We now focus on the following time-dependent PDE:

$$\mathcal{L}u(x,t) := \partial_t u(x,t) - \nabla\cdot(a(x)\nabla u(x,t)) = f(x,t), \quad x\in\Omega,\ t\in[0,T] \tag{1}$$
$$u(x,t) = 0, \qquad x\in\partial\Omega,\ t\in[0,T] \tag{2}$$
$$u(x,0) = u_0(x), \qquad x\in\Omega \tag{3}$$

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 $v_h(x) = \Phi(x)^*\boldsymbol{\gamma} := \sum_{i=1}^J \phi_i(x)\gamma_i$. Note that $\Phi(x) := (\phi_1(x),\dots,\phi_J(x))^T$. We take the prior on u to be:

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

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

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

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:

$$I_i(s)g := \int_\Omega \phi_i(x)g(x,s)\mathrm{d}x \tag{6}$$

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:

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

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

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

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

$$m_1 = m_0 + V_0\tilde{A}_\delta^*(\tilde{A}_\delta V_0\tilde{A}_\delta^*)^{-1}(F^1 - \tilde{A}_\delta m_0) \tag{8}$$
$$V_1 = V_0 - V_0\tilde{A}_\delta^*(\tilde{A}_\delta V_0\tilde{A}_\delta^*)^{-1}\tilde{A}_\delta V_0 \tag{9}$$

We now rewrite the mean update equation as follows:

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

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 $U^n(x) = \Phi(x)^*\boldsymbol{\alpha}^n$. The update rule for the vector of coefficients $\boldsymbol{\alpha}^n$ is given by:

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

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^*:

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

In our mean update rule m_0 plays the role of $\Phi^*\boldsymbol{\alpha}^0$ and m_1 plays the role of $\Phi^*\alpha^1$. In fact, we have $m_0 = \Phi^*\boldsymbol{\gamma}$ and so we can consider $\boldsymbol{\alpha}^0 = \boldsymbol{\gamma}$. 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:

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

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

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:

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

Using this we can now compute:

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

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:

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

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

$$M_{ij} := \int_\Omega \phi_i(x)\phi_j(x)\mathrm{d}x \tag{16}$$

$$A_{ij} := \int_\Omega a(x)\nabla\phi_i(x)\nabla\phi_j(x)\mathrm{d}x \tag{17}$$

Using this result we can deduce that:

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

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

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:

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

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

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

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:

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

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

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

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

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

Thus we can compute:

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

Putting this all together we obtain:

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

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:

$$\boldsymbol{\gamma} \longmapsto (M+\delta A)^{-1}M\boldsymbol{\gamma} + \delta(M+\delta A)^{-1}F^1 \tag{20}$$

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:

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

We can now move on to computing the covariance V_1. We start by computing $\tilde{A}_\delta V_0$. Computing this involves determining how $I_j(\delta)\mathcal{L}V_0$ acts on functions g for $j = 1,\ldots,J$. We have:

$$I_j(\delta)\mathcal{L}V_0 g = \int_\Omega \phi_j(x)(\mathcal{L}V_0 g)(x,\delta)\mathrm{d}x$$

Now recalling that $V_0 g(x,\delta) = \int_\Omega\int_0^T k_{ys}^{x\delta} g(y,s)\mathrm{d}s\mathrm{d}y$ we deduce:

$$\begin{aligned}(\mathcal{L}V_0 g)(x,\delta) &= \int_\Omega\int_0^T (\mathcal{L}k_{ys})(x,\delta)g(y,s)\mathrm{d}s\mathrm{d}y\\ &= \int_\Omega\int_0^T\left(\partial_1 k(\delta,s)\sum_{i=1}^J\lambda_i\phi_i(x)\phi_i(y) - k(\delta,s)\sum_{i=1}^J\lambda_i\nabla_x\cdot(a(x)\nabla_x\phi_i(x))\phi_i(y)\right)g(y,s)\mathrm{d}s\mathrm{d}y\end{aligned}$$

We can now perform the integration to obtain:

$$
\begin{aligned}
I_j(\delta)\mathcal{L}V_0 g &= \int_\Omega \phi_j(x)\left(\int_\Omega\int_0^T\left(\partial_1 k(\delta,s)\sum_{i=1}^J \lambda_i\phi_i(x)\phi_i(y) - k(\delta,s)\sum_{i=1}^J \lambda_i\nabla_x\cdot(a(x)\nabla_x\phi_i(x))\phi_i(y)\right)g(y,s)\mathrm{d}s\mathrm{d}y\right)\mathrm{d}x \\
&= \int_\Omega\int_0^T\left(\partial_1 k(\delta,s)\sum_{i=1}^J\lambda_i M_{ij}\phi_i(y)g(y,s) + k(\delta,s)\sum_{i=1}^J\lambda_i A_{ij}\phi_i(y)g(y,s)\right)\mathrm{d}s\mathrm{d}y \\
&= \sum_{i=1}^J\lambda_i\int_\Omega\int_0^T(\partial_1 k(\delta,s)M_{ij} + k(\delta,s)A_{ij})\phi_i(y)g(y,s)\mathrm{d}s\mathrm{d}y \\
&= \sum_{i=1}^J\lambda_i\left[\int_0^T M_{ij}\partial_1 k(\delta,s)(I_i(s)g)\mathrm{d}s + \int_0^T A_{ij}k(\delta,s)(I_i(s)g)\mathrm{d}s\right] \\
&= \sum_{i=1}^J\lambda_i\left[M_{ij}\left(\int_0^T\partial_1 k(\delta,s)(\mathcal{I}_s g)\mathrm{d}s\right)_i + A_{ij}\left(\int_0^T k(\delta,s)(\mathcal{I}_s g)\mathrm{d}s\right)_i\right] \\
&= \left(M\Lambda\int_0^T\partial_1 k(\delta,s)(\mathcal{I}_s g)\mathrm{d}s + A\Lambda\int_0^T k(\delta,s)(\mathcal{I}_s g)\mathrm{d}s\right)_j
\end{aligned}
$$

Thus we can deduce:

$$
\tilde{A}_\delta V_0 g = M\Lambda\int_0^T\partial_1 k(\delta,s)(\mathcal{I}_s g)\mathrm{d}s + A\Lambda\int_0^T k(\delta,s)(\mathcal{I}_s g)\mathrm{d}s \tag{22}
$$

Having worked this out we can now compute the second term in the formula for V_1 by utilising (18) as follows:

$$
\begin{aligned}
(V_0\tilde{A}_\delta^*(\tilde{A}_\delta V_o\tilde{A}_\delta^*)^{-1}\tilde{A}_\delta V_0 g)(x,t) &= \partial_2 k(t,\delta)\Phi(x)^*\Lambda M(Q\Lambda Q)^{-1}\left[M\Lambda\int_0^T\partial_1 k(\delta,s)(\mathcal{I}_s g)\mathrm{d}s + A\Lambda\int_0^T k(\delta,s)(\mathcal{I}_s g)\mathrm{d}s\right] \\
&\quad + k(t,\delta)\Phi(x)^*\Lambda A(Q\Lambda Q)^{-1}\left[M\Lambda\int_0^T\partial_1 k(\delta,s)(\mathcal{I}_s g)\mathrm{d}s + A\Lambda\int_0^T k(\delta,s)(\mathcal{I}_s g)\mathrm{d}s\right]
\end{aligned}
$$

To proceed we now utilise the specific choice of $k(s,t)$ to work out:

$$
\begin{aligned}
\int_0^T k(\delta,s)(\mathcal{I}_s g)\mathrm{d}s &= \delta\int_0^T s(\mathcal{I}_s g)\mathrm{d}s = \delta\boldsymbol{\nu}_g \\
\int_0^T\partial_1 k(\delta,s)(\mathcal{I}_s g) &= \int_0^T s(\mathcal{I}_s g)\mathrm{d}s = \boldsymbol{\nu}_g
\end{aligned}
$$

where $\boldsymbol{\nu}g := \int_0^T s(\mathcal{I}_s g)\mathrm{d}s$. Thus we have:

$$
\begin{aligned}
M\Lambda\int_0^T\partial_1 k(\delta,s)(\mathcal{I}_s g)\mathrm{d}s + A\Lambda\int_0^T k(\delta,s)(\mathcal{I}_s g)\mathrm{d}s &= M\Lambda\boldsymbol{\nu}_g + \delta A\Lambda\boldsymbol{\nu}_g \\
&= (M+\delta A)\Lambda\boldsymbol{\nu}_g \\
&= Q\Lambda\boldsymbol{\nu}_g
\end{aligned}
$$

We can now finish the computation of the second term of V_1:

$$\begin{aligned}
(V_0\tilde{A}_\delta^*(\tilde{A}_\delta V_o\tilde{A}_\delta^*)^{-1}\tilde{A}_\delta V_0 g)(x,t) &= t\Phi(x)^*\Lambda M(Q\Lambda Q)^{-1}Q\Lambda\boldsymbol{\nu}_g + t\delta\Phi(x)^*\Lambda A(Q\Lambda Q)^{-1}Q\Lambda\boldsymbol{\nu}_g \\
&= t\Phi(x)^*\Lambda(M+\delta A)(Q\Lambda Q)^{-1}Q\Lambda\boldsymbol{\nu}_g \\
&= t\Phi(x)^*\Lambda Q Q^{-1}\Lambda^{-1}Q^{-1}Q\Lambda\boldsymbol{\nu}_g \\
&= t\Phi(x)^*\Lambda\boldsymbol{\nu}_g \\
&= \sum_{i=1}^{J} t\lambda_i\phi_i(x)(\boldsymbol{\nu}_g)_i \\
&= \sum_{i=1}^{J} \lambda_i t\phi_i(x)\int_0^T s(I_i(s)g)\mathrm{d}s \\
&= \int_\Omega\int_0^T\sum_{i=1}^{J}\lambda_i\phi_i(x)\phi_i(y)tsg(y,s)\mathrm{d}s\mathrm{d}y \\
&= \int_\Omega\int_0^T\sum_{i=1}^{J}\lambda_i\phi_i(x)\phi_i(y)k(t,s)g(y,s)\mathrm{d}s\mathrm{d}y = (V_0 g)(x,t)
\end{aligned}$$

Thus, we can conclude: $V_0\tilde{A}_\delta^*(\tilde{A}_\delta V_o\tilde{A}_\delta^*)^{-1}\tilde{A}_\delta V_0 = V_0$ and so:

$$V_1 = V_0 - V_0\tilde{A}_\delta^*(\tilde{A}_\delta V_o\tilde{A}_\delta^*)^{-1}\tilde{A}_\delta V_0 = 0 \tag{23}$$

Thus, for a fixed realisation of f we have:

$$u|\{\tilde{A}_\delta u = F^1, f\} \sim \mathcal{N}(m_1, V_1) \tag{24}$$

where the function m_1 is given by:

$$m_1(x,s) = \Phi(x)^*[(M+\delta A)^{-1}(M+(\delta-s)A)\boldsymbol{\gamma} + s(M+\delta A)^{-1}F^1] \tag{25}$$

and where:

$$V_1 = 0$$

Thus if we evaluate the posterior (24) at the time $t_1 = \delta$ we will obtain a point mass at the spatial function $m_1(x,\delta) = \Phi(x)^*[(M+\delta A)^{-1}M\boldsymbol{\gamma} + \delta(M+\delta A)^{-1}F^1]$ just like the backward-Euler Galerkin method yields.

We now have the following distribution (for a fixed realisation of f):

$$u|\{\tilde{A}_\delta u = F^1, f\} \sim \mathcal{N}(m_1, 0) = \delta_{m_1} \tag{26}$$

We will now marginalise (26) over f in order to obtain the averaged conditional distribution. To do this we will need the following Lemma (which we prove below):

Lemma 0.1. Let $f \sim \mathcal{N}(\bar{f}, K)$ where we assume that this Gaussian measure is on a Hilbert space of functions $\mathcal{H}_1 \subset \mathbb{R}^{\mathcal{X}}$. Suppose that for a fixed realisation of f we have

$$y|f \sim \mathcal{N}(Lf + c, 0) = \delta_{Lf}$$

where L is a bounded linear operator from $\mathcal{H}_1$ to another Hilbert space $\mathcal{H}_2 \subset \mathbb{R}^{\mathcal{X}}$ (so y lies in $\mathcal{H}_2$) and where c is a deterministic function. Then marginalizing over f yields:

$$y \sim \mathcal{N}(L\bar{f} + c, LKL^*)$$

as the averaged distribution of y.

Proof: The fact that $y|f \sim \mathcal{N}(Lf + c, 0)$ is equivalent to saying that:

$$y = Lf + c + \tilde{y}$$

where $\tilde{y} \sim \mathcal{N}(0,0)(=\delta_0)$ is independent of f. Thus we have:

$$\begin{pmatrix} f \\ \tilde{y} \end{pmatrix} = \mathcal{N}\left(\begin{pmatrix} \bar{f} \\ 0 \end{pmatrix}, \begin{pmatrix} K & 0 \\ 0 & 0 \end{pmatrix}\right)$$

Since we can write:

$$y = \begin{pmatrix} L & 1 \end{pmatrix} \begin{pmatrix} f \\ \tilde{y} \end{pmatrix} + c$$

we deduce:

$$y \sim \mathcal{N}\left(L\bar{f} + c, \begin{pmatrix} L & 1 \end{pmatrix} \begin{pmatrix} K & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} L^* \\ 1 \end{pmatrix}\right) = \mathcal{N}(L\bar{f} + c, LKL^*)$$

as required. ■

We can now perform the marginalisation over f in (26). We first rewrite m_1 as follows:

$$\begin{aligned} m_1(x,s) &= \Phi(x)^*[Q^{-1}(M + (\delta - s)A)\boldsymbol{\gamma} + sQ^{-1}F^1] \\ &= \Phi(x)^*Q^{-1}(M + (\delta - s)A)\boldsymbol{\gamma} + s\Phi(x)^*Q^{-1}F^1 \\ &= c(x,s) + s\Phi(x)^*Q^{-1}\mathcal{I}_\delta f \end{aligned}$$

where $c(x,s) := \Phi(x)^*Q^{-1}(M + (\delta - s)A)\boldsymbol{\gamma}$ Thus, we have:

$$m_1 = \Psi^*Q^{-1}\mathcal{I}_\delta f + c = Lf + c$$

where $\Psi^* : \mathbb{R}^J \to L^2(\Omega \times [0,T])$ is the linear operator given by $(\Psi^*\boldsymbol{v})(x,s) := s\Phi(x)^*\boldsymbol{v}$ and $L := \Psi^*Q^{-1}\mathcal{I}_\delta$. Now both $\mathcal{I}_\delta$ and Q^{-1} are bounded linear operators. L will thus be a bounded linear operator provided Ψ^* is bounded. We check this:

$$\begin{aligned} \|\Psi^*\boldsymbol{v}\|^2 &= \int_\Omega \int_0^T |(\Psi^*\boldsymbol{v})(y,s)|^2 \mathrm{d}s\mathrm{d}y \\ &= \int_\Omega \int_0^T s^2 |\Phi(y)^*\boldsymbol{v}|^2 \mathrm{d}s\mathrm{d}y \\ &\leq T^2 \int_\Omega \left| \sum_{i=1}^J \phi_i(y) v_i \right|^2 \mathrm{d}y \\ &\leq T^2 \int_\Omega \sum_{i=1}^J |\phi_i(y)|^2 \sum_{j=1}^J |v_j|^2 \mathrm{d}y \\ &= T^2 \|\boldsymbol{v}\|^2 \sum_{i=1}^J \|\phi_i\|^2 \end{aligned}$$

and so $\|\Psi^*\boldsymbol{v}\| \leq T\left(\sum_{i=1}^J \|\phi_i\|^2\right)^{1/2} \|\boldsymbol{v}\|$. This implies: $\|\Psi^*\|_\infty \leq T\left(\sum_{i=1}^J \|\phi_i\|^2\right) < \infty$ and so Ψ^* is a bounded linear operator. We can thus use the Lemma given above to conclude that the marginalised distribution is:

$$\int u|\{\tilde{A}_\delta u = F^1, f\}\mathrm{d}f \sim \mathcal{N}(L\bar{f} + c, LKL^*) = \mathcal{N}(\bar{m}_1, \bar{V}_1)$$

where we have denoted $\bar{m}_1 := L\bar{f} + c$ and $\bar{V}_1 := LKL^*$. The mean is thus the following function:

$$\bar{m}_1(x,s) = \Phi(x)^*[Q^{-1}(M + (\delta - s)A)\boldsymbol{\gamma} + sQ^{-1}\bar{F}^1]$$

where $\bar{F}^1 := \mathcal{I}_\delta \bar{f}$. The covariance operator is given by:

$$\begin{aligned} \bar{V}_1 &= \Psi^*Q^{-1}\mathcal{I}_\delta K \mathcal{I}_\delta^* Q^{-1}\Psi \\ &= \Psi^* H^{(1)} \Psi \end{aligned}$$

where $H^{(1)} := Q^{-1}\mathcal{I}_\delta K \mathcal{I}_\delta^* Q^{-1}$ and Ψ is the adjoint of Ψ^* and which takes a function g and gives the following vector: $\Psi g = \int_0^T s(\mathcal{I}_s g)\mathrm{d}s$. Thus, $\bar{V}_1$ acts on functions via:

$$(\bar{V}_1 g)(x,t) = \int_\Omega \int_0^T \left(ts \sum_{i=1}^J \sum_{j=1}^J H_{ij}^{(1)} \phi_i(x)\phi_j(y) \right) g(y,s)\mathrm{d}s\mathrm{d}y$$

My idea was to continue with this distribution $\mathcal{N}(\bar{m}_1, \bar{V}_1)$ as the new prior and to condition on $A_{2\delta}u = F^2$ where $F^2 := \mathcal{I}_{2\delta}f$. This doesn't produce the Backward Euler-Galerkin since the mean function $\bar{m}_1$ is no longer constant in time and also applying $\mathcal{I}_{2\delta}$ will make a matrix $(M + 2\delta A)$ appear in the update equation instead of $Q = (M + \delta A)$.