

We now focus on the following time-dependent PDE:

$$\mathcal{L}u(x, t) := \partial_t u(x, t) - \nabla \cdot (a(x) \nabla u(x, t)) = f(x, t), \quad x \in \Omega, \quad t \in [0, T] \quad (1)$$

$$u(x, t) = 0, \quad x \in \partial\Omega, \quad t \in [0, T] \quad (2)$$

$$u(x, 0) = u_0(x), \quad x \in \Omega \quad (3)$$

We will now set up a prior on the solution u to the above problem. To do so we first let $v_h \in S_h$ be some approximation of the initial condition $u_0(x)$ in the FEM space S_h . To be more specific we will assume that $v_h(x) = \Phi(x)^* \gamma := \sum_{i=1}^J \phi_i(x) \gamma_i$. Note that $\Phi(x) := (\phi_1(x), \dots, \phi_J(x))^T$. We take the prior on u to be:

$$u \sim \mathcal{N}(m_0, V_0) \quad (4)$$

where $m_0(x, t) := v_h(x) = \Phi^*(x) \gamma$ (m_0 is constant in time). The prior covariance operator V_0 is defined as follows:

$$(V_0 g)(x, t) = \int_{\Omega} \int_0^T \sum_{i=1}^J \lambda_i \phi_i(x) \phi_i(y) k(t, s) g(y, s) ds dy =: \int_{\Omega} \int_0^T k_{ys}^{xt} g(y, s) ds dy \quad (5)$$

where we have a general kernel $k(t, s)$ for time which will be taken to be a specific function later. We have also used the notation $k_{ys}^{xt} := \sum_{i=1}^J \lambda_i \phi_i(x) \phi_i(y) k(t, s)$ to make it clear which variables are held fixed and which we integrate against.

We now introduce the following operators $\mathcal{I}_s := (I_1(s), \dots, I_J(s))^T$ where:

$$I_i(s)g := \int_{\Omega} \phi_i(x) g(x, s) dx \quad (6)$$

We now introduce a uniform time grid:

$$t_n = n\delta, \quad n = 0, 1, \dots, N$$

where δ is the spacing between consecutive times and $N = \frac{T}{\delta}$ (assume that N is an integer).

To move from $t = t_0 = 0$ to $t = t_1 = \delta$ we condition on observing $\mathcal{I}_{\delta} \mathcal{L}u = \mathcal{I}_{\delta} f =: F^1$. Let $\tilde{A}_{\delta} := \mathcal{I}_{\delta} \mathcal{L}$. For a fixed realisation of f (and so of F^1) we thus seek the following conditional distribution:

$$u | \{\tilde{A}_{\delta} u = F^1, f\} \sim \mathcal{N}(m_1, V_1) \quad (7)$$

That this distribution is itself Gaussian follows from considering the following joint distribution:

$$\begin{pmatrix} u \\ \tilde{A}_{\delta} u \end{pmatrix} = \begin{pmatrix} I \\ \tilde{A}_{\delta} \end{pmatrix} u \sim \mathcal{N} \left(\begin{pmatrix} m_0 \\ \tilde{A}_{\delta} m_0 \end{pmatrix}, \begin{pmatrix} V_0 & V_0 \tilde{A}_{\delta}^* \\ \tilde{A}_{\delta} V_0 & \tilde{A}_{\delta} V_0 \tilde{A}_{\delta}^* \end{pmatrix} \right)$$

It follows that the conditional distribution is Gaussian and the mean and covariance are given by:

$$m_1 = m_0 + V_0 \tilde{A}_{\delta}^* (\tilde{A}_{\delta} V_0 \tilde{A}_{\delta}^*)^{-1} (F^1 - \tilde{A}_{\delta} m_0) \quad (8)$$

$$V_1 = V_0 - V_0 \tilde{A}_{\delta}^* (\tilde{A}_{\delta} V_0 \tilde{A}_{\delta}^*)^{-1} \tilde{A}_{\delta} V_0 \quad (9)$$

We now rewrite the mean update equation as follows:

$$m_1 = \left(1 - V_0 \tilde{A}_{\delta}^* (\tilde{A}_{\delta} V_0 \tilde{A}_{\delta}^*)^{-1} \tilde{A}_{\delta} \right) m_0 + V_0 \tilde{A}_{\delta}^* (\tilde{A}_{\delta} V_0 \tilde{A}_{\delta}^*)^{-1} F^1 \quad (10)$$

Written in this form this update equation can now be more easily compared to the backward-Euler Galerkin method update rule. This method involves the following approximations: $U^n \approx u(t_n)$ and $U^n(x) = \Phi(x)^* \alpha^n$. The update rule for the vector of coefficients α^n is given by:

$$\alpha^n = (M + \delta A)^{-1} M \alpha^{n-1} + \delta (M + \delta A)^{-1} \mathbf{b}^n \quad (11)$$

where $\mathbf{b}^n = \mathcal{I}_{t_n} f = F^n$. In order to compare this to our mean update rule we now project this into S_h by premultiplying by Φ^* :

$$\Phi^* \alpha^n = \Phi^* (M + \delta A)^{-1} M \alpha^{n-1} + \delta \Phi^* (M + \delta A)^{-1} \mathbf{b}^n \quad (12)$$

In our mean update rule m_0 plays the role of $\Phi^* \alpha^0$ and m_1 plays the role of $\Phi^* \alpha^1$. In fact, we have $m_0 = \Phi^* \gamma$ and so we can consider $\alpha^0 = \gamma$. This is exactly the initial condition for the coefficient vector in the backward-Euler Galerkin method. Comparing (12) with (10) we thus see that we would like to be able to show:

$$\Phi^*(M + \delta A)^{-1} M = \left(1 - V_0 \tilde{A}_\delta^* (\tilde{A}_\delta V_0 \tilde{A}_\delta^*)^{-1} \tilde{A}_\delta\right) \Phi^* \quad (13)$$

$$\delta \Phi^*(M + \delta A)^{-1} = V_0 \tilde{A}_\delta^* (\tilde{A}_\delta V_0 \tilde{A}_\delta^*)^{-1} \quad (14)$$

To make progress we must now start computing various terms needed for our mean and covariance update rules. We start with $V_0 \tilde{A}_\delta^*$. We have:

$$V_0 \tilde{A}_\delta^* = V_0 \mathcal{L}^*(I_1(\delta)^*, \dots, I_J(\delta)^*)$$

We can thus see that we need to be able to compute terms of form $V_0 \mathcal{L}^* I_i^*(\delta) = V_0(I_i(\delta) \mathcal{L})^*$. Now since the operator $I_i(\delta) \mathcal{L}$ takes in a function on $\Omega \times [0, T]$ and outputs a real number its adjoint should take in a real number and output a function on $\Omega \times [0, T]$. This adjoint should satisfy the following relation:

$$\alpha((I_i(\delta) \mathcal{L})g) = \int_\Omega \int_0^T ((I_i(\delta) \mathcal{L})^* \alpha)(x, t) g(x, t) dt dx \quad \forall g, \quad \forall \alpha \in \mathbb{R} \quad (15)$$

Using this we can now compute:

$$\begin{aligned} (V_0(I_i(\delta) \mathcal{L})^* \alpha)(x, s) &= \int_\Omega \int_0^T k_{yw}^{xs} ((I_i(\delta) \mathcal{L})^* \alpha)(y, w) dw dy \\ &= \alpha(I_i(\delta)(\mathcal{L}k^{xs})) \\ &= \alpha \int_\Omega \phi_i(y)(\mathcal{L}k^{xs})(y, \delta) dy \end{aligned}$$

We now work out $(\mathcal{L}k^{xs})(y, \delta)$ taking care to remember that x, s are fixed and so \mathcal{L} acts on the variables y, δ :

$$\begin{aligned} (\mathcal{L}k^{xs})(y, \delta) &= \partial_2 k(s, \delta) \sum_{j=1}^J \lambda_j \phi_j(x) \phi_j(y) - k(s, \delta) \nabla_y \cdot \left(a(y) \nabla_y \sum_{j=1}^J (\lambda_j \phi_j(x) \phi_j(y)) \right) \\ &= \partial_2 k(s, \delta) \sum_{j=1}^J \lambda_j \phi_j(x) \phi_j(y) - k(s, \delta) \sum_{j=1}^J \lambda_j \phi_j(x) \nabla_y \cdot (a(y) \nabla_y \phi_j(y)) \end{aligned}$$

So we can now compute:

$$\begin{aligned} (V_0(I_i(\delta) \mathcal{L})^* \alpha)(x, s) &= \alpha \int_\Omega \phi_i(y) \partial_2 k(s, \delta) \sum_{j=1}^J \lambda_j \phi_j(x) \phi_j(y) dy - \alpha \int_\Omega \phi_i(y) k(s, \delta) \sum_{j=1}^J \lambda_j \phi_j(x) \nabla_y \cdot (a(y) \nabla_y \phi_j(y)) dy \\ &= \alpha \partial_2 k(s, \delta) \sum_{j=1}^J \lambda_j \phi_j(x) M_{ji} + \alpha k(s, \delta) \sum_{j=1}^J \lambda_j \phi_j(x) A_{ji} \end{aligned}$$

where M is the Galerkin mass matrix and A is the Galerkin stiffness matrix, i.e. the matrices with entries given by:

$$M_{ij} := \int_\Omega \phi_i(x) \phi_j(x) dx \quad (16)$$

$$A_{ij} := \int_\Omega a(x) \nabla \phi_i(x) \nabla \phi_j(x) dx \quad (17)$$

Using this result we can deduce that:

$$(V_0 \tilde{A}_\delta^* \mathbf{v})(x, s) = \partial_2 k(s, \delta) \Phi(x)^* \Lambda M \mathbf{v} + k(s, \delta) \Phi(x)^* \Lambda A \mathbf{v} \quad (18)$$

for any $\mathbf{v} \in \mathbb{R}^J$, where $\Lambda = \text{diag}\{\lambda_i\}_{i=1}^J$.

We now move onto computing:

$$\begin{aligned}\tilde{A}_\delta V_0 \tilde{A}_\delta^* &= \mathcal{I}_\delta \mathcal{L} V_0 \mathcal{L}^* \mathcal{I}_\delta^* \\ &= \begin{pmatrix} I_1(\delta) \\ \vdots \\ I_J(\delta) \end{pmatrix} \mathcal{L} V_0 \mathcal{L}^* \begin{pmatrix} I_1(\delta)^* & \dots & I_J(\delta)^* \end{pmatrix}\end{aligned}$$

This operator has ij -th entry which is given by:

$$\begin{aligned}(\tilde{A}_\delta V_0 \tilde{A}_\delta^*)_{ij} \alpha &= I_i(\delta) \mathcal{L} V_0 \mathcal{L}^* I_j(\delta)^* \alpha \\ &= \int_{\Omega} \phi_i(x) [(\mathcal{L} V_0 (I_j(\delta) \mathcal{L})^* \alpha)(x, \delta)] dx \\ &= \int_{\Omega} \phi_i(x) \left[\alpha \partial_1 \partial_2 k(\delta, \delta) \sum_{l=1}^J \lambda_l \phi_l(x) M_{lj} + \alpha \partial_1 k(\delta, \delta) \sum_{l=1}^J \lambda_l \phi_l(x) A_{lj} \right. \\ &\quad \left. - \alpha \partial_2 k(\delta, \delta) \sum_{l=1}^J \lambda_l M_{lj} \nabla \cdot (a(x) \nabla \phi_l(x)) - \alpha k(\delta, \delta) \sum_{l=1}^J \lambda_l A_{lj} \nabla \cdot (a(x) \nabla \phi_l(x)) \right] dx \\ &= \alpha \partial_1 \partial_2 k(\delta, \delta) \sum_{l=1}^J \lambda_l M_{il} M_{lj} + \alpha \partial_1 k(\delta, \delta) \sum_{l=1}^J \lambda_l M_{il} A_{lj} \\ &\quad + \alpha \partial_2 k(\delta, \delta) \sum_{l=1}^J \lambda_l M_{lj} A_{il} + \alpha k(\delta, \delta) \sum_{l=1}^J \lambda_l A_{il} A_{lj} \\ &= \alpha \partial_1 \partial_2 k(\delta, \delta) (M \Lambda M)_{ij} + \alpha \partial_1 k(\delta, \delta) (M \Lambda A)_{ij} + \alpha \partial_2 k(\delta, \delta) (A \Lambda M)_{ij} + \alpha k(\delta, \delta) (A \Lambda A)_{ij}\end{aligned}$$

We can thus conclude that $\tilde{A}_\delta V_0 \tilde{A}_\delta^*$ is the $J \times J$ matrix given by:

$$\tilde{A}_\delta V_0 \tilde{A}_\delta^* = \partial_1 \partial_2 k(\delta, \delta) M \Lambda M + \partial_1 k(\delta, \delta) M \Lambda A + \partial_2 k(\delta, \delta) A \Lambda M + k(\delta, \delta) A \Lambda A \quad (19)$$

We now will choose a specific kernel $k(s, t)$ for the temporal part of our prior covariance. We will take $k(s, t) := st$. We thus have $\partial_1 k(s, t) = t$, $\partial_2 k(s, t) = s$, and $\partial_1 \partial_2 k(s, t) = 1$. So we have:

$$\begin{aligned}\tilde{A}_\delta V_0 \tilde{A}_\delta^* &= M \Lambda M + \delta M \Lambda A + \delta A \Lambda M + \delta^2 A \Lambda A \\ &= M \Lambda (M + \delta A) + \delta A \Lambda (M + \delta A) \\ &= (M \Lambda + \delta A \Lambda) (M + \delta A) \\ &= (M + \delta A) \Lambda (M + \delta A) = Q \Lambda Q\end{aligned}$$

where we have defined $Q := M + \delta A$. We can now finish computing m_1 . Evaluate it at $(x, t_1) = (x, \delta)$:

$$m_1(x, \delta) = ((1 - V_0 \tilde{A}_\delta^* (\tilde{A}_\delta V_0 \tilde{A}_\delta^*)^{-1} \tilde{A}_\delta) m_0)(x, \delta) + (V_0 \tilde{A}_\delta^* (\tilde{A}_\delta V_0 \tilde{A}_\delta^*)^{-1} F^1)(x, \delta)$$

We can now apply (18) to compute the last term:

$$\begin{aligned}(V_0 \tilde{A}_\delta^* (\tilde{A}_\delta V_0 \tilde{A}_\delta^*)^{-1} F^1)(x, \delta) &= (V_0 \tilde{A}_\delta^* (Q \Lambda Q)^{-1} F^1)(x, \delta) \\ &= \partial_2 k(\delta, \delta) \Phi(x)^* \Lambda M (Q \Lambda Q)^{-1} F^1 + k(\delta, \delta) \Phi(x)^* \Lambda A (Q \Lambda Q)^{-1} F^1 \\ &= \delta \Phi(x)^* \Lambda M (Q \Lambda Q)^{-1} F^1 + \delta^2 \Phi(x)^* \Lambda A (Q \Lambda Q)^{-1} F^1 \\ &= \delta \Phi(x)^* \Lambda (M + \delta A) (Q \Lambda Q)^{-1} F^1 \\ &= \delta \Phi(x)^* \Lambda Q Q^{-1} \Lambda^{-1} Q^{-1} F^1 \\ &= \delta \Phi(x)^* Q^{-1} F^1\end{aligned}$$

In order to apply (18) to compute the second term involving m_0 in the update rule we must first compute $\tilde{A}_\delta m_0 = \mathcal{I}_\delta \mathcal{L} m_0$. To do this we compute:

$$\begin{aligned}(\mathcal{L} m_0)(x, t) &= \partial_t m_0 - \nabla \cdot (a(x) \nabla m_0(x, t)) \\ &= -\nabla(a(x) \nabla \Phi(x)^* \gamma) \\ &= -\sum_{j=1}^J \gamma_j \nabla \cdot (a(x) \nabla \phi_j(x))\end{aligned}$$

Thus, the i -th entry of $\tilde{A}_\delta m_0$ can be computed as:

$$\begin{aligned} (\tilde{A}_\delta m_0)_i &= I_i(\delta) \mathcal{L} m_0 \\ &= \int_{\Omega} \phi_i(x) \left(- \sum_{j=1}^J \gamma_j \nabla \cdot (a(x) \nabla \phi_j(x)) \right) dx \\ &= - \sum_{j=1}^J \gamma_j A_{ij} = (A\gamma)_i \end{aligned}$$

So $\tilde{A}_\delta m_0 = A\gamma$. We can now apply (18) to compute the second term involving m_0 as follows:

$$\begin{aligned} ((V_0 \tilde{A}_\delta^* (\tilde{A}_\delta V_0 \tilde{A}_\delta^*)^{-1} \tilde{A}_\delta) m_0)(x, \delta) &= (V_0 \tilde{A}_\delta^* (Q\Lambda Q)^{-1} A\gamma)(x, \delta) \\ &= \partial_2 k(\delta, \delta) \Phi(x)^* \Lambda M (Q\Lambda Q)^{-1} A\gamma + k(\delta, \delta) \Phi(x)^* \Lambda A (Q\Lambda Q)^{-1} A\gamma \\ &= \delta \Phi(x)^* \Lambda M (Q\Lambda Q)^{-1} A\gamma + \delta^2 \Phi(x)^* \Lambda A (Q\Lambda Q)^{-1} A\gamma \\ &= \delta \Phi(x)^* \Lambda (M + \delta A) (Q\Lambda Q)^{-1} A\gamma \\ &= \delta \Phi(x)^* \Lambda Q Q^{-1} \Lambda^{-1} Q^{-1} A\gamma \\ &= \delta \Phi(x)^* Q^{-1} A\gamma \end{aligned}$$

Thus we can compute:

$$\begin{aligned} (1 - (V_0 \tilde{A}_\delta^* (\tilde{A}_\delta V_0 \tilde{A}_\delta^*)^{-1} \tilde{A}_\delta) m_0)(x, \delta) &= \Phi(x)^* \gamma - \delta \Phi(x)^* Q^{-1} A\gamma \\ &= \Phi(x)^* (I - \delta Q^{-1} A) \gamma \\ &= \Phi(x)^* Q^{-1} (Q - \delta A) \gamma \\ &= \Phi(x)^* Q^{-1} (M + \delta A - \delta A) \gamma \\ &= \Phi(x)^* Q^{-1} M \gamma \end{aligned}$$

Putting this all together we obtain:

$$\begin{aligned} m_1(x, t_1) &= m_1(x, \delta) = \Phi(x)^* Q^{-1} M \gamma + \delta \Phi(x)^* Q^{-1} F^1 \\ &= \Phi(x)^* (Q^{-1} M \gamma + \delta Q^{-1} F^1) \\ &= \Phi(x)^* ((M + \delta A)^{-1} M \gamma + \delta (M + \delta A)^{-1} F^1) \end{aligned}$$

Thus we see that performing this mean update and then evaluating at the time $t_1 = \delta$ we obtain that the coefficients of the $\{\phi_i(x)\}_{i=1}^J$ changes as follows:

$$\gamma \mapsto (M + \delta A)^{-1} M \gamma + \delta (M + \delta A)^{-1} F^1 \quad (20)$$

just like in the update equation for the backward-Euler Galerkin method.

Remark. Note that evaluating m_1 at (x, s) instead of at (x, δ) yields the following:

$$m_1(x, s) = \Phi(x)^* [(M + \delta A)^{-1} (M + (\delta - s)A) \gamma + s(M + \delta A)^{-1} F^1] \quad (21)$$

We can now move on to computing the covariance V_1 .