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

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

such that

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

Here, $\boldsymbol{y}(t), \boldsymbol{u}(t) \in \mathbb{R}^N$ are column vectors, and $|\cdot|$ is the Euclidean norm. This problem can be seen as the semi-discretization in space of some linear PDE-constrained optimization problem.

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

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

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

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

Equating it to zero gives

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

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

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

which can also be written in the integral form as

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

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

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

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

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

   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

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

   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

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

   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

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

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

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

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 $\nabla_{\boldsymbol{u}_2} J(\boldsymbol{u}_1, \boldsymbol{u}_2) = \nu \boldsymbol{u}_2^k + \boldsymbol{p}_2^k$.

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.