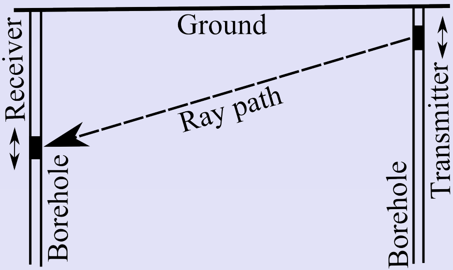
Gaussian priors

Example 5 (Matlab)



- Assume that two vertical boreholes have been drilled in the ground and the goal is to produce an image of electric permittivitivity ε of ground layers in between the holes.

Gaussian priors

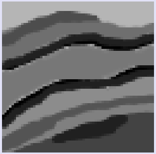
Example 5 (Matlab) continued

- The data are produced by transmitting a high power electromagnetic wave from one hole to the other where it is to be detected.
- The transmitter and receiver can move up and down in the holes and data are gathered using several distinct transmitter-receiver -location pairs.
- Recordings are done in both boreholes in order to obtain as complete set of data as possible.
- Transmitted rays are assumed to propagate linearly through the ground to the receiver. Possible reflections and refractions are neglected.

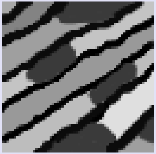
Gaussian priors

Example 5 (Matlab) continued

- The target of measurements is the phase (or phase shift) φ of the transmitted electromagnetic wave.
- The phase depends on permittivity according to the differential equation $d\varphi = c\varepsilon ds$, where c is a constant and ds is a differential path length.
- The exact ground layer structures are assumed to be the following 64×64 pixel images:



(a)



(b)

Gaussian priors

Example 5 (Matlab) continued

- The discretized unknown permittivity is denoted by X and the data by Y.
- These are related through the linear forward model $Y = AX + N$ (N = noise) in which A follows from $\varphi = \int_\ell d\varphi = c \int_\ell \varepsilon ds$ applied to a set of linear ray paths (ℓ = ray path) .
- The matrix A corresponds to sixteen different transmitter locations, eight equally spaced locations per a borehole, and to 64 equally spaced receiver positions per one transmitter location.
- Noise in the measurements is assumed to be zero-mean Gaussian white noise with standard deviation σ.

Gaussian priors

Example 5 (Matlab) continued

- The task is to find a MAP estimate corresponding to the following two priors:
  1. Gaussian white noise prior $X \sim \mathcal{N}(0, \alpha I)$,
  2. Gaussian structural prior $X \sim \mathcal{N}(0, \Gamma)$.
- In the latter prior $\Gamma = \alpha^2 (W^T W)^{-1}$ in which W is a discrete approximation of the directional differential operator $\partial_d = d_1 \partial_1 + d_2 \partial_2$.
- Direction $d = (d_1, d_2)$ is the *a priori* known direction of the ground layers, which in (a) and (b) makes an angle $\phi = \pi/9$ and $\phi = \pi/6$ with the ground (horizontal), respectively.

Gaussian priors

Example 5 (Matlab) continued

Solution

- In the case 1. (white noise prior), the MAP estimate coincides with the solution of Example 3, i.e. $x_{MAP} = [A^T A + (\sigma^2/\alpha^2) I]^{-1} A^T y$.
- In the case 2. (structural prior), the approach of Example 3 yields

$$
\begin{aligned}
x_{MAP} &= [A^T A + (\sigma^2/\alpha^2) W^T W]^{-1} A^T y \\
&= [W^T W^{T^{-1}} A^T A W^{-1} W + (\sigma^2/\alpha^2) W^T W]^{-1} A^T y \\
&= W^{-1} [W^{T^{-1}} A^T A W^{-1} + (\sigma^2/\alpha^2) I]^{-1} W^{T^{-1}} A^T y \\
&= W^{-1} [\widetilde{A}^T \widetilde{A} + (\sigma^2/\alpha^2) I]^{-1} \widetilde{A}^T y,
\end{aligned}
$$

with $\widetilde{A} = A W^{-1}$.

Gaussian priors

Example 5 (Matlab) continued

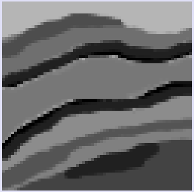
- Matrix W for $\phi = \pi/9$ and $\phi = \pi/6$ can be found as shown in (d)-part of Example 4.
- Since the dimensionality of the unknown (64×64) is somewhat higher than that of the data (16×64) it makes sense to write the formulas for the MAP estimates in the equivalent forms

$$
\begin{aligned}
(a) : x_{MAP} &= A^T[AA^T + (\sigma^2/\alpha^2)I]^{-1}y \\
(b) : x_{MAP} &= W^{-1}\widetilde{A}^T[\widetilde{A}\widetilde{A}^T + (\sigma^2/\alpha^2)I]^{-1}y,
\end{aligned}
$$

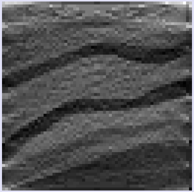
in which the matrix to be inverted is smaller. These formulas can be derived from the original ones, for example, using the singular value decomposition (SVD).

Gaussian priors

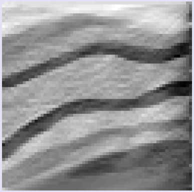
Example 5 (Matlab) continued



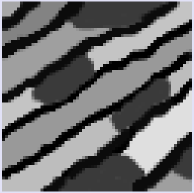
(a), exact



(a), white noise

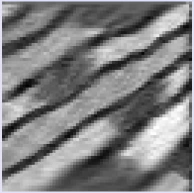


(a), structural



(b), exact

(b), white noise



(b), structural

Gaussian priors

Theorem 1

If the forward model is linear of the form $Y = AX + N$, where $A \in \mathbb{R}^{m \times n}$ is a known matrix, and X, Y and N are random variables, and if X and N are mutually independent Gaussian variables with probability densities

$$\begin{aligned} \pi_{pr}(x) &\propto \exp\left(-\frac{1}{2}(x - \overline{x}_{pr})^T \Gamma_{pr}^{-1}(x - \overline{x}_{pr})\right), \\ \pi_n(n) &\propto \exp\left(-\frac{1}{2}(n - \overline{n})^T \Gamma_{n}^{-1}(n - \overline{n})\right), \end{aligned}$$

then the resulting posterior density is Gaussian

$$\pi_{post}(x \mid y) \propto \exp\left(-\frac{1}{2}(x - \overline{x}_{post})^T \Gamma_{post}^{-1}(x - \overline{x}_{post})\right),$$

Gaussian priors

Theorem 1 continued

whose posterior mean and covariance matrix are given by

$$\begin{aligned}\overline{x}_{post} &= (\Gamma_{pr}^{-1} + A^T\Gamma_n^{-1}A)^{-1}(A^T\Gamma_n^{-1}(y-\overline{n}) + \Gamma_{pr}^{-1}\overline{x}_{pr}) \\ &= \overline{x}_{pr} + \Gamma_{pr}A^T(A\Gamma_{pr}A^T + \Gamma_n)^{-1}(y - A\overline{x}_{pr} - \overline{n}),\end{aligned}$$

$$\begin{aligned}\Gamma_{post} &= (\Gamma_{pr}^{-1} + A^T\Gamma_n^{-1}A)^{-1} \\ &= \Gamma_{pr} - \Gamma_{pr}A^T(A\Gamma_{pr}A^T + \Gamma_n)^{-1}A\Gamma_{pr}.\end{aligned}$$

Gaussian priors

Theorem 1 continued

Proof

The posterior density is given by

$$\begin{aligned}\pi(x \mid y) &\propto \pi_{pr}(x)\pi(y \mid x) = \pi_{pr}(x)\pi_n(y - Ax) \\ &\propto \exp\Big(-\frac{1}{2}(x - \overline{x}_{pr})^T \Gamma_{pr}^{-1}(x - \overline{x}_{pr}) \\ &\qquad -\frac{1}{2}(y - Ax - \overline{n})^T \Gamma_n^{-1}(y - Ax - \overline{n})\Big).\end{aligned}$$

This clearly is a Gaussian density, since $Q(x \mid y)$ in the argument of the exponent function $\pi(x \mid y) \propto \exp(-\frac{1}{2}Q(x \mid y))$ is a quadratic form. The posterior density can be normalized, since the prior $\pi_{pr}(x)$ can be normalized even though the $\pi(y \mid x)$ can be non-normalizable due to the possible non-triviality of the null-space of A.

Gaussian priors

Theorem 1 continued

The covariance matrix can be obtained from the highest order term of the quadratic form

$$\begin{aligned} Q(x \mid y) &= (x-\overline{x}_{pr})^T \Gamma_{pr}^{-1}(x-\overline{x}_{pr}) + (y-Ax-\overline{n})^T \Gamma_n^{-1}(y-Ax-\overline{n}) \\ &= x^T(\Gamma_{pr}^{-1} + A^T \Gamma_n^{-1} A)^{-1} x + \cdots , \end{aligned}$$

giving $\Gamma_{post} = (\Gamma_{pr}^{-1} + A^T \Gamma_n^{-1} A)^{-1}$. The second form given follows from the Sherman-Morrison-Woodbury formula stating that, for any invertible matrices B and C, it holds

$$(B + UCV)^{-1} = B^{-1} - B^{-1} U \left(C^{-1} + V B^{-1} U \right)^{-1} V B^{-1}.$$

(Direct product proves this. Try!)

Gaussian priors

Theorem 1 continued

Choosing $B = \Gamma_{pr}^{-1}$, $C = \Gamma_n^{-1}$ and $V = U^T = A$ yields

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

The mean can be found through $\widetilde{Q}(\widetilde{x} \mid \widetilde{y}) = Q(x \mid y)$ in which

$$\begin{aligned} \widetilde{x} &= W_{pr}(x - \overline{x}_{pr}) \\ \widetilde{y} &= W_n(y - A\overline{x}_{pr} - \overline{n}) \\ \widetilde{A} &= W_n A W_{pr}^{-1} \end{aligned}$$

with $W_{pr} = \sqrt{\Gamma_{pr}^{-1}}$ and $W_n = \sqrt{\Gamma_n^{-1}}$. (If the eigenvalue decomposition of a symmetric and positive definite matrix B is given by $B = VDV^T$, then $\sqrt{B} = V\sqrt{D}V^T$.) We have:

Gaussian priors

Theorem 1 continued

$$Q(x \mid y) = \widetilde{Q}(\widetilde{x} \mid \widetilde{y}) = (\widetilde{y} - \widetilde{A}\widetilde{x})^T(\widetilde{y} - \widetilde{A}\widetilde{x}) + \widetilde{x}^T\widetilde{x},$$

which is minimized by $\widetilde{x}_{MAP} = W_{pr}(x_{MAP} - \overline{x}_{pr})$, that is, $x_{MAP} = \overline{x}_{pr} + W_{pr}^{-1}\widetilde{x}_{MAP}$. For a Gaussian density maximum coincides with the mean, i.e., $\overline{x}_{post} = x_{MAP}$. Hence, based on Examples 3 and 5, we have

$$\begin{aligned}
\overline{x}_{post} = x_{MAP} &= \overline{x}_{pr} + W_{pr}^{-1}(\widetilde{A}^T\widetilde{A} + I)^{-1}\widetilde{A}^T\widetilde{y} \\
&= \overline{x}_{pr} + W_{pr}^{-1}(W_{pr}^{-1}A^T W_n^2 A W_{pr}^{-1} + I)^{-1}(W_{pr}{}^{-1}A^T W_n)W_n(y - A\overline{x}_{pr} - \overline{n}) \\
&= \overline{x}_{pr} + (W_{pr}W_{pr}^{-1}A^T W_n^2 A W_{pr}^{-1} W_{pr} + W_{pr}^2)^{-1}A^T(W_n^2)(y - A\overline{x}_{pr} - \overline{n})
\end{aligned}$$

Gaussian priors

Theorem 1 continued

$$=\overline{x}_{pr} + (A^T\Gamma_n^{-1}A + \Gamma_{pr}^{-1})^{-1}A^T\Gamma_n^{-1}(y - A\overline{x}_{pr} - \overline{n})$$
$$=\overline{x}_{pr}+(A^T\Gamma_n^{-1}A+\Gamma_{pr}^{-1})^{-1}(A^T\Gamma_n^{-1}(y-\overline{n})-(A^T\Gamma_n^{-1}A+\Gamma_{pr}^{-1})\overline{x}_{pr}+\Gamma_{pr}^{-1}\overline{x}_{pr})$$
$$=\overline{x}_{pr} - \overline{x}_{pr} + (\Gamma_{pr}^{-1} + A^T\Gamma_n^{-1}A)^{-1}(A^T\Gamma_n^{-1}(y - \overline{n}) + \Gamma_{pr}^{-1}\overline{x}_{pr})$$
$$=(\Gamma_{pr}^{-1} + A^T\Gamma_n^{-1}A)^{-1}(A^T\Gamma_n^{-1}(y - \overline{n}) + \Gamma_{pr}^{-1}\overline{x}_{pr}) = \overline{x}_{post}.$$

It holds that $(A^T\Gamma_n^{-1}A+\Gamma_{pr}^{-1})^{-1}A^T\Gamma_n^{-1}=\Gamma_{pr}A^T(A\Gamma_{pr}A^T+\Gamma_n)^{-1}$. Namely, we have:

$$(A^T\Gamma_n^{-1}A+\Gamma_{pr}^{-1})^{-1}A^T\Gamma_n^{-1} = (W_{pr}\widetilde{A}^T\widetilde{A}W_{pr} + W_{pr}^2)^{-1}W_{pr}\widetilde{A}^TW_n$$

Gaussian priors

Theorem 1 continued

$$
\begin{aligned}
&= W_{pr}^{-1}(\widetilde{A}^T\widetilde{A} + I)^{-1} W_{pr}{}^{-1} W_{pr}\widetilde{A}^T W_n \\
&= W_{pr}^{-1}(\widetilde{A}^T\widetilde{A} + I)^{-1}\widetilde{A}^T W_n = W_{pr}^{-1}\widetilde{A}^T(\widetilde{A}\widetilde{A}^T + I)^{-1} W_n \\
&= W_{pr}^{-2} A^T W_n (W_n A W_{pr}^{-2} A^T W_n + I)^{-1} W_n \\
&= \Gamma_{pr} A^T (W_n^{-1} W_n A \Gamma_{pr} A^T W_n W_n{}^{-1} + W_n{}^{-2})^{-1} \\
&= \Gamma_{pr} A^T (A\Gamma_{pr} A^T + \Gamma_n)^{-1}.
\end{aligned}
$$

Above, equation $(\widetilde{A}^T\widetilde{A} + I)^{-1}\widetilde{A}^T = \widetilde{A}^T(\widetilde{A}\widetilde{A}^T + I)^{-1}$ follows from the singular value decomposition.

Gaussian priors

Theorem 1 continued

Thus, we have

$$(A^T\Gamma_n^{-1}A+\Gamma_{pr}^{-1})^{-1}A^T\Gamma_n^{-1} = \Gamma_{pr}A^T(A\Gamma_{pr}A^T + \Gamma_n)^{-1}$$

and, consequently, it holds

$$\begin{aligned}\overline{x}_{post} &= \overline{x}_{pr} + (A^T\Gamma_n^{-1}A + \Gamma_{pr}^{-1})^{-1}A^T\Gamma_n^{-1}(y - A\overline{x}_{pr} - \overline{n}) \\ &= \overline{x}_{pr} + \Gamma_{pr}A^T(A\Gamma_{pr}A^T + \Gamma_n)^{-1}(y - A\overline{x}_{pr} - \overline{n}).\end{aligned}$$

Examples of non-Gaussian priors

- Frequently used non-Gaussian priors include, for example,
  1. ℓ^1-prior (impulse prior) of the form $\pi_{pr}(x) \propto \exp(-\alpha \sum_{i=1}^{n} |x_i|)$, which can be used to reconstruct well-localized objects,
  2. total variation (TV) prior $\pi_{pr}(f) \propto \exp\left(-\alpha \int_\Omega |\nabla f| \, d\Omega \right)$, beneficial in reconstructing images with sharp edges,
  3. hierarchical prior $\pi_{pr}(x) = \int \pi_{pr}(x \mid \gamma) \pi_{hyper}(\gamma) \, d\gamma$, in which $\pi_{pr}(x \mid \gamma)$ is conditional on hyperparameter γ with the prior (hyperprior) $\pi_{hyper}(\gamma)$.
  4. prior constraints limiting the unknown, e.g., to a relatively small region of interest (ROI).

Examples of non-Gaussian priors

- Gaussian priors usually lead to more or less smooth reconstructions.
- If sharp structures need to be produced, one can use, e.g., ℓ^1-prior, if the goal is to find focal objects or impulses.
- Total variation (TV) prior can be used if the goal is to reconstruct edge structures, for example, to preserve the edges of a noisy image. With a high value of parameter α, TV-prior it is likely to produce a cartoon-like effect.
- Hierarchical priors can be advantageous, e.g., when the prior information can be classified to different hierarchy levels, e.g., schools, classes, students.
- Using a region of interest (ROI) can be useful, for instance, to speed up exploration of the posterior density.