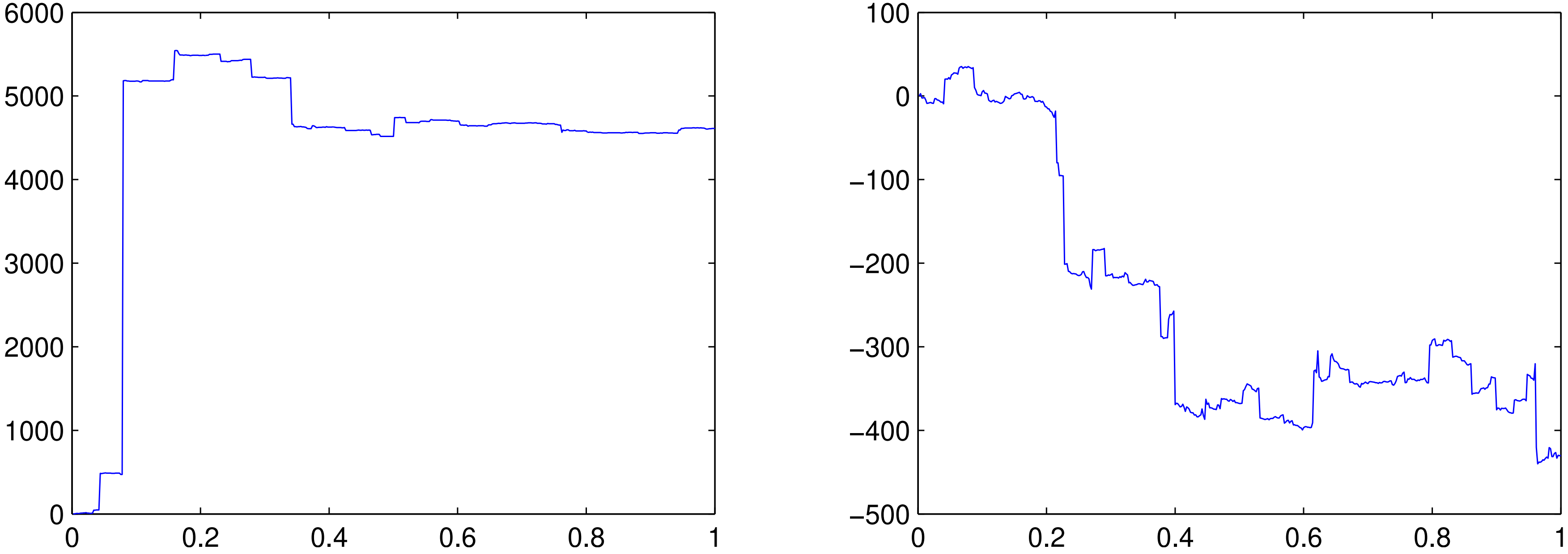
$$\tilde{x} = A\xi,$$

where $A \in \mathbb{R}^{N \times N}$ is a lower triangular matrix such that $A_{jk} = 1$ for $j \geq k$. Hence, it follows, e.g., from the change of variables rule for probability densities that

$$\pi_{\mathrm{pr}}(\xi) = \left(\frac{\alpha}{\pi}\right)^N \prod_{j=1}^{N} \frac{1}{1 + \alpha^2 \xi_j^2}.$$

In particular, due to the product form of $\pi_{\mathrm{pr}}(\xi)$, the components of Ξ are mutually independent, and can thus be drawn from a one-dimensional Cauchy density.

Subsequently, a random draw from the distribution of X can be constructed by recalling that $x_0 = 0$ and using the relation $\tilde{x} = A\xi$.


6000
5000
4000
3000
2000
1000
0
0
0.2
0.4
0.6
0.8
1
100
0
−100
−200
−300
−400
−500
0
0.2
0.4
0.6
0.8
1

Sample-based densities

Assume that we have a large sample of realizations of a random variable $X \in \mathbb{R}^n$:

$$S = \{x^1, x^2, \ldots, x^N\}.$$

One way to construct a prior density for X is to approximate $\pi(x)$ based on S.

Estimates of the mean and the covariance:

$$E\{X\} \approx \frac{1}{N} \sum_{j=1}^{N} x^j =: \bar{x},$$

$$\mathrm{cov}(X) = E\{XX^{\mathrm{T}}\} - E\{X\}E\{X\}^{\mathrm{T}} \approx \frac{1}{N} \sum_{j=1}^{N} x^j (x^j)^{\mathrm{T}} - \bar{x}\bar{x}^{\mathrm{T}} =: \Gamma.$$

(Notice that Γ is not the unbiased sample covariance estimator, but let us anyway follow the notation of the text book.)

The eigenvalue decomposition of Γ is

$$\Gamma = UDU^{\mathrm{T}},$$

where $U \in \mathbb{R}^{n\times n}$ is orthogonal and has the eigenvectors of Γ as its columns, and $D \in \mathbb{R}^{n\times n}$ is diagonal with the eigenvalues $d_1 \geq \ldots \geq d_n \geq 0$ as its diagonal entries. (Note that Γ is clearly symmetric and positive semi-definite, and thus it has a full set of eigenvectors with non-negative eigenvalues.)

The vectors x^j, $j = 1, \ldots, N$, are typically 'somewhat similar' and the matrix Γ can consequently be singular or almost singular: The eigenvalues often satisfy $d_j \approx 0$ for $j > r$, where $1 < r < n$ is some cut-off index. In other words, the difference $X - E\{X\}$ does not seem to vary much in the direction of the eigenvectors $u_{r+1}, \ldots, u_n$.

Assume this is the case. Then, one can postulate that the values of the random variable $X - E(X)$ lie ‘with a high probability’ in the subspace spanned by the first r eigenvectors of Γ. One way of trying to state this information quantitatively, is to introduce a *subspace prior*

$$\pi(x) \propto \exp\left(-\alpha \|(I - P)(x - \bar{x})\|^2\right),$$

where P is the orthogonal projector $\mathbb{R}^n \to \mathrm{span}\{u_1, \ldots, u_r\}$. The parameter $\alpha > 0$ controls how much $X - \bar{x}$ is allowed to vary from the subspace $\mathrm{span}\{u_1, \ldots, u_r\}$. (Take note that such a subspace prior is not a probability density in the traditional sense.)

If Γ is not almost singular, the inverse Γ^{-1} can be computed stably. In this case, the most straightforward way of approximating the (prior) probability density of X is to introduce the Gaussian approximation:

$$\pi_{\mathrm{pr}}(x) \propto \exp\left(-\frac{1}{2}(x-\bar{x})^{\mathrm{T}}\Gamma^{-1}(x-\bar{x})\right).$$

Depending on the higher order statistics of X, this may or may not provide a good approximation for the distribution of X.

Posterior density and a simple linear model

Consider a linear system of equations with noisy right hand side,

$$y = Ax + e, \qquad x \in \mathbb{R}^n,\ y, e \in \mathbb{R}^m,\ A \in \mathbb{R}^{m \times n}.$$

The corresponding stochastic extension reads

$$Y = AX + E,$$

where X, Y and E are random variables.

A very common assumption: X and E are independent and Gaussian,

$$X \sim \mathcal{N}(0, \gamma^2 \Gamma), \qquad E \sim \mathcal{N}(0, \sigma^2 I),$$

where we have assumed that both X and E have zero mean. (If this was not the case, the means could be subtracted from the respective random variables.)

The covariance of the noise indicates that the components of Y are contaminated by independent and identically distributed Gaussian random variables of variance σ^2. On the other hand, the prior distribution of X is assumed to have a bit more structure: Γ need not be diagonal and the parameter γ^2 is introduced for controlling the 'magnitude' of the (prior) covariance.

In other words, the prior density is of the form

$$\pi_{\mathrm{pr}}(x) \propto \exp\left(-\frac{1}{2\gamma^2} x^{\mathrm{T}} \Gamma^{-1} x\right),$$

and assuming that the noise level σ^2 is known, the likelihood function reads as

$$\pi(y \,|\, x) \propto \exp\left(-\frac{1}{2\sigma^2} \|y - Ax\|^2\right).$$

It follows from the Bayes formula that the posterior density is

$$\begin{aligned}\pi(x \,|\, y) &\propto \pi_{\mathrm{pr}}(x)\pi(y \,|\, x) \\ &\propto \exp\left(-\frac{1}{2\gamma^2}x^{\mathrm{T}}\Gamma^{-1}x - \frac{1}{2\sigma^2}\|y - Ax\|^2\right) \\ &= \exp(-V(x \,|\, y)),\end{aligned}$$

where

$$V(x \,|\, y) = \frac{1}{2\gamma^2}x^{\mathrm{T}}\Gamma^{-1}x + \frac{1}{2\sigma^2}\|y - Ax\|^2.$$

If Γ is symmetric and positive definite, so is Γ^{-1}. Hence, we can introduce a Cholesky factorization:

$$\Gamma^{-1} = R^{\mathrm{T}}R.$$

With this notation,

$$x^{\mathrm{T}}\Gamma^{-1}x = x^{\mathrm{T}}R^{\mathrm{T}}Rx = \|Rx\|^2,$$

and we define

$$T(x) = 2\sigma^2 V(x \,|\, y) = \|y - Ax\|^2 + \delta\|Rx\|^2, \qquad \delta := \frac{\sigma^2}{\gamma^2}.$$

The functional T is sometimes referred to as the *Tikhonov functional*.

Recall that the *maximum a posteriori (MAP)* estimator maximizes the posterior probability density of the unknowns:

$$x_{\mathrm{MAP}} = \arg\max_{x \in \mathbb{R}^n} \pi(x \,|\, y).$$

In our setting,

$$x_{\mathrm{MAP}} = \arg\min V(x \,|\, y) \quad \text{because} \quad V(x \,|\, y) = -\log \pi(x \,|\, y).$$

With the help of the Tikhonov functional, this reads

$$x_{\mathrm{MAP}} = \arg\min\, T(x) = \arg\min \left(\|y - Ax\|^2 + \delta \|Rx\|^2 \right).$$

Recall that the Tikhonov regularized solution of $y = Ax$ — with the penalty term $\|Rx\|$ — is the minimizer of $T(x)$. In consequence, the Tikhonov regularized solution and x_{MAP} coincide if the regularization parameter is chosen to be $\delta = \sigma^2/\gamma^2$.

n-variate Gaussian densities

Definition. *Let*

$$\Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} \in \mathbb{R}^{n\times n}$$

be a positive definite and symmetric matrix, with $\Gamma_{11} \in \mathbb{R}^{k\times k}$*,* $k < n$*,* $\Gamma_{22} \in \mathbb{R}^{(n-k)\times(n-k)}$*, and* $\Gamma_{21} = \Gamma_{12}^{\mathrm{T}} \in \mathbb{R}^{(n-k)\times k}$*. We define the Schur complement* $\tilde{\Gamma}_{jj}$ *of* Γ_{jj}*,* $j = 1, 2$*, by the formulas*

$$\tilde{\Gamma}_{22} = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}, \qquad \tilde{\Gamma}_{11} = \Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}$$

Observe that the definition of Γ implies that Γ_{jj}, $j = 1, 2$, are symmetric, positive definite and, in particular, invertible. In consequence, the Schur complements are well defined and symmetric.

Lemma. *Let* Γ *be a matrix that satisfies the assumptions of the previous definition. Then, the Schur complements* $\tilde{\Gamma}_{jj}$, $j = 1, 2$, *are invertible matrices and, furthermore,*

$$\Gamma^{-1} = \begin{bmatrix} \tilde{\Gamma}_{22}^{-1} & -\tilde{\Gamma}_{22}^{-1}\Gamma_{12}\Gamma_{22}^{-1} \\ -\tilde{\Gamma}_{11}^{-1}\Gamma_{21}\Gamma_{11}^{-1} & \tilde{\Gamma}_{11}^{-1} \end{bmatrix}.$$

Proof: We prove first that the Schur complements are invertible: Consider the determinant of Γ,

$$|\Gamma| = \begin{vmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{vmatrix} \neq 0.$$

By subtracting the first row multiplied by $\Gamma_{21}\Gamma_{11}^{-1}$ from the second one, we find that

$$|\Gamma| = \begin{vmatrix} \Gamma_{11} & \Gamma_{12} \\ 0 & \Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12} \end{vmatrix} = |\Gamma_{11}||\tilde{\Gamma}_{11}|,$$

implying that $|\tilde{\Gamma}_{11}| \neq 0$. In the same way, we can also show that $|\tilde{\Gamma}_{22}| \neq 0$.

The proof of the second assertion of the lemma follows from the Gaussian elimination: Consider the linear system

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

By solving for x_2 in the second equation, we get

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

Substituting this formula into the first equation, then gives us

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

or equivalently

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

which verifies the first row of claimed representation of Γ^{-1}. The second row of the representation follows by reversing the roles of x_1 and x_2.

Remark: Since Γ is a symmetric matrix, so is Γ^{-1}. In consequence, we have the identity

$$\tilde{\Gamma}_{11}^{-1}\Gamma_{21}\Gamma_{11}^{-1} = (\tilde{\Gamma}_{22}^{-1}\Gamma_{12}\Gamma_{22}^{-1})^{\mathrm{T}} = \Gamma_{22}^{-1}\Gamma_{21}\tilde{\Gamma}_{22}^{-1}.$$

Theorem. *Let $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^m$ be two Gaussian random variables whose joint probability density $\pi \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_+$ is of the form*

$$\pi(x,y) \propto \exp\left(-\frac{1}{2} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \right).$$

Then, the probability density of X conditioned on $Y = y$, i.e., $\pi(x \,|\, y) \colon \mathbb{R}^n \to \mathbb{R}_+$, is of the form

$$\pi(x \,|\, y) \propto \exp\left(-\frac{1}{2}(x - \bar{x})^{\mathrm{T}} \tilde{\Gamma}_{22}^{-1} (x - \bar{x}) \right),$$

where

$$\bar{x} = x_0 + \Gamma_{12}\Gamma_{22}^{-1}(y - y_0).$$

Proof: For simplicity, let us assume that $x_0 = 0$ and $y_0 = 0$.

Due the representation of the joint covariance matrix Γ^{-1} provided by the previous Lemma and the remark that followed, we may write

$$\begin{aligned}\pi(x,y) &\propto \exp\left(-\frac{1}{2}\big(x^{\mathrm{T}}\tilde{\Gamma}_{22}^{-1}x - 2x^{\mathrm{T}}\tilde{\Gamma}_{22}^{-1}\Gamma_{12}\Gamma_{22}^{-1}y + y^{\mathrm{T}}\tilde{\Gamma}_{11}^{-1}y\big)\right) \\ &= \exp\left(-\frac{1}{2}\big((x-\Gamma_{12}\Gamma_{22}^{-1}y)^{\mathrm{T}}\tilde{\Gamma}_{22}^{-1}(x-\Gamma_{12}\Gamma_{22}^{-1}y)+c\big)\right),\end{aligned}$$

where $c = y^{\mathrm{T}}(\tilde{\Gamma}_{11}^{-1} - \Gamma_{22}^{-1}\Gamma_{21}\tilde{\Gamma}_{22}^{-1}\Gamma_{12}\Gamma_{22}^{-1})y$. Hence, it follows that

$$\pi(x\,|\,y) \propto \pi(x,y) \propto \exp\left(-\frac{1}{2}(x-\Gamma_{12}\Gamma_{22}^{-1}y)^{\mathrm{T}}\tilde{\Gamma}_{22}^{-1}(x-\Gamma_{12}\Gamma_{22}^{-1}y)\right),$$

where the proportionality constants depend on y but not on x. This proves the claim. □

Theorem. *Let X and Y be Gaussian random variables with a joint probability density as in the previous theorem. Then, the marginal density of X is*

$$\pi(x) = \int_{\mathbb{R}^m} \pi(x,y)dy \propto \exp\left(-\frac{1}{2}(x-x_0)^{\mathrm{T}}\Gamma_{11}^{-1}(x-x_0)\right).$$

Proof: The proof is slightly more complicated than the previous one. It can be found in the textbook by Kaipio and Somersalo.

Linear inverse problem

Assume that we have a linear model with additive noise,

$$Y = AX + E,$$

where $A \in \mathbb{R}^{m\times n}$ is a known matrix, and $X \in \mathbb{R}^n$ and $Y, E \in \mathbb{R}^m$ are random variables. Assume furthermore that X and E are mutually independent Gaussian variables with probability densities

$$\pi_{\mathrm{pr}}(x) \propto \exp\left(-\frac{1}{2}(x - x_0)^{\mathrm{T}}\Gamma_{\mathrm{pr}}^{-1}(x - x_0)\right),$$

and

$$\pi_{\mathrm{noise}}(e) \propto \exp\left(-\frac{1}{2}(e - e_0)^{\mathrm{T}}\Gamma_{\mathrm{noise}}^{-1}(e - e_0)\right).$$

With this information, we get from the Bayes formula that the posterior distribution of X conditioned on $Y = y$ is

$$\pi(x \,|\, y) \propto \pi_{\mathrm{pr}}(x)\pi(y \,|\, x) = \pi_{\mathrm{pr}}(x)\pi_{\mathrm{noise}}(y - Ax)$$
$$\propto \exp\left(-\frac{1}{2}(x - x_0)^{\mathrm{T}}\Gamma_{\mathrm{pr}}^{-1}(x - x_0) - \frac{1}{2}(y - Ax - e_0)^{\mathrm{T}}\Gamma_{\mathrm{noise}}^{-1}(y - Ax - e_0)\right)$$

The explicit form of this posterior distribution, i.e., the form that shows the posterior mean and covariance explicitly, can be calculated in a straightforward but tedious manner by ‘completing the squares’ with respect to x. However, we may also use the first of the two theorems presented on the previous few slides.

Since X and E are Gaussian, so is Y, and we have

$$E\left\{\begin{bmatrix} X \\ Y \end{bmatrix}\right\} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}, \qquad y_0 = Ax_0 + e_0$$

Furthermore, using the fact that X and E are independent, we deduce that

$$E\left\{(X - x_0)(X - x_0)^{\mathrm{T}}\right\} = \Gamma_{\mathrm{pr}},$$

$$\begin{aligned} E\left\{(Y - y_0)(Y - y_0)^{\mathrm{T}}\right\} &= E\left\{\big(A(X-x_0)+(E-e_0)\big)\big(A(X-x_0)+(E-e_0)\big)^{\mathrm{T}}\right\} \\ &= A\Gamma_{\mathrm{pr}}A^{\mathrm{T}} + \Gamma_{\mathrm{noise}}, \end{aligned}$$

$$\begin{aligned} E\left\{(X - x_0)(Y - y_0)^{\mathrm{T}}\right\} &= E\left\{(X - x_0)\big(A(X - x_0) + (E - e_0)\big)^{\mathrm{T}}\right\} \\ &= \Gamma_{\mathrm{pr}}A^{\mathrm{T}}. \end{aligned}$$

Hence, we get

$$\mathrm{cov}\begin{bmatrix} X \\ Y \end{bmatrix} = E\left\{ \begin{bmatrix} X - x_0 \\ Y - y_0 \end{bmatrix} \begin{bmatrix} X - x_0 \\ Y - y_0 \end{bmatrix}^{\mathrm{T}} \right\} = \begin{bmatrix} \Gamma_{\mathrm{pr}} & \Gamma_{\mathrm{pr}} A^{\mathrm{T}} \\ A\Gamma_{\mathrm{pr}} & A\Gamma_{\mathrm{pr}} A^{\mathrm{T}} + \Gamma_{\mathrm{noise}} \end{bmatrix}.$$

The joint probability density of X and Y is thus of the form

$$\pi(x,y) \propto \exp\left(-\frac{1}{2} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \Gamma_{\mathrm{pr}} & \Gamma_{\mathrm{pr}} A^{\mathrm{T}} \\ A\Gamma_{\mathrm{pr}} & A\Gamma_{\mathrm{pr}} A^{\mathrm{T}} + \Gamma_{\mathrm{noise}} \end{bmatrix}^{-1} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \right).$$

Using the first of the above two theorems, we can thus write the posterior density of X conditioned on $Y = y$.

Theorem. *Assume that $X \in \mathbb{R}^n$ and $E \in \mathbb{R}^m$ are mutually independent Gaussian random variables,*

$$X \sim \mathcal{N}(x_0, \Gamma_{\mathrm{pr}}), \quad E \sim \mathcal{N}(e_0, \Gamma_{\mathrm{noise}})$$

and $\Gamma_{\mathrm{pr}} \in \mathbb{R}^{n \times n}$ and $\Gamma_{\mathrm{noise}} \in \mathbb{R}^{m \times m}$ are positive definite. Assume further that we have a linear model $Y = AX + E$ for a noisy measurement Y, where $A \in \mathbb{R}^{m \times n}$ is a known matrix. Then, the posterior probability density of X given the measurement $Y = y$ is

$$\pi(x \,|\, y) \propto \exp\left(-\frac{1}{2}(x - \bar{x})^{\mathrm{T}} \Gamma_{\mathrm{post}}^{-1} (x - \bar{x})\right),$$

where

$$\bar{x} = x_0 + \Gamma_{\mathrm{pr}} A^{\mathrm{T}} (A \Gamma_{\mathrm{pr}} A^{\mathrm{T}} + \Gamma_{\mathrm{noise}})^{-1} (y - A x_0 - e_0),$$

and

$$\Gamma_{\mathrm{post}} = \Gamma_{\mathrm{pr}} - \Gamma_{\mathrm{pr}} A^{\mathrm{T}} (A \Gamma_{\mathrm{pr}} A^{\mathrm{T}} + \Gamma_{\mathrm{noise}})^{-1} A \Gamma_{\mathrm{pr}}.$$

Remark: It holds that

$$\Gamma_{\mathrm{pr}} - \Gamma_{\mathrm{post}} = \Gamma_{\mathrm{pr}} A^{\mathrm{T}} (A \Gamma_{\mathrm{pr}} A^{\mathrm{T}} + \Gamma_{\mathrm{noise}})^{-1} A \Gamma_{\mathrm{pr}},$$

which is a positive semi-definite matrix. Loosely speaking, this means that the prior density is wider than the posterior, i.e., the measurement decreases the uncertainty in the whereabouts of X.

Remark: As already mentioned, the explicit forms of the mean and the covariance of the Gaussian posterior density for this linear model can also be derived directly. This way we get alternative representations for the posterior covariance matrix

$$\Gamma_{\mathrm{post}} = (\Gamma_{\mathrm{pr}}^{-1} + A^{\mathrm{T}} \Gamma_{\mathrm{noise}}^{-1} A)^{-1}$$

and the posterior mean

$$\bar{x} = \Gamma_{\mathrm{post}} (A^{\mathrm{T}} \Gamma_{\mathrm{noise}}^{-1} (y - e_0) + \Gamma_{\mathrm{pr}}^{-1} x_0).$$

Gaussian white noise prior and Tikhonov regularization

Consider the simple *Gaussian white noise prior* case, $X \sim \mathcal{N}(0, \gamma^2 I)$, and assume also that the noise is white noise, i.e., $E \sim (0, \sigma^2 I)$. In this particular case the mean of the posterior distribution given by the above theorem turns into

$$\bar{x} = \gamma^2 A^{\mathrm{T}}(\gamma^2 A A^{\mathrm{T}} + \sigma^2)^{-1} y = A^{\mathrm{T}}(A A^{\mathrm{T}} + \delta I)^{-1} y,$$

where $\delta = \sigma^2/\gamma^2$.

It can be shown that this form is equivalent to the Tikhonov regularized solution

$$x_\delta = (A^{\mathrm{T}} A + \delta I)^{-1} A^{\mathrm{T}} y,$$

which is not very surprising, as we have already deduced at the previous lecture that $x_{\mathrm{MAP}} = x_\delta$ for $\delta = \sigma^2/\gamma^2$ and, on the other hand, $x_{\mathrm{CM}} = x_{\mathrm{MAP}}$ for a Gaussian posterior distribution.