Model problem: Let $\hat{\boldsymbol{y}} \in L^2(0,T;\mathbb{R}^N)$ a given target and $\gamma, \nu > 0$ two penalization parameters. We minimize the cost functional

$$\min_{\boldsymbol{y},\boldsymbol{u}} J(\boldsymbol{y},\boldsymbol{u}) := \frac{1}{2}\int_0^T |\boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t)|^2\,dt + \frac{\gamma}{2}|\boldsymbol{y}(T) - \hat{\boldsymbol{y}}(T)|^2 + \frac{\nu}{2}\int_0^T |\boldsymbol{u}(t)|^2\,dt, \tag{1}$$

such that

$$\dot{\boldsymbol{y}}(t) + A\boldsymbol{y}(t) = \boldsymbol{u}(t), \quad \boldsymbol{y}(0) = \boldsymbol{y}_0. \tag{2}$$

Here, $\boldsymbol{y}(t), \hat{\boldsymbol{y}}(t), \boldsymbol{u}(t) \in \mathbb{R}^N$ are column vectors, $|\cdot|$ denotes the Euclidean norm, $A \in \mathbb{R}^{N\times N}$ can be seen as the discretization matrix of some spatial operators, and $\boldsymbol{y}_0 \in \mathbb{R}^N$ is a given initial vector. This problem can be seen as the semi-discretization in space of some linear PDE-constrained optimization problems.

First-order optimality system: The standard approach to treat this problem is to introduce a Lagrange multiplier $\boldsymbol{p} \in L^2(0,T;\mathbb{R}^N)$, and write the Lagrangian as

$$\mathcal{L}(\boldsymbol{y},\boldsymbol{p},\boldsymbol{u}) = J(\boldsymbol{y},\boldsymbol{u}) - \int_0^T \boldsymbol{p}^T(t)\left(\dot{\boldsymbol{y}}(t) + A\boldsymbol{y}(t) - \boldsymbol{u}(t)\right)\,dt.$$

For any variation $\delta\boldsymbol{y} \in C^\infty(0,T;\mathbb{R}^N)$ with $\delta\boldsymbol{y}(0) = 0$, we have

$$\begin{aligned}\partial_{\boldsymbol{y}}\mathcal{L}[\delta\boldsymbol{y}] = &\int_0^T \left(\boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t)\right)^T \delta\boldsymbol{y}(t)\,dt + \gamma\left(\boldsymbol{y}(T) - \hat{\boldsymbol{y}}(T)\right)^T \delta\boldsymbol{y}(T)\\ &+ \int_0^T \dot{\boldsymbol{p}}^T(t)\delta\boldsymbol{y}(t)\,dt - \int_0^T (A^T\boldsymbol{p})^T(t)\delta\boldsymbol{y}(t)\,dt - \boldsymbol{p}^T(T)\delta\boldsymbol{y}(T) + \boldsymbol{p}^T(0)\delta\boldsymbol{y}(0).\end{aligned}$$

Equating it to zero gives

$$-\dot{\boldsymbol{p}}(t) + A^T\boldsymbol{p}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t), \quad \boldsymbol{p}(T) = \gamma\left(\boldsymbol{y}(T) - \hat{\boldsymbol{y}}(T)\right).$$

Similarly, for any variation $\delta\boldsymbol{u} \in C^\infty(0,T;\mathbb{R}^N)$, we have

$$\partial_{\boldsymbol{u}}\mathcal{L}[\delta\boldsymbol{u}] = \int_0^T \nu\boldsymbol{u}(t)^T\delta\boldsymbol{u}(t)\,dt + \int_0^T \boldsymbol{p}^T(t)\delta\boldsymbol{u}(t)\,dt.$$

Equating it to zero gives

$$\boldsymbol{p}(t) + \nu\boldsymbol{u}(t) = 0.$$

We then obtain the optimality system

$$\begin{aligned}\dot{\boldsymbol{y}}(t) + A\boldsymbol{y}(t) &= \boldsymbol{u}(t), & \boldsymbol{y}(0) &= \boldsymbol{y}_0,\\ -\dot{\boldsymbol{p}}(t) + A^T\boldsymbol{p}(t) &= \boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t), & \boldsymbol{p}(T) &= \gamma\left(\boldsymbol{y}(T) - \hat{\boldsymbol{y}}(T)\right),\\ \nu\boldsymbol{u}(t) &= -\boldsymbol{p}(t),\end{aligned} \tag{3}$$

which can also be written in the integral form as

$$\begin{aligned}&\boldsymbol{y}(t) = \boldsymbol{y}_0 + \int_0^t \left(-A\boldsymbol{y}(t) + \boldsymbol{u}(t)\right)\,dt,\\ &\boldsymbol{p}(t) = \boldsymbol{p}(T) + \int_t^T \left(-A^T\boldsymbol{p}(t) + \boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t)\right)\,dt,\\ &\boldsymbol{u}(t) = -\boldsymbol{p}(t)/\nu, \quad t \in (0,T).\end{aligned}$$

Reduced cost functional: To enter in the same optimization framework proposed in [1], we introduce the solution operator $\mathcal{S} : L^2(0,T;\mathbb{R}^N) \to L^2(0,T;\mathbb{R}^N)$ such that $\mathcal{S}\boldsymbol{u} = \boldsymbol{y}$. Substituting $\mathcal{S}\boldsymbol{u}$ into the cost functional gives

$$J(\boldsymbol{u}) = J(\mathcal{S}\boldsymbol{u},\boldsymbol{u}) = \frac{1}{2}\int_0^T |\mathcal{S}\boldsymbol{u}(t) - \hat{\boldsymbol{y}}(t)|^2\,dt + \frac{\gamma}{2}|\mathcal{S}\boldsymbol{u}(T) - \hat{\boldsymbol{y}}(T)|^2 + \frac{\nu}{2}\int_0^T |\boldsymbol{u}(t)|^2\,dt,$$

where we still use J to denote the reduced cost functional. We can then write the optimization problem as $\min_{\boldsymbol{u}} J(\boldsymbol{u})$ and use the elimination approach proposed in [1].

Decomposition: There are two different ways to decompose the problem and write

$$J(\boldsymbol{u}_1, \boldsymbol{u}_2).$$

If we decompose the control variable $\boldsymbol{u}$ into two controls $\boldsymbol{u}_1(t) \in \mathbb{R}^{N_1}$ and $\boldsymbol{u}_2(t) \in \mathbb{R}^{N_2}$ with $N_1 + N_2 = N$. This is exactly in the spirit of [1], which corresponds to a space decomposition. In this case, the functional $J : L^2(0,T;\mathbb{R}^{N_1}) \times L^2(0,T;\mathbb{R}^{N_2}) \to \mathbb{R}$. Instead, one can also decompose the time interval $(0,T)$ into two subintervals $Q_1 := (T_0, T_1)$ and $Q_2 := (T_1, T_2)$ with $T_0 = 0$, $T_2 = T$ and $T_1 \in (0,T)$. This corresponds to a time decomposition, and the two associated controls $\boldsymbol{u}_1(t) \in \mathbb{R}^N$, $t \in Q_1$ and $\boldsymbol{u}_2(t) \in \mathbb{R}^N$, $t \in Q_2$. In this case, the functional $J : L^2(T_0,T_1;\mathbb{R}^N) \times L^2(T_1,T_2;\mathbb{R}^N) \to \mathbb{R}$.

As we are interested in the second case, let us derive the algorithm following the approach proposed in [1]. Assuming that for every $\boldsymbol{u}_1$ the equation $\nabla_{\boldsymbol{u}_2} J(\boldsymbol{u}_1, \boldsymbol{u}_2) = 0$ admits a unique solution $\boldsymbol{u}_2$ and that $\nabla_{\boldsymbol{u}_2\boldsymbol{u}_2} J(\boldsymbol{u}_1, \boldsymbol{u}_2)$ is invertible for every $(\boldsymbol{u}_1, \boldsymbol{u}_2)$, then applying the implicit function theorem, there exists a continuously differentiable mapping $h : L^2(T_0,T_1;\mathbb{R}^N) \to L^2(T_1,T_2;\mathbb{R}^N)$ such that we can eliminate $\boldsymbol{u}_2$ and obtain $\nabla_{\boldsymbol{u}_2} J(\boldsymbol{u}_1, h(\boldsymbol{u}_1)) = 0$. We may apply the Newton iteration to solve the reduced optimality condition $F(\boldsymbol{u}_1) = \nabla_{\boldsymbol{u}_1} J(\boldsymbol{u}_1, h(\boldsymbol{u}_1)) = 0$. For iteration index $k = 0, 1, \ldots$, one solves

$$\boldsymbol{u}_1^{k+1} = \boldsymbol{u}_1^k - \left(JF(\boldsymbol{u}_1^k) \right)^{-1} \nabla_{\boldsymbol{u}_1} J(\boldsymbol{u}_1^k, h(\boldsymbol{u}_1^k)),$$

which is exactly what has been shown in [1, Eq. 4].

As also discussed in [1], our ultimate goal is to solve the optimization problem. Thus, one can also perform such variable elimination on the objective function J, that is, $\tilde{J}(\boldsymbol{u}_1) := J(\boldsymbol{u}_1, h(\boldsymbol{u}_1))$, and then apply a gradient descent method to solve the minimization problem as

$$\boldsymbol{u}_1^{k+1} = \boldsymbol{u}_1^k - \alpha \nabla \tilde{J}(\boldsymbol{u}_1^k).$$

with α the step size satisfies $\tilde{J}(\boldsymbol{u}_1^k - \alpha \nabla \tilde{J}(\boldsymbol{u}_1^k)) < \tilde{J}(\boldsymbol{u}_1^k)$.

Algorithm: The solving process can be resumed as:

1. For a given $\boldsymbol{u}_1^k$, one can find the state variable $\boldsymbol{y}_1^k$ with

$$\boldsymbol{y}_1^k(t) = \boldsymbol{y}_0 + \int_0^t \left(-A\boldsymbol{y}_1^k(t) + \boldsymbol{u}_1^k(t) \right) dt.$$

   This consists in applying the solution operator $\mathcal{S}_1 \boldsymbol{u}_1^k$.

2. Using the fact that $\boldsymbol{y}_2^k(T_1) = \boldsymbol{y}_1^k(T_1)$, one solves the system

$$\begin{aligned} \boldsymbol{y}_2^k(t) &= \boldsymbol{y}_2^k(T_1) + \int_{T_1}^t \left(-A\boldsymbol{y}_2^k(t) + \boldsymbol{u}_2^k(t) \right) dt, \\ \boldsymbol{p}_2^k(t) &= \boldsymbol{p}_2^k(T_2) + \int_t^{T_2} \left(-A^T \boldsymbol{p}_2^k(t) + \boldsymbol{y}_2^k(t) - \hat{\boldsymbol{y}}(t) \right) dt, \\ \boldsymbol{u}_2^k(t) &= -\boldsymbol{p}_2^k(t)/\nu, \quad t \in (T_1, T_2). \end{aligned}$$

   The above part consists in evaluating the mapping $\boldsymbol{u}_2^k = h(\boldsymbol{u}_1^k)$.

3. Update the control variable in Q_1 with

$$\boldsymbol{u}_1^{k+1} = \boldsymbol{u}_1^k - \alpha \left(\nabla_{\boldsymbol{u}_1} \tilde{J} \left(\boldsymbol{u}_1^k, h(\boldsymbol{u}_1^k) \right) + \nabla_{\boldsymbol{u}_2} \tilde{J} \left(\boldsymbol{u}_1^k, h(\boldsymbol{u}_1^k) \right) h'(\boldsymbol{u}_1^k) \right).$$

   As the implicit mapping h fulfils $\nabla_{\boldsymbol{u}_2} J(\boldsymbol{u}_1^k, h(\boldsymbol{u}_1^k)) = 0$. The update can then be written as

$$\boldsymbol{u}_1^{k+1} = \boldsymbol{u}_1^k - \alpha \nabla_{\boldsymbol{u}_1} \tilde{J} \left(\boldsymbol{u}_1^k, h(\boldsymbol{u}_1^k) \right) = \boldsymbol{u}_1^k - \alpha(\boldsymbol{p}_1^k + \nu \boldsymbol{u}_1^k),$$

where $\boldsymbol{p}_1^k$ is given by

$$\boldsymbol{p}_1^k(t) = \boldsymbol{p}_1^k(T_1) + \int_t^{T_1} \left(-A^T \boldsymbol{p}_1^k(t) + \boldsymbol{y}_1^k(t) - \hat{\boldsymbol{y}}(t) \right) dt,$$

with $\boldsymbol{p}_1^k(T_1) = \boldsymbol{p}_2^k(T_1)$.

Here, we are in the case with an exact solve of $\nabla_{\boldsymbol{u}_2} J(\boldsymbol{u}_1, \boldsymbol{u}_2) = 0$, as $\nabla_{\boldsymbol{u}_2} J(\boldsymbol{u}_1, \boldsymbol{u}_2) = \nu \boldsymbol{u}_2^k + \boldsymbol{p}_2^k$.

Discretize-then-optimize: We describe here the discretize-then-optimize approach to solve the problem (1)-(2). Let $0 = t_0 < t_1 < \ldots < t_M = T$ with uniform time step $\Delta t = T/M$. Denote $\boldsymbol{y}_m \approx \boldsymbol{y}(t_m)$, $\hat{\boldsymbol{y}}_m \approx \hat{\boldsymbol{y}}(t_m)$, $\boldsymbol{u}_m \approx \boldsymbol{u}(t_m)$ and $\boldsymbol{p}_m \approx \boldsymbol{p}(t_m)$. Applying the Crank-Nicolson time integration method for (2) gives

$$\frac{\boldsymbol{y}_{m+1} - \boldsymbol{y}_m}{\Delta t} + A \frac{\boldsymbol{y}_{m+1} + \boldsymbol{y}_m}{2} = \frac{\boldsymbol{u}_{m+1} + \boldsymbol{u}_m}{2} \Leftrightarrow \left(I_N + \frac{\Delta t}{2} A \right) \boldsymbol{y}_{m+1} - \left(I_N - \frac{\Delta t}{2} A \right) \boldsymbol{y}_m = \frac{\Delta t}{2} (\boldsymbol{u}_{m+1} + \boldsymbol{u}_m), \quad (4)$$

for $m = 0, \ldots, M-1$ and a given $\boldsymbol{y}_0$. To keep consistence with the Crank-Nicolson method, we use the trapezodial rule for numerical integration of the cost function (1) and find

$$J_M(\boldsymbol{y}, \boldsymbol{u}) := \frac{\Delta t}{4} \sum_{m=0}^{M-1} \left(|\boldsymbol{y}_{m+1} - \hat{\boldsymbol{y}}_{m+1}|^2 + |\boldsymbol{y}_m - \hat{\boldsymbol{y}}_m|^2 \right) + \frac{\gamma}{2} |\boldsymbol{y}_M - \hat{\boldsymbol{y}}_M|^2 + \frac{\nu \Delta t}{4} \sum_{m=0}^{M-1} \left(|\boldsymbol{u}_{m+1}|^2 + |\boldsymbol{u}_m|^2 \right). \quad (5)$$

The discrete Lagrangian then reads

$$\mathcal{L} = J_M - \sum_{m=0}^{M-1} \boldsymbol{p}_{m+1}^T \left(\left(I_N + \frac{\Delta t}{2} A \right) \boldsymbol{y}_{m+1} - \left(I_N - \frac{\Delta t}{2} A \right) \boldsymbol{y}_m - \frac{\Delta t}{2} (\boldsymbol{u}_{m+1} + \boldsymbol{u}_m) \right). \quad (6)$$

To obtain the discrete adjoint equation, one needs to do the "discrete integration by parts" in (6), that is

$$\begin{aligned} &\sum_{m=0}^{M-1} \boldsymbol{p}_{m+1}^T \left(\left(I_N + \frac{\Delta t}{2} A \right) \boldsymbol{y}_{m+1} - \left(I_N - \frac{\Delta t}{2} A \right) \boldsymbol{y}_m \right) \\ &= \sum_{m=1}^{M} \boldsymbol{p}_m^T \left(I_N + \frac{\Delta t}{2} A \right) \boldsymbol{y}_m - \sum_{m=0}^{M-1} \boldsymbol{p}_{m+1}^T \left(I_N - \frac{\Delta t}{2} A \right) \boldsymbol{y}_m \\ &= \sum_{m=1}^{M-1} \left(\left(I_N + \frac{\Delta t}{2} A \right)^T \boldsymbol{p}_m - \left(I_N - \frac{\Delta t}{2} A \right)^T \boldsymbol{p}_{m+1} \right)^T \boldsymbol{y}_m + \boldsymbol{p}_M^T \left(I_N + \frac{\Delta t}{2} A \right) \boldsymbol{y}_M - \boldsymbol{p}_1^T \left(I_N - \frac{\Delta t}{2} A \right) \boldsymbol{y}_0. \end{aligned}$$

Meanwhile, we re-write the sum over m of $\boldsymbol{y}_m$ in (5),

$$\frac{\Delta t}{4} \sum_{m=0}^{M-1} |\boldsymbol{y}_{m+1} - \hat{\boldsymbol{y}}_{m+1}|^2 + \frac{\Delta t}{4} \sum_{m=0}^{M-1} |\boldsymbol{y}_m - \hat{\boldsymbol{y}}_m|^2 = \frac{\Delta t}{2} \sum_{m=1}^{M-1} |\boldsymbol{y}_m - \hat{\boldsymbol{y}}_m|^2 + \frac{\Delta t}{4} |\boldsymbol{y}_M - \hat{\boldsymbol{y}}_M|^2 + \frac{\Delta t}{4} |\boldsymbol{y}_0 - \hat{\boldsymbol{y}}_0|^2.$$

We derive now the discrete adjoint equation

$$\partial_{\boldsymbol{y}_m} \mathcal{L} = -\left(\left(I_N + \frac{\Delta t}{2} A \right)^T \boldsymbol{p}_m - \left(I_N - \frac{\Delta t}{2} A \right)^T \boldsymbol{p}_{m+1} \right) + \Delta t (\boldsymbol{y}_m - \hat{\boldsymbol{y}}_m), \quad m = 1, \ldots, M-1,$$

with the final condition

$$\partial_{\boldsymbol{y}_M} \mathcal{L} = -\left(I_N + \frac{\Delta t}{2} A \right)^T \boldsymbol{p}_M + \frac{\Delta t}{2} (\boldsymbol{y}_M - \hat{\boldsymbol{y}}_M) + \gamma (\boldsymbol{y}_M - \hat{\boldsymbol{y}}_M).$$

We treat in a similar way of the sum related to $\boldsymbol{u}_m$ in (5)-(6),

$$\begin{aligned} \frac{\nu \Delta t}{4} \sum_{m=0}^{M-1} |\boldsymbol{u}_{m+1}|^2 + \frac{\nu \Delta t}{4} \sum_{m=0}^{M-1} |\boldsymbol{u}_m|^2 &= \frac{\nu \Delta t}{2} \sum_{m=1}^{M-1} |\boldsymbol{u}_m|^2 + \frac{\nu \Delta t}{4} |\boldsymbol{u}_M|^2 + \frac{\nu \Delta t}{4} |\boldsymbol{u}_0|^2, \\ \frac{\Delta t}{2} \sum_{m=0}^{M-1} \boldsymbol{p}_{m+1}^T \boldsymbol{u}_{m+1} + \frac{\Delta t}{2} \sum_{m=0}^{M-1} \boldsymbol{p}_{m+1}^T \boldsymbol{u}_m &= \frac{\Delta t}{2} \sum_{m=1}^{M-1} (\boldsymbol{p}_m^T + \boldsymbol{p}_{m+1}^T) \boldsymbol{u}_m + \frac{\Delta t}{2} \boldsymbol{p}_M^T \boldsymbol{u}_M + \frac{\Delta t}{2} \boldsymbol{p}_1^T \boldsymbol{u}_0. \end{aligned}$$

This then gives the discrete optimality condition

$$\partial_{\boldsymbol{u}_m}\mathcal{L}=\frac{\Delta t}{2}(\boldsymbol{p}_m+\boldsymbol{p}_{m+1})+\nu\Delta t\boldsymbol{u}_m,\quad m=1,\dots,M-1,\quad \partial_{\boldsymbol{u}_M}\mathcal{L}=\frac{\nu\Delta t}{2}\boldsymbol{u}_M+\frac{\Delta t}{2}\boldsymbol{p}_M,\quad \partial_{\boldsymbol{u}_0}\mathcal{L}=\frac{\nu\Delta t}{2}\boldsymbol{u}_0+\frac{\Delta t}{2}\boldsymbol{p}_1.$$

Equating these partial derivatives to zero gives the discrete optimality system using Crank-Nicolson method,

$$\begin{aligned}
\left(I_N+\frac{\Delta t}{2}A\right)\boldsymbol{y}_{m+1}-\left(I_N-\frac{\Delta t}{2}A\right)\boldsymbol{y}_m&=\frac{\Delta t}{2}(\boldsymbol{u}_{m+1}+\boldsymbol{u}_m), && m=0,\dots,M-1,\\
\left(I_N+\frac{\Delta t}{2}A^T\right)\boldsymbol{p}_m-\left(I_N-\frac{\Delta t}{2}A^T\right)\boldsymbol{p}_{m+1}&=\Delta t(\boldsymbol{y}_m-\hat{\boldsymbol{y}}_m), && m=1,\dots,M-1,\\
\left(I_N+\frac{\Delta t}{2}A^T\right)\boldsymbol{p}_M&=\left(\frac{\Delta t}{2}+\gamma\right)(\boldsymbol{y}_M-\hat{\boldsymbol{y}}_M),\\
\nu\boldsymbol{u}_m&=-\frac{\boldsymbol{p}_m+\boldsymbol{p}_{m+1}}{2}, && m=1,\dots,M-1,\\
\nu\boldsymbol{u}_M&=-\boldsymbol{p}_M,\quad \nu\boldsymbol{u}_0=-\boldsymbol{p}_1.
\end{aligned}\tag{7}$$

We can substitute $\boldsymbol{u}_m$ by $\boldsymbol{p}_m$ to obtain the discrete reduced optimality system

$$\begin{aligned}
\left(I_N+\frac{\Delta t}{2}A\right)\boldsymbol{y}_{m+1}-\left(I_N-\frac{\Delta t}{2}A\right)\boldsymbol{y}_m&=-\frac{\Delta t}{4\nu}(\boldsymbol{p}_m+2\boldsymbol{p}_{m+1}+\boldsymbol{p}_{m+2}), && m=1,\dots,M-2,\\
\left(I_N+\frac{\Delta t}{2}A\right)\boldsymbol{y}_1-\left(I_N-\frac{\Delta t}{2}A\right)\boldsymbol{y}_0&=-\frac{\Delta t}{4\nu}(3\boldsymbol{p}_1+\boldsymbol{p}_2),\\
\left(I_N+\frac{\Delta t}{2}A\right)\boldsymbol{y}_M-\left(I_N-\frac{\Delta t}{2}A\right)\boldsymbol{y}_{M-1}&=-\frac{\Delta t}{4\nu}(3\boldsymbol{p}_M+\boldsymbol{p}_{M-1}),\\
\left(I_N+\frac{\Delta t}{2}A^T\right)\boldsymbol{p}_m-\left(I_N-\frac{\Delta t}{2}A^T\right)\boldsymbol{p}_{m+1}&=\Delta t(\boldsymbol{y}_m-\hat{\boldsymbol{y}}_m), && m=1,\dots,M-1,\\
\left(I_N+\frac{\Delta t}{2}A^T\right)\boldsymbol{p}_M&=\left(\frac{\Delta t}{2}+\gamma\right)(\boldsymbol{y}_M-\hat{\boldsymbol{y}}_M).
\end{aligned}\tag{8}$$

Denote $\boldsymbol{U}_1=(\boldsymbol{y}_1,\dots,\boldsymbol{y}_M)^T$ and $\boldsymbol{U}_2=(\boldsymbol{p}_1,\dots,\boldsymbol{p}_M)^T$. The all-at-once block matrix form is given by

$$\tilde{A}\boldsymbol{U}=\boldsymbol{F},\quad \tilde{A}=\begin{bmatrix}\tilde{A}_{11}&\tilde{A}_{12}\\ \tilde{A}_{21}&\tilde{A}_{22}\end{bmatrix},\quad \boldsymbol{U}=\begin{bmatrix}\boldsymbol{U}_1\\ \boldsymbol{U}_2\end{bmatrix},\quad \boldsymbol{F}=\begin{bmatrix}\boldsymbol{F}_1\\ \boldsymbol{F}_2\end{bmatrix},$$

with

$$\tilde{A}_{11}=\begin{bmatrix}I_N+\frac{\Delta t}{2}A&&&&\\ I_N-\frac{\Delta t}{2}A&I_N+\frac{\Delta t}{2}A&&&\\ &\ddots&\ddots&&\\ &&I_N-\frac{\Delta t}{2}A&I_N+\frac{\Delta t}{2}A&\\ &&&I_N-\frac{\Delta t}{2}A&I_N+\frac{\Delta t}{2}A\end{bmatrix},\quad \tilde{A}_{12}=\begin{bmatrix}\frac{3\Delta t}{4\nu}I_N&\frac{\Delta t}{4\nu}I_N&&&\\ \frac{\Delta t}{4\nu}I_N&\frac{\Delta t}{2\nu}I_N&\frac{\Delta t}{4\nu}I_N&&\\ &\ddots&\ddots&\ddots&\\ &&\frac{\Delta t}{4\nu}I_N&\frac{\Delta t}{2\nu}I_N&\frac{\Delta t}{4\nu}I_N\\ &&&\frac{\Delta t}{4\nu}I_N&\frac{3\Delta t}{4\nu}I_N\end{bmatrix},$$

$$\tilde{A}_{21}=\begin{bmatrix}-\Delta tI_N&&&&\\ &-\Delta tI_N&&&\\ &&\ddots&&\\ &&&-\Delta tI_N&\\ &&&&-\left(\frac{\Delta t}{2}+\gamma\right)I_N\end{bmatrix},$$

$$\tilde{A}_{22}=\begin{bmatrix}I_N+\frac{\Delta t}{2}A^T&-\left(I_N-\frac{\Delta t}{2}A^T\right)&&&\\ &I_N+\frac{\Delta t}{2}A^T&-\left(I_N-\frac{\Delta t}{2}A^T\right)&&\\ &&\ddots&\ddots&\\ &&&I_N+\frac{\Delta t}{2}A^T&-\left(I_N-\frac{\Delta t}{2}A^T\right)\\ &&&&I_N+\frac{\Delta t}{2}A^T\end{bmatrix},$$

and

$$\boldsymbol{F}_1 = \begin{bmatrix} \left(I_N - \frac{\Delta t}{2} A^T\right) \boldsymbol{y}_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \boldsymbol{F}_2 = \begin{bmatrix} -\Delta t \hat{\boldsymbol{y}}_1 \\ \vdots \\ -\Delta t \hat{\boldsymbol{y}}_{M-1} \\ -\left(\frac{\Delta t}{2} + \gamma\right) \hat{\boldsymbol{y}}_M \end{bmatrix}.$$

Note that the discrete optimality system obtained in (7) is different from applying directly the Crank-Nicolson method to discretize the optimality system (3).

References

[1] G. Ciaramella and T. Vanzan, *Variable reduction as a nonlinear preconditioning approach for optimization problems*, in Domain Decomposition Methods in Science and Engineering XXVIII, Springer Cham, 2025.