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 $\Phi(x) := (\phi_1(x),\ldots,\phi_J(x))^T$. 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 $\mathcal{I}_s := (I_1(s),\ldots,I_J(s))^T$ 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:

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

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 $Q := M + \delta A$.