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:

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

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