

OPTIMAL DIRICHLET CONTROL OF PARTIAL DIFFERENTIAL EQUATIONS ON NETWORKS*

MARTIN STOLL[†] AND MAX WINKLER[‡]

Abstract. Differential equations on metric graphs can describe many phenomena in the physical world but also the spread of information on social media. To efficiently compute the optimal setup of the differential equation for a given desired state is a challenging numerical analysis task. In this work, we focus on the task of solving an optimization problem subject to a linear differential equation on a metric graph with the control defined on a small set of Dirichlet nodes. We discuss the discretization by finite elements and provide rigorous error bounds as well as an efficient preconditioning strategy to deal with the large-scale case. We show in various examples that the method performs very robustly.

Key words. complex networks, optimal Dirichlet control, preconditioning, saddle point systems, error estimation

AMS subject classifications. 65F08, 65F50, 65M60, 65N15,

1. Introduction. Graphs and networks¹ are ubiquitous in the modeling of physical phenomena, the representation of data, and other applications [7]. While the graph Laplacian is a crucial tool in the analysis of such networks [5, 10, 56, 61, 64, 67], there are many examples where it is often not sufficient to only reflect the binary relationships of being connected or not.

In order to model complex phenomena, one can extend the concept of a graph to a so-called metric graph, where the differential operator describing the dynamics goes beyond the usually applied graph Laplacian [43, 45]. A metric graph is a graph where a length is assigned to each edge, and we define a differential operator, often called the *Hamiltonian*, on the graph domain. There are many contributions dealing with PDEs on metric graphs, often tailored to particular applications. In this paper, we focus on elliptic and parabolic problems often found in the modeling of carbon nano-structures or groundwater flow (cf. [34, 47]), but besides that, also hyperbolic problems have been studied intensively in the literature and cover many interesting applications like gas dynamics and traffic flow [12, 13, 15, 16, 21, 24, 25, 27, 30, 53, 69].

Of particular interest in the context of PDE models is the control or identification of a source term or parameter such that a desired or measured state is achieved. These questions lead to optimal control² problems governed by PDE models, which received much attention over the last years; cf. [32, 33, 66]. Optimal control problems for ODE or PDE models on networks are studied, e.g., in [11, 26, 37] mainly from an analytical perspective.

For the numerical treatment of these kind of problems, a couple of difficulties arise. While for the discretization of PDEs in classical domains, a minimal angle condition constrains the number of degrees of freedom that are connected to a vertex/mesh-point, this is not the case when considering complex networks with dense adjacency structure or with highly varying degree distributions, e.g., hubs in social networks. In these cases, the resulting matrices stemming from the discretization are rather dense. Thus, it is a key objective in the

*Received October 17, 2020. Accepted April 15, 2021. Published online on June 7, 2021. Recommended by M. Benzi.

[†]Chair of Scientific Computing, Technische Universität Chemnitz, Reichenhainer Str. 41, 09107 Chemnitz, Germany (martin.stoll@math.tu-chemnitz.de).

[‡]Chair of Numerical Mathematics (PDEs), Technische Universität Chemnitz, Reichenhainer Str. 41, 09107 Chemnitz, Germany (max.winkler@math.tu-chemnitz.de).

¹We here do not distinguish between graphs and networks [4] even though networks are often found in the real world and graphs are often thought of as their mathematical representation.

²We here view optimal control problems synonymously with PDE-constrained optimization problems as discussed in [66] and understand that there is some ambiguity in the context of optimal control for dynamical systems.

numerical analysis to construct tailored discretizations and preconditioned iterative solvers for the resulting systems of linear equations taking into account the adjacency structure of the network as much as possible. As a first general work in this direction, we want to mention the work of Arioli and Benzi [3], where the discretization and iterative solution of elliptic and parabolic problems on graphs are studied.

In the present article we aim at an extension of this theory to Dirichlet control problems constrained by elliptic or parabolic problems on metric graphs. We derive optimality conditions for an optimal Dirichlet control problem with control constraints and construct a finite element discretization. Furthermore, sharp discretization error estimates for the approximate controls and the states in various norms are proved. Similar studies for related problems in bounded domains can be found in [1, 8, 40, 46, 52, 68], where, e.g., linear convergence of the approximate controls in the L^2 -norm can be proved, provided that the exact solutions are sufficiently regular. A surprising observation in our setting is that the approximate controls u_h converge even quadratically with respect to the mesh parameter $h > 0$, i.e., the estimate $|u - u_h| \leq Ch^2$ is satisfied. This is due to the fact that the control space is finite-dimensional and the solutions on graphs are more regular. Moreover, it will be confirmed that the discretization error estimate is independent of the adjacency structure of the graph.

As a second main result we construct preconditioned iterative solvers for the resulting discretized optimality conditions of control problems for elliptic and parabolic equations. We propose a Schur-complement-based preconditioner [29, 39, 49, 50, 54, 63] and confirm that the iterative method performs not only robustly with respect to changes in the regularization and discretization parameter but also with respect to the adjacency pattern of the metric graph.

Finally, we want to mention that there are additional techniques that might allow for a more efficient solution of PDEs and control problems in complex networks like model reduction techniques [9, 42] or homogenization [28], but the setup there is for now different from ours.

The model we use in this paper did not originate from a particular application as our focus is on the development of a full framework ranging from discretization, error estimation, to tailored iterative solvers. Our model provides a sufficient amount of complexity, and we believe, while only considering a linear constraint here, that our results carry over to the nonlinear case, which can then be used for real-world examples such as reaction-diffusion systems in chemical engineering [20] or traffic density.

The paper is structured as follows. First, we formulate the model state equation and summarize some preliminary results in Section 2. Moreover, we introduce a related optimal control problem and derive the optimality system. Our discretization strategy is studied in Section 3. There, we derive rigorous discretization error estimates for the finite element approximation of the state equation and the approximation of a discrete Kirchhoff operator, which appears in the discrete optimality system of the studied Dirichlet control problem. These results are then used for error estimates of the approximate solutions of the optimal control problem. An efficient solver for the optimality system and the corresponding preconditioners are investigated in Section 4. In Section 5 we also study the case that the state equation is parabolic. This requires a further discussion of a temporal discretization and an extension of our preconditioners to the new time dimension. Finally, in Section 6, we test the theoretically predicted behavior in several numerical experiments. To be more precise, we confirm that the discretization error estimates are sharp and that the preconditioned method is robust.

2. Optimal Dirichlet control problems on metric graphs.

2.1. Differential equations on metric graphs. We consider an undirected graph $\Gamma = (\mathcal{V}, \mathcal{E})$ consisting of a vertex set $\mathcal{V} = \{v_i\}_{i=1}^n$ and an edge set \mathcal{E} . Each edge $e \in \mathcal{E}$

connects a pair of nodes (v_a^e, v_b^e) with $v_a^e, v_b^e \in \mathcal{V}$. We denote by $m = |\mathcal{E}|$ the number of edges and by $n = |\mathcal{V}|$ the number of vertices.

Often the dynamics on the graph are based on the graph Laplacian defined as $\mathbf{L} = \mathbf{D} - \mathbf{W} = \mathbf{E}\mathbf{E}^\top$, where \mathbf{W} is the weight matrix of the undirected graph, \mathbf{E} its incidence matrix, and \mathbf{D} its degree matrix (cf. [67]). The graph Laplacian is in certain applications not sufficient to describe the intricate relationships as it only reflects information about the nodes being connected and the edge information encoded in the weight. A more sophisticated representation follows from the now introduced concept of metric graphs.

In a *metric graph* each edge is identified with an interval of length $L_e > 0$ on the real line, and to each interval we associate a differential operator, such as, e.g., the one-dimensional Schrödinger operator

$$(2.1) \quad (\mathcal{H}y)(x) := \left(-\frac{d^2}{dx^2} + c_0(x) \right) y(x)$$

with a potential function c_0 . Then, one can formulate a differential equation of the form

$$(\mathcal{H}y|_e)(x_e) = f|_e(x_e) \quad \text{for all } x_e \in (0, L_e),$$

where the functions $f|_e: (0, L_e) \rightarrow \mathbb{R}$, $e \in \mathcal{E}$, are given source terms. Here, x_e are local coordinates associated to the edge e . With a slight abuse of notation we will sometimes evaluate $y|_e$ at one of the vertices $v \in \mathcal{V}$ of e . Depending on the orientation of the local coordinate x_e , we then mean either $y|_e(0)$ or $y|_e(L_e)$.

Additionally, we can impose boundary or vertex conditions to couple these equations. There are of course several different vertex conditions, and we will distinguish among two different types. First, at vertices $v \in \mathcal{V}_\mathcal{K} \subset \mathcal{V}$, we have *homogeneous Neumann-Kirchhoff conditions*, i.e., there holds

$$(2.2) \quad (\mathcal{K}y)(v) := \sum_{e \in \mathcal{E}_v} \frac{d}{dx} y|_e(v) = 0,$$

with \mathcal{E}_v the edge set incident to the vertex v . At vertices $v \in \mathcal{V}_\mathcal{D} := \mathcal{V} \setminus \mathcal{V}_\mathcal{K}$, the solution y satisfies *Dirichlet conditions*

$$y|_e(v) = u_v, \quad \text{for all } e \in \mathcal{E}_v,$$

where $u \in \mathbb{R}^{n_\mathcal{D}}$, $n_\mathcal{D} := |\mathcal{V}_\mathcal{D}|$, is a vector containing the Dirichlet data. Note that if all nodes are Dirichlet nodes, then the problem decouples into a set of one-dimensional problems, and we exclude this trivial case by assuming $|\mathcal{V}_\mathcal{K}| \neq \emptyset$. Other boundary conditions are not discussed here but will be the subject of future research. For more information on vertex conditions in metric graphs, we refer to [7, 23, 60].

The Kirchhoff-Neumann conditions are the natural boundary conditions for the differential operator (2.1), as for each $v \in \mathcal{V}$ and test functions $\phi \in \otimes_{e \in \mathcal{E}} C^\infty(e)$ that are continuous in v and vanish at $v' \in \mathcal{V} \setminus \{v\}$, the formula

$$(2.3) \quad (\mathcal{K}y)(v) \phi(v) = \sum_{e \in \mathcal{E}_v} y|_e'(v) \phi|_e(v) = \sum_{e \in \mathcal{E}_v} \int_e [y'(x) \phi'(x) + y''(x) \phi(x)] dx$$

is satisfied. Here and in the following, we use the notation $\frac{d}{dx} y|_e(x) = y|_e'(x)$ and omit the subscript e unless the context requires otherwise. Throughout this article we consider the state

equation

$$(2.4) \quad \begin{cases} -y'' + c_0 y = f & \text{on all } e \in \mathcal{E}, \\ (\mathcal{K}y)(v) = 0 & \text{for } v \in \mathcal{V}_{\mathcal{K}}, \\ y(v) = u_v & \text{for } v \in \mathcal{V}_{\mathcal{D}}, \end{cases}$$

where $f \in \otimes_{e \in \mathcal{E}} L^2(e)$ and $u \in \mathbb{R}^{n_{\mathcal{D}}}$ are given data.

For an appropriate treatment of this boundary value problem, we require some further notation. We introduce the function spaces

$$L^2(\Gamma) = \bigotimes_{e \in \mathcal{E}} L^2(e) \quad \text{and} \quad H^1(\Gamma) = \bigotimes_{e \in \mathcal{E}} H^1(e) \cap C^0(\Gamma)$$

equipped with the norms

$$\|y\|_{L^2(\Gamma)}^2 := \sum_{e \in \mathcal{E}} \|y\|_{L^2(e)}^2, \quad \|y\|_{H^1(\Gamma)}^2 := \|y\|_{L^2(\Gamma)}^2 + \sum_{e \in \mathcal{E}} \|y'\|_{L^2(e)}^2.$$

Higher-order Sobolev spaces $H^m(\Gamma)$ with $m > 1$ are defined analogously. The $L^2(\Gamma)$ -inner product is denoted by $(f, g)_{L^2(\Gamma)} := \sum_{e \in \mathcal{E}} \int_e f(x) g(x) dx$. To establish the essential boundary conditions, we moreover define

$$H_{\mathcal{D}}^1(\Gamma) := \{y \in H^1(\Gamma) : y(v) = 0 \quad \forall v \in \mathcal{V}_{\mathcal{D}}\}.$$

In order to derive the weak formulation for the problem (2.4), we multiply each differential equation by a test function $w|_e \in H^1(e)$, integrate over e , and sum up the equations for all $e \in \mathcal{E}$. After an application of the integration-by-parts formula (2.3) and insertion of the vertex conditions, we arrive at

$$(2.5) \quad \text{Find } y \in H^1(\Gamma) : \quad y(v) = u_v, v \in \mathcal{V}_{\mathcal{D}}, \quad a(y, w) = (f, w)_{L^2(\Gamma)}, \quad \forall w \in H_{\mathcal{D}}^1(\Gamma),$$

with the bilinear form $a : H^1(\Gamma) \times H^1(\Gamma) \rightarrow \mathbb{R}$ defined by

$$a(y, w) := \sum_{e \in \mathcal{E}} \int_e (y'(x) w'(x) + c_0(x) y(x) w(x)) dx.$$

Throughout this article, the potential function c_0 belongs to $L^\infty(\Gamma) := \otimes_{e \in \mathcal{E}} L^\infty(e)$ and satisfies $c_0 \geq \underline{c} > 0$ a.e. in Γ . Moreover, the source term f belongs to $L^2(\Gamma)$. We can establish the following existence and regularity result:

LEMMA 2.1. *Let $\mathcal{V}_{\mathcal{D}} \neq \emptyset$. Then, for each $f \in L^2(\Gamma)$, $c_0 \in L^\infty(\Gamma)$, with $c_0 \geq \underline{c} > 0$, and $u \in \mathbb{R}^{n_{\mathcal{D}}}$, the problem (2.5) has a unique solution $y \in H^2(\Gamma)$ satisfying*

$$(2.6) \quad \|y\|_{H^2(\Gamma)} \leq 2 \sqrt{\frac{M}{\gamma}} \left(\|f\|_{L^2(\Gamma)} + C_{\Gamma} |u|_2 \right)$$

with $M := (1 + \|c_0\|_{L^\infty(\Gamma)})$, $\gamma := \min\{1, \underline{c}\}$, and

$$C_{\Gamma} := c \sqrt{\max_{v \in \mathcal{V}_{\mathcal{D}}} \text{card}(\mathcal{E}_v) \max_{e \in \mathcal{E}} (L_e^{-1} + L_e)}.$$

Here, $c > 0$ is a generic constant independent of u , f , c_0 , and Γ .

Proof. First, we define an extension y_u of u by

$$y_u|_e \in \mathcal{P}_1 \quad \forall e \in \mathcal{E}, \quad y_u(v) = \begin{cases} u_v & \forall v \in \mathcal{V}_{\mathcal{D}}, \\ 0 & \forall v \in \mathcal{V}_{\mathcal{K}}, \end{cases}$$

i.e., on each edge $e \in \mathcal{E}$, $y_u|_e$ is a linear function and satisfies the Dirichlet boundary conditions in $\mathcal{V}_{\mathcal{D}}$, as well as homogeneous Dirichlet conditions in $\mathcal{V}_{\mathcal{K}}$. The problem (2.5) can then be rewritten as

$$\text{Find } y_0 \in H_D^1(\Gamma): \quad a(y_0, w) = (f, w)_{L^2(\Gamma)} - a(y_u, w), \quad \forall w \in H_D^1(\Gamma),$$

and there holds $y = y_0 + y_u$. It is easy to show that a is bounded and coercive, i.e., the estimates

$$(2.7) \quad \begin{aligned} a(y, z) &\leq M \|y\|_{H^1(\Omega)} \|z\|_{H^1(\Omega)} \quad \forall y, z \in H^1(\Omega), \\ a(y, y) &\geq \gamma \|y\|_{H^1(\Omega)}^2 \quad \forall y \in H^1(\Omega), \end{aligned}$$

hold true. To show (2.7), we use the assumption that c_0 is uniformly positive. The result remains true for general $c_0 \geq 0$ and $y \in H_D^1(\Omega)$ provided that $\mathcal{V}_{\mathcal{D}} \neq \emptyset$. This follows from a Poincaré inequality on graphs and is proved in [36, Lemma 3.1]. The existence and uniqueness of a solution $y_0 \in H_D^1(\Gamma)$ can be concluded from the Lax-Milgram-Lemma, and there holds

$$(2.8) \quad \|y_0\|_{H^1(\Gamma)} \leq \sqrt{\frac{M}{\gamma}} \left(\|f\|_{L^2(\Gamma)} + \|y_u\|_{H^1(\Gamma)} \right).$$

It remains to bound the norm of y_u . To this end, let $e \in \mathcal{E}$ be arbitrary, and denote by $v, v' \in \mathcal{V}$ the endpoints of e . If $v, v' \in \mathcal{V}_{\mathcal{K}}$, then there holds $y_u|_e = 0$, and otherwise, if $v \in \mathcal{V}_{\mathcal{D}}$, then one easily computes

$$\|y_u\|_{L^2(e)} \leq \sqrt{\frac{L_e}{3}} |u_v - u_{v'}| \quad \text{and} \quad \|y'_u\|_{L^2(e)} = \sqrt{\frac{1}{L_e}} |u_v - u_{v'}|,$$

where we set $u_{v'} = 0$ in case of $v' \in \mathcal{V}_{\mathcal{K}}$. Summation over all $e \in \mathcal{E}$ yields

$$(2.9) \quad \begin{aligned} \|y_u\|_{H^1(\Gamma)}^2 &\leq c \sum_{v \in \mathcal{V}_{\mathcal{D}}} \sum_{e \in \mathcal{E}_v} |u_v - u_{v'}|^2 (L_e + L_e^{-1}) \\ &\leq c \max_{v \in \mathcal{V}_{\mathcal{D}}} \text{card}(\mathcal{E}_v) \max_{e \in \mathcal{E}} (L_e^{-1} + L_e) |u|_2^2. \end{aligned}$$

Together with (2.8) this yields the estimate

$$\|y\|_{H^1(\Gamma)} \leq \sqrt{\frac{M}{\gamma}} \left(\|f\|_{L^2(\Gamma)} + C_{\Gamma} |u|_2 \right).$$

As the functions $y|_e$ solve $-y''_e + c_0 y|_e = f|_e$ in e with fixed boundary data, the $H^2(e)$ -regularity follows from standard arguments. Moreover, it is easy to deduce the following a priori estimate

$$\|y\|_{H^2(\Gamma)}^2 \leq \|c_0\|_{L^\infty(\Gamma)} \|y\|_{L^2(\Gamma)} + \|f\|_{L^2(\Gamma)} \leq 2 \sqrt{\frac{M}{\gamma}} \left(\|f\|_{L^2(\Gamma)} + C_{\Gamma} |u|_2 \right),$$

where we have inserted the $L^2(e)$ -norm estimate in the last step. \square

When considering the estimate (2.9) we can observe in which way the graph structure influences the norm of the solution. In particular, the constant grows when there are extremely small or long edges, when the function values u_v in the Dirichlet nodes vary sharply, and when the cardinality of \mathcal{E}_v for some $v \in \mathcal{V}_D$ is large. Note that this effects, for instance, hubs in social networks, where the latter case can occur.

2.2. The optimal control problem. As an extension of the above, we now consider an optimal control problem of the form

$$(2.10a) \quad \text{Minimize } \frac{1}{2} \|y - \bar{y}\|_{L^2(\Gamma)}^2 + \frac{\beta}{2} |u|_2^2 \quad \text{over } y \in H^1(\Gamma), u \in \mathbb{R}^{n_D}$$

subject to the constraints

$$(2.10b) \quad \begin{cases} -y'' + c_0 y = f & \text{on all } e \in \mathcal{E}, \\ (\mathcal{K}y)(v) = 0 & v \in \mathcal{V}_K, \\ y(v) = u_v & v \in \mathcal{V}_D, \end{cases}$$

and

$$(2.10c) \quad u_a \leq u \leq u_b,$$

with a desired state $\bar{y} \in L^2(\Gamma)$ that shall be actuated and control bounds $u_a, u_b \in \mathbb{R}^{n_D}$ with $u_a \leq u_b$ (understood component-wise). Here, $|\cdot|_2$ is the Euclidean norm in \mathbb{R}^{n_D} . Other choices for the objective functional are certainly possible, e.g., one could fit the outflow at selected nodes $v \in \mathcal{V}$ by $|(\mathcal{K}y)(v) - \bar{y}_v|$. Moreover, when a sparse control is desired, the ℓ_1 -norm [62] could be used to measure control costs. The regularization parameter β is assumed to be positive. This assumption already guarantees the existence and uniqueness of a minimizer [66]. The state equation is understood in the weak sense (2.5). We can decompose the state into a part depending linearly on u and a constant contribution, i.e., $y = y_u + y_f$, with $y_u \in H^1(\Gamma)$, $y_u(v) = u_v$, for $v \in \mathcal{V}_D$, and $y_f \in H_D^1(\Gamma)$ satisfying

$$a(y_u, w) = 0 \quad \text{and} \quad a(y_f, w) = (f, w)_{L^2(\Gamma)}$$

for all $w \in H_D^1(\Gamma)$. This decomposition allows us to introduce a linear control-to-state operator $S: \mathbb{R}^{n_D} \rightarrow L^2(\Gamma)$ defined by $u \mapsto S(u) := y_u$, and one can eliminate the state equation in the optimal control problem (2.10). Then, we arrive at the reduced optimization problem

$$(2.11) \quad \text{Minimize } j(u) := \frac{1}{2} \|Su + y_f - \bar{y}\|_{L^2(\Gamma)}^2 + \frac{\beta}{2} |u|_2^2 \quad \text{s.t. } u \in U_{\text{ad}},$$

with $U_{\text{ad}} := \{u \in \mathbb{R}^{n_D} : u_a \leq u \leq u_b\}$ the set of admissible controls.

As S is linear, the functional j is quadratic and hence Fréchet-differentiable and convex. By differentiation using the chain rule, we can derive the optimality condition which is the variational inequality

$$(2.12) \quad (Su + y_f - \bar{y}, S(w - u))_{L^2(\Gamma)} + \beta u^\top (w - u) \geq 0 \quad \forall w \in U_{\text{ad}}.$$

Due to the convexity of j , this condition is also sufficient for u being the unique global solution of (2.11). In order to derive a more handy form of the optimality condition, we investigate the adjoint operator of S .

LEMMA 2.2. *The adjoint operator $S^* : L^2(\Gamma) \rightarrow \mathbb{R}^{n_D}$ has the representation $S^* = -\mathcal{K} \circ P$, where \mathcal{K} is defined in (2.2) and $P : L^2(\Gamma) \rightarrow H_D^1(\Gamma)$ is defined by $y \mapsto P(y) := p$, with $p \in H_D^1(\Gamma)$ solving*

$$a(w, p) = (y, w)_{L^2(\Gamma)} \quad \forall w \in H_D^1(\Gamma).$$

Proof. Standard regularity results for elliptic problems imply that P even maps onto $H^2(\Gamma) \cap H_D^1(\Gamma)$. Consequently, p solves $-p'' + c_0 p = y$ almost everywhere in Γ . Together with the integration-by-parts formula (2.3), we obtain for all $z \in \mathbb{R}^{n_D}$

$$(y, Sz)_{L^2(\Gamma)} = (-p'' + c_0 p, Sz)_{L^2(\Gamma)} = a(Sz, p) - (\mathcal{K}p)^\top z.$$

Due to $p \in H_D^1(\Gamma)$ there holds $a(Sz, p) = 0$. This implies $S^*y = -\mathcal{K}p$ and proves the assertion. \square

As a consequence, we can rewrite the optimality condition (2.12) using the equivalence

$$(Su + y_f - \bar{y}, Sw)_{L^2(\Gamma)} = (S^*(Su + y_f - \bar{y}), w)_{\mathbb{R}^{n_D}} = -(\mathcal{K}p, w)_{\mathbb{R}^{n_D}},$$

with $p \in H_D^1(\Gamma)$ the *adjoint state* being the weak solution of

$$\left\{ \begin{array}{ll} \left(-\frac{d^2}{dx^2} + c_0 \right) p = y - \bar{y} & \text{on all } e \in \mathcal{E}, \\ (\mathcal{K}p)(v) = 0 & v \in \mathcal{V}_{\mathcal{K}}, \\ p(v) = 0 & v \in \mathcal{V}_{\mathcal{D}}, \end{array} \right.$$

with $y = Su + y_f$. This allows for a reformulation of the optimality condition (2.12) as

$$(\beta u - \mathcal{K}p)^\top (w - u) \geq 0 \quad \forall w \in U_{\text{ad}}.$$

To summarize the previous investigations we state the following theorem:

THEOREM 2.3. *The pair $(u, y) \in U_{\text{ad}} \times H^1(\Gamma)$ is the unique global solution of (2.10a)–(2.10b) if and only if some adjoint state $p \in H_D^1(\Gamma)$ exists such that the system*

$$(2.13a) \quad y(v) = u_v, v \in \mathcal{V}_{\mathcal{D}}, \quad a(y, w) = (f, w)_{L^2(\Gamma)} \quad \forall w \in H_D^1(\Gamma),$$

$$(2.13b) \quad a(w, p) = (y - \bar{y}, w)_{L^2(\Gamma)} \quad \forall w \in H_D^1(\Gamma),$$

$$(2.13c) \quad (\beta u - \mathcal{K}p)^\top (w - u) \geq 0 \quad \forall w \in U_{\text{ad}}$$

is satisfied.

3. Discretization and error analysis for the optimal control problem. In [3] the authors study a finite element discretization of the variational problem (2.5) and their approach will also form the basis of our investigations. For the convenience of the reader we briefly repeat this discretization approach. On each edge $e \in \mathcal{E}$ we introduce an equidistant grid with vertices $\{v_a^e = v_0^e, v_1^e, \dots, v_{n_e}^e = v_b^e\}$. This induces a decomposition of the one-dimensional edge e into intervals $I_k^e := (v_k^e, v_{k+1}^e)$, $k = 0, \dots, n_e - 1$. The global finite element space is defined by

$$V_h := \{y_h \in C(\Gamma) : y_h|_{I_k^e} \in \mathcal{P}_1, k = 0, \dots, n_e - 1, e \in \mathcal{E}\}.$$

Here, the discretization parameter $h > 0$ is the maximal length of the intervals I_k^e , for $k = 0, \dots, n_e - 1$, $e \in \mathcal{E}$. By ψ_j^e , $j = 1, \dots, n_e - 1$, and ϕ_v , $v \in \mathcal{V}$, we denote the nodal basis functions of V_h satisfying $\psi_j^e(v_j^e) = 1$, for $j = 1, \dots, n_e - 1$, and $\phi_v(v) = 1$, for $v \in \mathcal{V}$. Hence, each function $y_h \in V_h$ can be represented by

$$(3.1) \quad y_h(x) = \sum_{e \in \mathcal{E}} \sum_{j=1}^{n_e-1} y_j^e \psi_j^e(x) + \sum_{v \in \mathcal{V}} y_v \phi_v(x).$$

The Galerkin approximation of (2.5) then reads:

Find $y_h \in V_h$, with $y_h(v) = u_v$ for $v \in \mathcal{V}_D$, such that

$$(3.2) \quad a(y_h, w_h) = (f, w_h)_{L^2(\Gamma)} \quad \forall w_h \in V_{h,D},$$

with $V_{h,D} := V_h \cap H_D^1(\Gamma)$.

Analogous to the continuous case, we split the discrete state y_h into $y_{u,h} := S_h u$, where

$$S_h : \mathbb{R}^{n_D} \rightarrow V_h$$

denotes the *discrete harmonic extension* (harmonic with respect to the Hamiltonian \mathcal{H}) of the Dirichlet data $u \in \mathbb{R}^{n_D}$ and a function $y_{f,h} \in V_{h,D}$ satisfying homogeneous Dirichlet conditions in the nodes \mathcal{V}_D . To be more precise, we have

$$(3.3) \quad y_{u,h}(v) = u_v, \quad v \in \mathcal{V}_D, \quad a(y_{u,h}, w_h) = 0 \quad \forall w_h \in V_{h,D},$$

$$(3.4) \quad a(y_{f,h}, w_h) = (f, w_h)_{L^2(\Gamma)} \quad \forall w_h \in V_{h,D},$$

and a simple computation shows that $y_h = y_{h,u} + y_{h,f}$. Finally, we can formulate the following discretized optimal control problem:

Minimize over $u_h \in \mathbb{R}^{n_D}, y_h \in V_h$,

$$(3.5a) \quad J_h(y_h, u_h) := \frac{1}{2} \|y_h - \bar{y}\|_{L^2(\Gamma)}^2 + \frac{\beta}{2} |u_h|_2^2,$$

subject to the constraints

$$(3.5b) \quad y_h(v) = u_{h,v}, \quad v \in \mathcal{V}_D, \quad a(y_h, v_h) = (f, v_h)_{L^2(\Gamma)} \quad \forall v_h \in V_{h,D},$$

$$(3.5c) \quad u_a \leq u_h \leq u_b.$$

As in the continuous case discussed in Section 2, we can derive a necessary and sufficient optimality condition. In the elliptic case the control is finite-dimensional and as a result does not require any discretization. First, we define the solution operator $P_h : L^2(\Gamma) \rightarrow V_{h,D}$ of the discretized adjoint equation by $P_h(y) = p_h$, with

$$(3.6) \quad a(w_h, p_h) = (y, w_h)_{L^2(\Gamma)} \quad \forall w_h \in V_{h,D}.$$

For $p_h = P_h(y)$, the discretized Kirchhoff-Neumann operator $\mathcal{K}_h : V_h \rightarrow \mathbb{R}^{n_D}$ is defined in a variational sense by

$$(3.7) \quad (\mathcal{K}_h p_h)(v) = a(\phi_v, p_h) - (y, \phi_v)_{L^2(\Gamma)} \quad \forall v \in \mathcal{V}_D.$$

LEMMA 3.1. *The adjoint of the operator $S_h : \mathbb{R}^{n_D} \rightarrow V_h$ can be represented by*

$$S_h^* = -\mathcal{K}_h \circ P_h : L^2(\Gamma) \rightarrow \mathbb{R}^{n_D}.$$

Proof. Let $y \in L^2(\Gamma)$, $z \in \mathbb{R}^{n_D}$, and $p_h = P_h(y)$ be arbitrary. We confirm, using the definitions of P_h , \mathcal{K}_h , and S_h as well as the properties $w_h := S_h z - \sum_{v \in \mathcal{V}_D} z_v \phi_v \in V_{h,D}$ and $p_h \in V_{h,D}$, that

$$\begin{aligned} (y, S_h z)_{L^2(\Gamma)} &= (y, S_h z - \sum_{v \in \mathcal{V}_D} z_v \phi_v)_{L^2(\Gamma)} + (y, \sum_{v \in \mathcal{V}_D} z_v \phi_v)_{L^2(\Gamma)} \\ &= a(S_h z - \sum_{v \in \mathcal{V}_D} z_v \phi_v, p_h) + (y, \sum_{v \in \mathcal{V}_D} z_v \phi_v)_{L^2(\Gamma)} \\ &= \sum_{v \in \mathcal{V}_D} z_v (-a(\phi_v, p_h) + (y, \phi_v)_{L^2(\Gamma)}) \\ &= - \sum_{v \in \mathcal{V}_D} z_v (\mathcal{K}_h p_h)(v) = -z^\top (\mathcal{K}_h p_h). \end{aligned}$$

This implies the desired result as $\mathcal{K}_h p_h = (\mathcal{K}_h \circ P_h)y$. \square

Analogous to the continuous case investigated in Section 2, we can derive an optimality system for (3.5a)–(3.5b) which reads:

Find $u_h \in U_{\text{ad}}$, $y_h \in V_h$, with $y_h(v) = u_{h,v} \forall v \in \mathcal{V}_{\mathcal{D}}$, $p_h \in V_{h,\mathcal{D}}$, such that

$$(3.8a) \quad a(y_h, w_h) = (f, w_h)_{L^2(\Gamma)} \quad \forall w_h \in V_{h,\mathcal{D}},$$

$$(3.8b) \quad a(w_h, p_h) = (y_h - \bar{y}, w_h)_{L^2(\Gamma)} \quad \forall w_h \in V_{h,\mathcal{D}},$$

$$(3.8c) \quad (\beta u_h - \mathcal{K}_h p_h)^\top (w_h - u_h) = 0 \quad \forall w_h \in U_{\text{ad}}.$$

This is a discrete version of the optimality system (2.13). Note that $\mathcal{K}_h(p_h) \neq \mathcal{K}(p_h)$. Using the exact Kirchhoff-Neumann operator \mathcal{K} in the discrete optimality system is basically possible, but as we have seen, only the variational Kirchhoff-Neumann map \mathcal{K}_h yields the favorable property that the approaches “first optimize then discretize” and “first discretize then optimize” lead to the same discrete optimality system.

In the remainder of this section we study discretization error estimates for the approximation of the control $|u - u_h|$ and the states $\|y - y_h\|_{H^1(\Omega)}$. The most technical part of the error analysis is the proof of an estimate for the approximate Kirchhoff-Neumann operator \mathcal{K}_h . The proof is based on a duality argument and requires the boundedness of the discrete control-to-state operator.

LEMMA 3.2. *Assume $c_0 \geq \underline{c} > 0$. Then the operator S_h satisfies the boundedness estimate*

$$(3.9) \quad \|S_h u\|_{H^1(\Gamma)} \leq 2 \left(\frac{M}{\gamma} \right)^{3/2} C_\Gamma |u|_2$$

for arbitrary $u \in \mathbb{R}^{n_{\mathcal{D}}}$.

Proof. Recall the extension $y_u \in H_{\mathcal{D}}^1(\Gamma)$ of the Dirichlet data $u \in \mathbb{R}^{n_{\mathcal{D}}}$ introduced in the proof of Lemma 2.1. As y_u is edge-wise linear, there holds $y_u \in V_h$. Due to coercivity (2.7) and $V_h \subset H^1(\Omega)$ as well as $S_h u - y_u \in V_{h,\mathcal{D}}$, we then deduce that

$$\gamma \|S_h u\|_{H^1(\Gamma)}^2 \leq a(S_h u, S_h u) = a(S_h u, y_u) \leq M \|S_h u\|_{H^1(\Gamma)} \|y_u\|_{H^1(\Gamma)}.$$

Insertion of (2.6) yields the property (3.9). \square

As an important consequence, we also obtain the boundedness of the adjoint S_h^* and thus for the discrete Kirchhoff-Neumann operator \mathcal{K}_h .

LEMMA 3.3. *For arbitrary $y \in L^2(\Gamma)$ there holds the stability estimate*

$$|S_h^* y|_2 = |\mathcal{K}_h(P_h y)|_2 \leq 2 \left(\frac{M}{\gamma} \right)^{3/2} C_\Gamma \|y\|_{L^2(\Gamma)}.$$

Proof. First, we apply the definitions of \mathcal{K}_h from (3.7) and P_h from (3.6), taking into account that $\sum_{v \in \mathcal{V}_{\mathcal{D}}} u_v \phi_v - S_h u \in V_{h,\mathcal{D}}$, and obtain

$$\begin{aligned} |\mathcal{K}_h(P_h y)|_2 &= \sup_{\substack{u \in \mathbb{R}^{n_{\mathcal{D}}} \\ |u|_2=1}} |u^\top \mathcal{K}_h(P_h y)| \\ &= \sup_{\substack{u \in \mathbb{R}^{n_{\mathcal{D}}} \\ |u|_2=1}} \left| \sum_{v \in \mathcal{V}_{\mathcal{D}}} u_v [a(\phi_v, P_h y) - (y, \phi_v)_{L^2(\Gamma)}] \right| \\ &= \sup_{\substack{u \in \mathbb{R}^{n_{\mathcal{D}}} \\ |u|_2=1}} |a(S_h u, P_h y) - (y, S_h u)_{L^2(\Gamma)}|. \end{aligned}$$

The first term on the right-hand side vanishes due to the definition of S_h (see (3.3)) and $P_h y \in V_{h,D}$. The last term can be bounded by means of the Cauchy-Schwarz inequality and the stability estimate for S_h proved in Lemma 3.2. \square

Now, we are in the position to derive the following general error estimate:

LEMMA 3.4. *Let $(u, y) \in \mathbb{R}^{n_D} \times H^1(\Gamma)$ and $(u_h, y_h) \in \mathbb{R}^{n_D} \times V_h$ be the solutions of (2.10) and (3.5), respectively. Then the error estimate*

$$\beta |u - u_h|_2 \leq 2 \left(\frac{M}{\gamma} \right)^{3/2} C_\Gamma \|y - \tilde{y}_h\|_{L^2(\Gamma)} + |\mathcal{K}p - \mathcal{K}_h \tilde{p}_h|_2$$

is satisfied, where p is the adjoint state corresponding to u , and $\tilde{y}_h = S_h u + y_{f,h}$ and $\tilde{p}_h = P_h(y - \bar{y})$ are the Ritz projections of y and p .

Proof. Summation of the optimality conditions (2.13c) with $w = u_h$ and (3.8c) with $w_h = u$ and using $y_h = S_h u_h$ as well as $p_h = P_h(S_h u_h + y_{f,h} - \bar{y})$ leads to

$$\begin{aligned} \beta |u - u_h|_2^2 &\leq (\mathcal{K}p - \mathcal{K}_h p_h, u - u_h) \\ &= (\mathcal{K}p - \mathcal{K}_h(P_h(y - \bar{y})) + \mathcal{K}_h(P_h(y - (S_h u + y_{f,h}))) \\ &\quad + \mathcal{K}_h(P_h(S_h(u - u_h))))^\top (u - u_h) \\ &= (\mathcal{K}p - \mathcal{K}_h \tilde{p}_h)^\top (u - u_h) + (S_h^*(y - \tilde{y}_h))^\top (u - u_h) \\ &\quad - (S_h^*(S_h(u - u_h)))^\top (u - u_h). \end{aligned} \tag{3.10}$$

Note that in the first step, we simply introduced the intermediate functions $\mathcal{K}_h(P_h y)$ and $\mathcal{K}_h(P_h(S_h u))$. In the last step we inserted the Ritz projections $\tilde{y}_h = S_h u + y_{f,h}$ and $\tilde{p}_h = P_h(y - \bar{y})$ of y and p , respectively, as well as $S_h^* = -\mathcal{K}_h \circ P_h$.

For the first two terms on the right-hand side of (3.10), we get with the Cauchy-Schwarz inequality

$$\begin{aligned} (\mathcal{K}p - \mathcal{K}_h \tilde{p}_h)^\top (u - u_h) &\leq |\mathcal{K}p - \mathcal{K}_h \tilde{p}_h|_2 |u - u_h|_2, \\ S_h^*(y - \tilde{y}_h)^\top (u - u_h) &\leq 2 \left(\frac{M}{\gamma} \right)^{3/2} C_\Gamma \|y - \tilde{y}_h\|_{L^2(\Gamma)} |u - u_h|_2. \end{aligned}$$

In the latter estimate we used the stability properties of S_h^* proved in Lemma 3.3. The last term on the right-hand side of (3.10) is non-positive and can be neglected. \square

It remains to derive error estimates for the two terms on the right-hand side of the general estimate derived in Lemma 3.4. The first term is a simple $L^2(\Gamma)$ -error, and an estimate can be simply obtained from the $H^1(\Gamma)$ -error estimate

$$|y - \tilde{y}_h|_{H^1(\Gamma)} \leq c h |y|_{H^2(\Gamma)} \tag{3.11}$$

proved in [3, Theorem 3.2] and an application of the Aubin-Nitsche method. These arguments imply

$$\|y - \tilde{y}_h\|_{L^2(\Gamma)} \leq c \sqrt{\frac{M^3}{\gamma}} h^2 |y|_{H^2(\Gamma)}. \tag{3.12}$$

In both estimates the constant c is independent of y , u , f , and Γ . The second term, namely the error estimate for the discrete Kirchhoff-Neumann operator, can also be deduced by a duality argument.

THEOREM 3.5. *For given $y \in L^2(\Gamma)$, let $p \in H_D^1(\Gamma)$ be the solution of*

$$a(w, p) = (y, w)_{L^2(\Gamma)} \quad \forall w \in H_D^1(\Gamma),$$

and denote by $\tilde{p}_h \in V_{h,D}$ its Ritz-projection, i.e., $a(w_h, p - \tilde{p}_h) = 0$ for all $w_h \in V_{h,D}$. Then the error estimate

$$|\mathcal{K}p - \mathcal{K}_h \tilde{p}_h|_2 \leq c \frac{M}{\gamma} C_\Gamma h^2 \|y\|_{L^2(\Gamma)}$$

holds with a constant $c > 0$ independent of h, y, p , and Γ .

Proof. Let $u \in \mathbb{R}^{n_D}$ be an arbitrary vector. Applying the integration-by-parts formula (2.3), which implies

$$(\mathcal{K}p)^\top u = \sum_{v \in \mathcal{V}_D} u_v [a(\phi_v, p) - (y, \phi_v)_{L^2(\Gamma)}],$$

the definition of \mathcal{K}_h from (3.7), the Galerkin orthogonality using $\sum_{v \in \mathcal{V}_D} u_v \phi_v - S_h u \in V_{h,D}$, and the definition of S yield

$$\begin{aligned} (\mathcal{K}p - \mathcal{K}_h \tilde{p}_h)^\top u &= \sum_{v \in \mathcal{V}_D} u_v a(\phi_v, p - \tilde{p}_h) = a(S_h u, p - \tilde{p}_h) = a(S_h u - S u, p - \tilde{p}_h) \\ &\leq c \|p - \tilde{p}_h\|_{H^1(\Gamma)} \|S u - S_h u\|_{H^1(\Gamma)} \leq c \frac{M}{\gamma} C_\Gamma h^2 \|y\|_{L^2(\Gamma)} |u|_2. \end{aligned}$$

The last step of the previous estimate follows from the $H^1(\Gamma)$ -estimate (3.11) and Lemma 2.1. The assertion finally follows from

$$|\mathcal{K}p - \mathcal{K}_h \tilde{p}_h|_2 = \sup_{\substack{u \in \mathbb{R}^{n_D} \\ |u|_2=1}} |(\mathcal{K}p - \mathcal{K}_h \tilde{p}_h)^\top u| \leq c \frac{M}{\gamma} C_\Gamma h^2 \|y\|_{L^2(\Gamma)}. \quad \square$$

Now we are in a position to state the main result of this section.

THEOREM 3.6. *Let $f, \bar{y} \in L^2(\Gamma)$, and $\beta > 0$ be arbitrary. The solutions $(u, y) \in \mathbb{R}^{n_D} \times H^1(\Gamma)$ and $(u_h, y_h) \in \mathbb{R}^{n_D} \times V_h$ of (2.10) and (3.5), respectively, satisfy the error estimate*

$$\begin{aligned} |u - u_h|_2 + C_\Gamma^{-1} \left(\|y - y_h\|_{L^2(\Gamma)} + h \|y - y_h\|_{H^1(\Gamma)} \right) \\ \leq c(M, \gamma) \beta^{-1} C_\Gamma h^2 \left(\|f\|_{L^2(\Gamma)} + C_\Gamma |u|_2 + \|y - \bar{y}\|_{L^2(\Gamma)} \right) \end{aligned}$$

with a constant $c > 0$ independent of h .

Proof. The estimate for the control follows after insertion of the bound (3.12) and the estimate of Theorem 3.5 into the one from Lemma 3.4. To obtain an estimate for the state, we apply the triangle inequality

$$\|y - y_h\| \leq \|y - (S_h u + y_{f,h})\| + \|S_h(u - u_h)\|,$$

where we used the decomposition $y_h = S_h u_h + y_{f,h}$; compare (3.3)–(3.4). Note that $S_h u + y_{f,h}$ is the Ritz projection of y , and thus, the first term on the right-hand side can be bounded by means of (3.11) for the $L^2(\Gamma)$ -norm and (3.12) for the $H^1(\Gamma)$ -norm. An estimate for the second term follows from the stability of S_h shown in Lemma 3.2 and the estimate derived for the controls. \square

4. Structure of the system matrices and preconditioning. We now discuss the structure of the discretized equations based on the considerations from the previous section as well as the results given in [3]. The finite element discretization may also be interpreted as an extension of the original graph Γ to the finite element nodes in the interior of each edge. This yields an *extended graph*, which provides a lot of structure that we may exploit for both the assembly of the finite element system matrices and the preconditioning of the saddle point system (3.8).

4.1. Assembly of finite element matrices. We use a numbering of the nodes such that the *interior nodes* come first followed by the original graph nodes; cf. (4.1). We note that the incidence matrix for the *interior nodes* of an edge $e \in \mathcal{E}$ is structured and can be written as

$$\mathbf{E}_e = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(n_e-1) \times n_e}.$$

The incidence matrix for all interior nodes on all edges is then given by

$$\mathbf{E}_i = \mathbf{blkdiag}(\{\mathbf{E}_e\}_{e \in \mathcal{E}}) \in \mathbb{R}^{m(n_e-1) \times m n_e},$$

where n_e is the same for all edges for simplicity. To also incorporate the original graph nodes, we now introduce the matrices

$$\mathbf{E}^+ = \frac{1}{2}(\mathbf{E} + |\mathbf{E}|), \quad \mathbf{E}^- = \frac{1}{2}(\mathbf{E} - |\mathbf{E}|),$$

with $|\mathbf{E}|$ being the component-wise absolute value of the incidence matrix \mathbf{E} of the original graph. We now establish the incidence matrix for the extended graph based on the original graph nodes. For this consider

$$\widehat{\mathbf{E}}_j^+ = \mathbf{E}_j^+ \otimes \underbrace{[1, 0, \dots, 0]}_{\in \mathbb{R}^{1 \times n_{e_j}}} = [\mathbf{E}_j^+, \mathbf{0}, \dots, \mathbf{0}] \in \mathbb{R}^{n \times n_{e_j}},$$

which allows us to incorporate outgoing edges, where \mathbf{E}_j^+ indicates the j -th column of the matrix \mathbf{E}^+ and $n_{e_j} = n_e$ is the number of internal nodes on the edge e_j . Here, the index j runs from 1 to $m = |\mathcal{E}|$. For the incoming edges we use

$$\widehat{\mathbf{E}}_j^- = \mathbf{E}_j^- \otimes [0, 0, \dots, 1] = [\mathbf{0}, \dots, \mathbf{0}, \mathbf{E}_j^-] \in \mathbb{R}^{n \times n_e}.$$

We see that due to the incorporation of the interior nodes, this somewhat stretches the incidence matrix of the original graph. As a result the part of the incidence matrix representing the original nodes becomes

$$\mathbf{E}_v = \left[\widehat{\mathbf{E}}_1^+ + \widehat{\mathbf{E}}_1^-, \widehat{\mathbf{E}}_2^+ + \widehat{\mathbf{E}}_2^-, \dots, \widehat{\mathbf{E}}_m^+ + \widehat{\mathbf{E}}_m^- \right] \in \mathbb{R}^{n \times m n_e}.$$

We then obtain the incidence matrix of the extended graph as

$$\widetilde{\mathbf{E}} = \begin{bmatrix} \mathbf{E}_i \\ \mathbf{E}_v \end{bmatrix} \in \mathbb{R}^{m(n_e-1)+n \times m n_e}.$$

In order to build the finite element matrices, we incorporate the mesh-parameter h_e via

$$\begin{aligned} \mathbf{W}_E &= \mathbf{blkdiag} \left(\{ h_e^{-1} \mathbf{I}_{n_e} \}_{e \in \mathcal{E}} \right) \in \mathbb{R}^{(n_i+m) \times (n_i+m)}, \\ \widehat{\mathbf{W}}_E &= \mathbf{blkdiag} \left(\{ h_e \mathbf{I}_{n_e} \}_{e \in \mathcal{E}} \right) \in \mathbb{R}^{(n_i+m) \times (n_i+m)}, \end{aligned}$$

with $n_i = \sum_{e \in \mathcal{E}} (n_e - 1)$, and in the case that all edges have the same number of interior nodes, there holds $n_i = m(n_e - 1)$. As was shown in [3], the stiffness and mass matrices \mathbf{A} and \mathbf{M} have the following structure

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} \mathbf{E}_i \\ \mathbf{E}_v \end{bmatrix} \mathbf{W}_E \begin{bmatrix} \mathbf{E}_i^\top & \mathbf{E}_v^\top \end{bmatrix} = \begin{bmatrix} \mathbf{E}_i \mathbf{W}_E \mathbf{E}_i^\top & \mathbf{E}_i \mathbf{W}_E \mathbf{E}_v^\top \\ \mathbf{E}_v \mathbf{W}_E \mathbf{E}_i^\top & \mathbf{E}_v \mathbf{W}_E \mathbf{E}_v^\top \end{bmatrix}, \\ \mathbf{M} &= \frac{1}{6} \left(\left| \widetilde{\mathbf{E}} \right| \widehat{\mathbf{W}}_E \left| \widetilde{\mathbf{E}} \right|^\top + \mathbf{diag} \left(\left\{ \left(\left| \widetilde{\mathbf{E}} \right| \widehat{\mathbf{W}}_E \left| \widetilde{\mathbf{E}} \right|^\top \right)_{i,i} \right\}_{i=1}^{n_i+m} \right) \right). \end{aligned}$$

With the discretization of both the mass and the stiffness terms, we are now able to obtain the matrix representation of the optimization problem, where the mass and stiffness matrix are split according to free and Dirichlet-control variables. The mass matrix incorporating the term c_0 can be assembled in a similar fashion (cf. [3]), and we refer to it as \mathbf{M}_{c_0} . Finally, the system matrix representing the discrete bilinear form on the left-hand side of (3.2) is denoted by

$$\mathbf{K} = \mathbf{A} + \mathbf{M}_{c_0}.$$

4.2. The discrete optimality system. We want to further discuss the discrete optimality system (3.8) from a linear algebra perspective aiming at efficient solvers for the optimal control problem. To simplify the presentation we omit the control constraints so that the optimality condition turns into a system of linear equations. The control-constrained case can be discussed in a similar way using the techniques introduced in [65]. We start with the ansatz (3.1), which allows for a representation of a function $y_h \in V_h$ by the coefficient vectors $\mathbf{y} \in \mathbb{R}^N$, $N = m(n_e - 1) + n_K + n_D$, which are sorted in the form

$$(4.1) \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_I \\ \mathbf{y}_K \\ \mathbf{y}_D \end{bmatrix} \hat{=} \begin{bmatrix} \text{coefficients related to } \psi_j^e, j = 1, \dots, n_e - 1, e \in \mathcal{E} \\ \text{coefficients related to } \phi_v, v \in \mathcal{V}_K \\ \text{coefficients related to } \phi_v, v \in \mathcal{V}_D \end{bmatrix}.$$

In the following, we directly incorporate the Dirichlet conditions by $\mathbf{y}_D = \mathbf{u}$. Furthermore, we may split the matrices \mathbf{K} and \mathbf{M} into

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{II} & \mathbf{K}_{IK} & \mathbf{K}_{ID} \\ \mathbf{K}_{KI} & \mathbf{K}_{KK} & 0 \\ \mathbf{K}_{DI} & 0 & \mathbf{K}_{DD} \end{bmatrix} \quad \text{and} \quad \mathbf{M} = \begin{bmatrix} \mathbf{M}_{II} & \mathbf{M}_{IK} & \mathbf{M}_{ID} \\ \mathbf{M}_{KI} & \mathbf{M}_{KK} & 0 \\ \mathbf{M}_{DI} & 0 & \mathbf{M}_{DD} \end{bmatrix},$$

respectively. Note that we assumed $n_e \geq 1$ to get the zero blocks. Moreover, under this assumption, \mathbf{K}_{KK} , \mathbf{K}_{DD} , \mathbf{M}_{KK} , and \mathbf{M}_{DD} are diagonal matrices. In order to get a description of the discrete optimality system (3.8) in a matrix-vector notation, we distinguish only among Dirichlet nodes and free nodes and thus define

$$\mathbf{y}_F = \begin{bmatrix} \mathbf{y}_I \\ \mathbf{y}_K \end{bmatrix}, \quad \mathbf{K}_{FF} = \begin{bmatrix} \mathbf{K}_{II} & \mathbf{K}_{IK} \\ \mathbf{K}_{KI} & \mathbf{K}_{KK} \end{bmatrix}, \quad \mathbf{K}_{FD} = \begin{bmatrix} \mathbf{K}_{ID} \\ 0 \end{bmatrix}, \quad \mathbf{K}_{DF} = \mathbf{K}_{FD}^\top.$$

Analogous definitions are used for the mass matrix \mathbf{M} .

Furthermore, we define the load vector \mathbf{f} by $\mathbf{f}^\top \mathbf{v} = (f, v_h)_{L^2(\Gamma)}$ and the vector $\bar{\mathbf{y}}$ by $\bar{\mathbf{y}}^\top \mathbf{v} = (\bar{y}, v_h)_{L^2(\Gamma)}$. These vectors are decomposed as well in the form

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}_F \\ \mathbf{f}_D \end{bmatrix}, \quad \bar{\mathbf{y}} = \begin{bmatrix} \bar{\mathbf{y}}_F \\ \bar{\mathbf{y}}_D \end{bmatrix}.$$

With the previously introduced matrices and vectors we can write the optimality system (3.8) in the form

$$(4.2) \quad \begin{bmatrix} \mathbf{M}_{FF} & \mathbf{M}_{FD} & \mathbf{K}_{FF}^\top \\ \mathbf{M}_{DF} & \mathbf{M}_{DD} + \beta \mathbf{I} & \mathbf{K}_{FD}^\top \\ \mathbf{K}_{FF} & \mathbf{K}_{FD} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{y}_F \\ \mathbf{u} \\ \mathbf{p}_F \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{y}}_F \\ \bar{\mathbf{y}}_D \\ \mathbf{f}_F \end{bmatrix}.$$

Note that the sign of the adjoint state in this formulation is different from that in (3.8). This results in a symmetric system matrix.

In fact, the system (4.2) is a saddle point or KKT matrix [6, 17]. For complex networks with many connections it is infeasible to work with direct solvers [14] due to the higher complexity and fill-in issues. While one could use non-standard conjugate gradient methods [29, 55, 59], we here employ MINRES [48] as a tailored scheme for symmetric and indefinite matrices or GMRES [58] for the case of a nonsymmetric preconditioned systems. Of course, the performance of this scheme will rely on the distribution of the eigenvalues, and we need to improve the performance by introducing a suitably chosen preconditioner \mathbf{P} . Its design is discussed in the next part.

4.3. Preconditioning. It is well known [44] that an ideal preconditioner for saddle point systems is given by

$$\mathbf{P}_{\text{ideal}} = \begin{bmatrix} \mathbf{M}_{FF} & \mathbf{M}_{FD} & \mathbf{0} \\ \mathbf{M}_{DF} & \mathbf{M}_{DD} + \beta \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S} \end{bmatrix}$$

with the Schur-complement defined as

$$\mathbf{S} = [\mathbf{K}_{FF} \quad \mathbf{K}_{FD}] \begin{bmatrix} \mathbf{M}_{FF} & \mathbf{M}_{FD} \\ \mathbf{M}_{DF} & \mathbf{M}_{DD} + \beta \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{K}_{FF}^\top \\ \mathbf{K}_{FD}^\top \end{bmatrix}.$$

While this preconditioner is attractive in producing an optimally clustered spectrum of the preconditioned matrix, it is in general very expensive to apply as the matrix \mathbf{S} is typically dense. A more practical choice is obtained when we consider a block-diagonal approximation

$$\mathbf{P}_{\text{ideal}} \approx \begin{bmatrix} \mathbf{M}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_1 \end{bmatrix}$$

with

$$\begin{bmatrix} \mathbf{M}_{FF} & \mathbf{M}_{FD} \\ \mathbf{M}_{DF} & \mathbf{M}_{DD} + \beta \mathbf{I} \end{bmatrix} \approx \begin{bmatrix} \mathbf{M}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2 \end{bmatrix} \text{ and } \mathbf{S}_1 \approx \mathbf{S}.$$

To make these approximation more precise, we first focus on approximating the mass matrix block and note that again an ideal preconditioner according to [44] is

$$\begin{bmatrix} \mathbf{M}_{FF} & \mathbf{M}_{FD} \\ \mathbf{M}_{DF} & \mathbf{M}_{DD} + \beta \mathbf{I} \end{bmatrix} \approx \begin{bmatrix} \mathbf{M}_{FF} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_M \end{bmatrix}, \text{ with } \mathbf{S}_M = \mathbf{M}_{DD} + \beta \mathbf{I} - \mathbf{M}_{DF} \mathbf{M}_{FF}^{-1} \mathbf{M}_{FD}.$$

Many authors [49, 50, 70] have studied the behavior of the iterative solver with respect to a changing regularization parameter by designing a preconditioner that proved robust to the varying regularization parameter. We here want to study the behavior of the eigenvalues with respect to β . For this we recall from [57] the eigenvalue bounds for a general saddle point problem,

$$\begin{aligned} \text{sp}(\mathcal{A}) \subset & \left[\frac{1}{2} \left(\lambda_{\min}(\mathbf{A}) - \sqrt{\lambda_{\min}(\mathbf{A})^2 + 4\sigma_{\max}(\mathbf{B})} \right), \right. \\ & \left. \frac{1}{2} \left(\lambda_{\max}(\mathbf{A}) - \sqrt{\lambda_{\max}(\mathbf{A})^2 + 4\sigma_{\min}(\mathbf{B})} \right) \right] \\ & \cup \left[\lambda_{\min}(\mathbf{A}), \frac{1}{2} \left(\lambda_{\max}(\mathbf{A}) + \sqrt{\lambda_{\max}(\mathbf{A})^2 + 4\sigma_{\min}(\mathbf{B})} \right) \right], \end{aligned}$$

where in standard saddle point notation \mathcal{A} refers to the saddle point matrix in (4.2) with the blocks

$$\mathbf{A} = \begin{bmatrix} \mathbf{M}_{FF} & \mathbf{M}_{FD} \\ \mathbf{M}_{DF} & \mathbf{M}_{DD} + \beta \mathbf{I} \end{bmatrix}, \quad \mathbf{B} = [\mathbf{K}_{FF} \quad \mathbf{K}_{FD}].$$

From the eigenvalue bounds we see that the dependence on β comes from the eigenvalues of \mathbf{A} , which we want to study now in some more detail considering that this matrix is symmetric. There holds

$$\mathbf{x}^T \left(\begin{bmatrix} \mathbf{M}_{FF} & \mathbf{M}_{FD} \\ \mathbf{M}_{DF} & \mathbf{M}_{DD} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \beta \mathbf{I} \end{bmatrix} \right) \mathbf{x} = \mathbf{x}^T \begin{bmatrix} \mathbf{M}_{FF} & \mathbf{M}_{FD} \\ \mathbf{M}_{DF} & \mathbf{M}_{DD} \end{bmatrix} \mathbf{x} + \mathbf{x}^T \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \beta \mathbf{I} \end{bmatrix} \mathbf{x},$$

and as a result we see that the eigenvalues are bounded from above and below by the sum of the maximal and minimal eigenvalues of the individual matrices, respectively. This yields

$$\lambda_{\min}(\mathbf{M}) \leq \lambda(\mathbf{A}) \leq \lambda_{\max}(\mathbf{M}) + \beta.$$

Obviously, the regularization parameter β has no influence on the smallest eigenvalue of the $(1, 1)$ -block of the saddle point matrix, and only for large values of β the maximal eigenvalues of \mathbf{A} are influenced.

For small values of β our preconditioning strategy needs to focus on the robustness with respect to the mesh-parameter h . To verify this claim we measure the estimated condition numbers of the saddle point matrix \mathcal{A} with respect to a changing regularization and mesh parameter for the benchmark problem on the L-shaped FD graph from Section 6.1. The computational results reported in Table 4.1 confirm the expected behavior.

It remains to construct suitable approximations of the Schur complement matrices \mathbf{S} and \mathbf{S}_M . First, we approximate \mathbf{S} by

$$\begin{aligned} \mathbf{S} & \approx [\mathbf{K}_{FF} \quad \mathbf{K}_{FD}] \begin{bmatrix} \mathbf{M}_{FF} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_M \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{K}_{FF}^\top \\ \mathbf{K}_{FD}^\top \end{bmatrix} \\ & = \mathbf{K}_{FF} \mathbf{M}_{FF}^{-1} \mathbf{K}_{FF}^\top + \mathbf{K}_{FD} \mathbf{S}_M^{-1} \mathbf{K}_{FD}^\top. \end{aligned}$$

TABLE 4.1

Condition numbers for a graph generated from a finite difference discretization network on an L-shaped domain in dependence of β and the number of unknowns n_{DOF} of the FE discretization.

N_{DOF}	244	2484	22132
β			
10^{-1}	8.7e+01	4.3e+02	2.1e+03
10^{-2}	9.2e+01	4.1e+02	2.0e+03
10^{-3}	8.8e+01	5.0e+02	2.2e+03
10^{-4}	9.3e+01	4.6e+02	1.9e+03
10^{-5}	9.5e+01	5.0e+02	2.0e+03
10^{-6}	9.2e+01	4.5e+02	2.1e+03
10^{-7}	9.4e+01	3.9e+02	2.0e+03

This approximation is still not very practical to work with since we would have to invert a sum of matrix products. An efficient approximation of this sum follows from the choices $\mathbf{M}_{FF} \approx \mathbf{D}_M := \text{diag}(\mathbf{M}_{FF})$ and $\mathbf{S}_M \approx \mathbf{D}_{S_M} := \mathbf{M}_{DD} + \beta \mathbf{I} - \mathbf{M}_{DF} \mathbf{D}_M^{-1} \mathbf{M}_{FD}$, which yield

$$\mathbf{K}_{FF} \mathbf{M}_{FF}^{-1} \mathbf{K}_{FF}^\top + \mathbf{K}_{FD} \mathbf{S}_M^{-1} \mathbf{K}_{FD}^\top \approx \mathbf{K}_{FF} \mathbf{D}_M^{-1} \mathbf{K}_{FF}^\top + \mathbf{K}_{FD} \mathbf{D}_{S_M}^{-1} \mathbf{K}_{FD}^\top.$$

Still, this approximation is difficult to work with as it involves a sum of terms. As we have illustrated earlier, the dependence on β is rather benign, and the simple approximations

$$\mathbf{S} \approx \mathbf{K}_{FF} \mathbf{D}_M^{-1} \mathbf{K}_{FF}^\top \quad \text{or} \quad \mathbf{S} \approx \mathbf{K}_{FF} \mathbf{M}_{FF}^{-1} \mathbf{K}_{FF}^\top$$

perform very well in numerical experiments.

Alternatively, we want to construct a more sophisticated preconditioner based on a matching approach [49, 51]

$$\mathbf{K}_{FF} \mathbf{D}_M^{-1} \mathbf{K}_{FF}^\top + \mathbf{K}_{FD} \mathbf{D}_{S_M}^{-1} \mathbf{K}_{FD}^\top \approx (\mathbf{K}_{FF} + \mathbf{N}) \mathbf{D}_M^{-1} (\mathbf{K}_{FF} + \mathbf{N})^\top,$$

which requires

$$\mathbf{N} \mathbf{D}_M^{-1} \mathbf{N}^\top = \mathbf{K}_{FD} \mathbf{D}_{S_M}^{-1} \mathbf{K}_{FD}^\top.$$

This equivalence would hold true for

$$\mathbf{N} = (\mathbf{K}_{FD} \mathbf{D}_{S_M}^{-1} \mathbf{K}_{FD}^\top)^{1/2} \mathbf{D}_M^{1/2}.$$

However, since we do not want to take the square root of a possibly very large matrix, we approximate the Schur-complement yielding

$$\mathbf{N} = (\mathbf{D}_{KDK})^{1/2} \mathbf{D}_M^{1/2},$$

where \mathbf{D}_{KDK} is the lumped version of $\mathbf{K}_{FD} \mathbf{D}_{S_M}^{-1} \mathbf{K}_{FD}^\top$. While this approach fits well within the preconditioning framework for a symmetric solver like MINRES [48], it would be possible to avoid this using a non-symmetric matching strategy [49]. Let us look at the eigenvalue distribution of the preconditioned saddle point matrix.

In Figure 4.1 we illustrate the eigenvalue distribution of the overall preconditioners equipped with the different Schur-complement approximations, where we have used

$$\mathbf{P}_i = \begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{S_M} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_i \end{bmatrix}$$

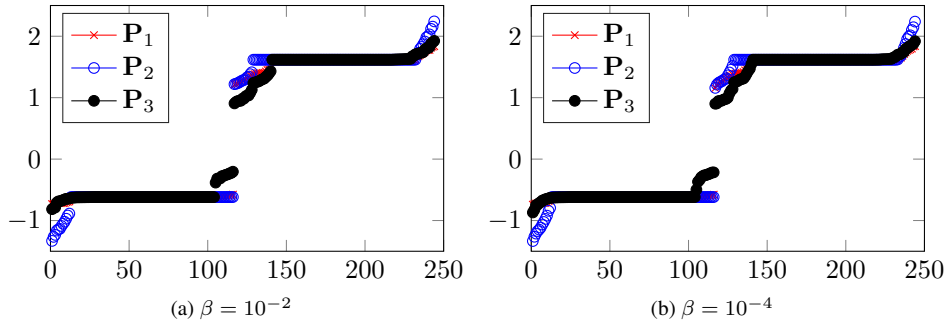


FIG. 4.1. Eigenvalue plots of the preconditioned KKT matrix.

with

$$\begin{aligned}
 \mathbf{S}_1 &= \mathbf{K}_{FF}\mathbf{M}^{-1}\mathbf{K}_{FF}^\top + \mathbf{K}_{FD}\mathbf{D}_{SM}^{-1}\mathbf{K}_{FD}^\top \\
 \mathbf{S}_2 &= \mathbf{K}_{FF}\mathbf{M}^{-1}\mathbf{K}_{FF}^\top \\
 \mathbf{S}_3 &= (\mathbf{K}_{FF} + \mathbf{N})\mathbf{D}_M^{-1}(\mathbf{K}_{FF} + \mathbf{N})^\top.
 \end{aligned}$$

The plot shows that there is no visible dependence on the parameter β when it comes to the different preconditioners. As a result, we settle for the simplest one, \mathbf{P}_2 , which will be the preconditioner of choice for the remainder of this paper. In the case of control constraints as discussed earlier, the linear systems maintain their saddle point structure, and preconditioning remains essential. We refrain from a detailed discussion here and note that a combination of the preconditioner proposed in this section and the technique used in [65] will lead to an efficient method for the control-constrained case as well.

5. Dirichlet control problems with parabolic state equation.

5.1. Analysis of the continuous optimal control problem. As an extension of the results from the previous sections, we study an optimal control problem governed by a parabolic state equation. As a model problem, we consider

$$(5.1) \quad \text{Minimize } J(y, u) := \frac{1}{2} \int_0^T \|y(t) - \bar{y}(t)\|_{L^2(\Gamma)}^2 dt + \frac{\beta}{2} \int_0^T |u(t)|_2^2 dt$$

over $y \in L^2(I; L^2(\Gamma))$ and $u \in L^2(I; \mathbb{R}^{n_D})$, with a desired state $\bar{y} \in L^2(I; L^2(\Gamma))$ and the time horizon $I := (0, T)$. The function spaces are defined by the usual Bochner integrals [19, Chapter 7]. The state equation in the current setting reads

$$(5.2) \quad \begin{cases} \partial_t y - y'' + c_0 y = f & \text{in } I \times e, e \in \mathcal{E}, \\ y = u_v & \text{on } I \times \mathcal{V}_D, \\ \mathcal{K}y = 0 & \text{on } I \times \mathcal{V}_K, \\ y(0, \cdot) = 0 & \text{in } \Gamma. \end{cases}$$

Due to the low a priori regularity of the control, $u \in L^2(I; \mathbb{R}^{n_D})$, the state equation is not well-posed in the weak sense. Thus, as usual for Dirichlet control problems with controls in L^2 , we have to consider the state equation in a very weak sense. We follow the approach

from [35] and define

$$(5.3) \quad \left\{ \begin{array}{l} \text{Find } y \in L^2(I; L^2(\Gamma)) \text{ with } y(0) = 0: \\ \langle \partial_t y(t), w \rangle_{X^* \times X} + (y(t), -w'' + c_0 w)_{L^2(\Gamma)} \\ \qquad \qquad \qquad = (f(t), w)_{L^2(\Gamma)} - \sum_{v \in \mathcal{V}_{\mathcal{D}}} u_v(t) (\mathcal{K}w)(v), \\ \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \text{for all } w \in X, \end{array} \right.$$

with

$$X := \{w \in H^2(\Gamma) \cap H_D^1(\Gamma) : (\mathcal{K}w)(v) = 0 \quad \text{for all } v \in \mathcal{V}_{\mathcal{K}}\}.$$

LEMMA 5.1. *For each $f \in L^2(I; L^2(\Gamma))$ and $u \in L^2(I; \mathbb{R}^{n_{\mathcal{D}}})$, the problem (5.3) has a unique solution $y \in L^2(I; L^2(\Gamma)) \cap H^1(I; X^*) \cap C(\bar{I}; H_D^1(\Gamma)^*)$ depending continuously on the input data.*

We omit the proof as the result follows from the arguments in [35, Theorem 2.1]. Similar results for a slightly different weak formulation with ∂_t also acting on the test function can be found in [22, Lemma 1]. We also refer to [2, 40], where the existence of very weak solutions of elliptic equations is discussed.

As in the elliptic case we introduce the control-to-state operator $S : L^2(I; \mathbb{R}^{n_{\mathcal{D}}}) \rightarrow L^2(I; L^2(\Gamma))$ mapping a control u to the solution of (5.3) with $f \equiv 0$. The state can be represented by

$$y = Su + y_f$$

with $y_f \in L^2(I; H_D^1(\Gamma)) \cap H^1(I; H_D^1(\Gamma)^*) \hookrightarrow C(\bar{I}; L^2(\Gamma))$ being the weak solution of (5.3) with $u \equiv 0$.

Existence and uniqueness of a solution of (5.1)–(5.3) follow directly from the arguments of [35, Proposition 3.1]. The necessary and sufficient optimality condition reads as follows.

THEOREM 5.2. *Let $f, \bar{y} \in L^2(I; L^2(\Gamma))$. The optimal control problem (5.1)–(5.2) has a unique solution $(y, u) \in L^2(I; L^2(\Gamma)) \times L^2(I; \mathbb{R}^{n_{\mathcal{D}}})$. Moreover, there exists an adjoint state $p \in L^2(I; X) \cap H^1(I; L^2(\Gamma))$ being the weak solution of the adjoint equation*

$$(5.4) \quad \left\{ \begin{array}{ll} -\partial_t p - p'' + c_0 p = y - \bar{y} & \text{in } I \times \Gamma, \\ p = 0 & \text{on } I \times \mathcal{V}_{\mathcal{D}}, \\ \mathcal{K}p = 0 & \text{on } I \times \mathcal{V}_{\mathcal{K}}, \\ p(T, \cdot) = 0 & \text{in } \Gamma, \end{array} \right.$$

such that

$$\beta u - \mathcal{K}p = 0 \quad \text{a. e. on } I \times \mathcal{V}_{\mathcal{D}}$$

is satisfied. This optimality condition is also sufficient.

Proof. First, we confirm the relation $S^* = -\mathcal{K} \circ P$ with the solution operator of (5.4), $P : L^2(I; L^2(\Gamma)) \rightarrow L^2(I; X) \cap H^1(I; L^2(\Gamma))$. For arbitrary functions $u \in L^2(I; \mathbb{R}^{n_{\mathcal{D}}})$ and $z \in L^2(I; L^2(\Gamma))$, we obtain with $y_u = Su$ and $p = Pz$ the equality

$$\begin{aligned} (Su, z)_{L^2(I \times \Gamma)} &= (y_u, -\partial_t p - p'' + c_0 p)_{L^2(I \times \Gamma)} \\ &= \int_0^T \langle \partial_t y_u(t), p(t) \rangle_{X^* \times X} dt - \langle y_u(T), p(T) \rangle_{H_D^1(\Gamma)^* \times H_D^1(\Gamma)} \\ &\qquad \qquad \qquad + (y_u, -p'' + c_0 p)_{L^2(I \times \Gamma)} \\ &= -(u, \mathcal{K}p)_{L^2(I; \mathbb{R}^{n_{\mathcal{D}}})}, \end{aligned}$$

where (5.3) and $p(T, \cdot) = 0$ are used in the last step. Consequently, the reduced objective $j(u) = J(S(u), u)$ is differentiable, and the standard necessary optimality condition reads $(j'(u), v)_{L^2(I; \mathbb{R}^{n_D})} = (S^*(Su + y_f - \bar{y}), v)_{L^2(I; L^2(\Gamma))} + \beta(u, v)_{L^2(I; \mathbb{R}^{n_D})} = 0$ for all functions $v \in L^2(I; \mathbb{R}^{n_D})$. Substituting S and S^* implies the result. \square

We cannot expect much regularity for the state related to a given control in $L^2(I; \mathbb{R}^{n_D})$. However, for the optimal solution, higher regularity can be deduced.

LEMMA 5.3. *The optimal control, state, and adjoint state of (5.1)–(5.2) possess the regularity*

$$\begin{aligned} u &\in H^{1/4}(I; \mathbb{R}^{n_D}), \\ y &\in L^2(I; H^1(\Gamma)) \cap H^{1/2}(I; L^2(\Gamma)), \\ p &\in L^2(I; X) \cap H^1(I; L^2(\Gamma)). \end{aligned}$$

Proof. The desired result follows from bootstrapping arguments like in [22, Theorem 3.4] and [35, Theorem 3.4]. From $u \in L^2(I; \mathbb{R}^{n_D})$ and $y \in L^2(I \times \Gamma)$, we conclude with [19, §7, Theorem 5(ii)] that $p \in L^2(I; H^2(\Gamma)) \cap H^1(I; L^2(\Gamma))$. A trace theorem then gives $u = (1/\beta)\mathcal{K}p \in H^{1/4}(I; \mathbb{R}^{n_D})$. The regularity of the state is obtained from parabolic regularity results in interpolation spaces [38]. \square

REMARK 5.4. As in the proof of [35, Theorem 3.4] it is even possible to show that the optimal state satisfies $y \in H^1(I; H_D^1(\Gamma)^*)$, which is hence a solution of (5.2) in the usual weak sense. With these arguments, the authors in [35] justify a direct discretization of the standard weak form of the state equation, which we will also use in the following.

Moreover, one can even show higher regularity of the optimal solution, more precisely, $u, \mathcal{K}p \in H^{1/2}(I; \mathbb{R}^{n_D})$ and $y \in L^2(I; H^{3/2-\varepsilon}(\Gamma)) \cap H^{3/4-\varepsilon}(I; L^2(\Gamma))$ with arbitrary $\varepsilon > 0$, provided that the desired state satisfies $\bar{y} \in L^2(I; H^1(\Gamma)) \cap H^{1/2}(I; L^2(\Gamma))$.

5.2. Discretization. In this section we investigate a discretization approach for the optimal control problem based on a temporal discontinuous Galerkin scheme and standard finite elements for the spatial variables. We refer to [41], where such an approach was studied first in the context of optimal control problems.

For the time discretization we introduce an equidistant time grid $0 = t_0 < t_1 < \dots < t_{n_t} = T$ with discretization parameter $\tau := t_{k+1} - t_k$, $k = 0, \dots, n_t - 1$, and define the intervals $I_0 := \{0\}$ and $I_k := (t_{k-1}, t_k]$, for $k = 1, \dots, n_t$. The trial and test spaces for a semi-discrete scheme are defined by

$$V_\tau(X) = \{v \in L^\infty(I; X) : v|_{I_k} \in \mathcal{P}_0(X) \text{ for all } k = 1, \dots, n_t\},$$

where X is an arbitrary Banach space. Each function $w_\tau \in V_\tau(X)$ can be represented by the piecewise constant (in time) functions $w_k := w_\tau|_{I_k} \in X$. The coupling of the time intervals is realized by introducing additional jump terms in the bilinear form. Hence, we define $\llbracket w_\tau \rrbracket_k = w_{k+1} - w_k$, for $k = 1, \dots, n_t - 1$, and $\llbracket w_\tau \rrbracket_0 = w_1$. The DG(0) semi-discretization of the state equation (5.2) then reads:

Find $y_\tau \in V_\tau(H^1(\Gamma))$ with $y_\tau(\cdot, v) = Q_\tau(u_v)$, $v \in \mathcal{V}_D$:

$$\begin{aligned} B(y_\tau, w_\tau) &:= \sum_{k=1}^{n_t} \int_{I_k} a(y_k, w_k) \, dt + \sum_{k=0}^{n_t-1} (\llbracket y_\tau \rrbracket_k, w_{k+1})_{L^2(\Gamma)} \\ &= (f, w_\tau)_{L^2(I \times \Gamma)} \quad \forall w_\tau \in V_\tau(H_D^1(\Gamma)). \end{aligned}$$

Here, $Q_\tau : L^2(I; X) \rightarrow V_\tau(X)$ denotes the temporal L^2 -projection. The term for $k = 0$ in the second sum, namely $(\llbracket y_\tau \rrbracket_0, w_1)_{L^2(\Gamma)} = (y_1, w_1)_{L^2(\Gamma)}$, is used to incorporate the initial

conditions. If the integrals over t are approximated by the rectangle rule, then this approach is equivalent to the well-known backward Euler method.

Note that the bilinear form can be equivalently expressed as

$$B(y_\tau, w_\tau) = \sum_{k=1}^{n_t} \int_{I_k} a(y_k, w_k) dt - \sum_{k=1}^{n_t} (y_k, \llbracket w_\tau \rrbracket_k)_{L^2(\Gamma)},$$

with $\llbracket w_\tau \rrbracket_{n_t} := -y_{n_t}$. When changing the role of trial and test functions, the latter term realizes a discretization of a parabolic problem for $w \approx w_\tau$ going backward in time. The term for $k = n_t$ realizes a homogeneous terminal condition at $t = T$. This fits with the structure of the adjoint equation (5.4).

Next, we insert our spatial discretization from Section 3 and arrive at the fully discrete scheme:

Find $y_{\tau h} \in V_\tau(V_h)$ with $y_{\tau h}(\cdot, v) = Q_\tau(u_v)$, $v \in \mathcal{V}_D$:

$$(5.5) \quad B(y_{\tau h}, w_{\tau h}) = (f, w_{\tau h})_{L^2(I \times \Gamma)} \quad \forall w_{\tau h} \in V_\tau(V_{h,D}),$$

where $w_{kh} := w_{\tau h}|_{I_k} \in V_h$, for $k = 1, \dots, n_t$.

In order to discretize the optimal control problem, we use a control approximation $u_{\tau h} \in V_\tau(\mathbb{R}^{n_D})$, which is piecewise constant in time as well. The discretized version of the control problem (5.1) then reads

$$(5.6) \quad \text{Minimize } \frac{1}{2} \|y_{\tau h} - \bar{y}\|_{L^2(I; L^2(\Gamma))}^2 + \frac{\beta}{2} \|u_{\tau h}\|_{L^2(I; \mathbb{R}^{n_D})}^2$$

over $y_{\tau h} \in V_\tau(V_h)$ and $u_{\tau h} \in V_\tau(\mathbb{R}^{n_D})$ subject to (5.5). An optimality system of (5.6) can be deduced with the Lagrange formalism. To be more precise, one can show that the pair $(y_{\tau h}, u_{\tau h}) \in V_\tau(V_h) \times V_\tau(\mathbb{R}^{n_D})$ is the optimal solution if and only if some discrete adjoint state $p_{\tau h} \in V_\tau(V_{h,D})$ exists satisfying

$$(5.7) \quad \begin{cases} y_{\tau h}(t, v) = u_{\tau h, v}(t) & \forall t \in I, v \in \mathcal{V}_D, \\ B(y_{\tau h}, w_{\tau h}) = (f, w_{\tau h})_{L^2(I; L^2(\Gamma))} & \forall w_{\tau h} \in V_\tau(V_{h,D}), \\ B(w_{\tau h}, p_{\tau h}) = (y_{\tau h} - \bar{y}, w_{\tau h})_{L^2(I; L^2(\Gamma))} & \forall w_{\tau h} \in V_\tau(V_{h,D}), \\ (\beta u_{\tau h}, z_\tau)_{L^2(I; \mathbb{R}^{n_D})} = (\mathcal{K}_{\tau h}(p_{\tau h}), z_\tau)_{L^2(I; \mathbb{R}^{n_D})} & \forall z_\tau \in V_\tau(\mathbb{R}^{n_D}). \end{cases}$$

The discrete Kirchhoff-Neumann operator $\mathcal{K}_{\tau h}: V_\tau(V_h) \rightarrow V_\tau(\mathbb{R}^{n_D})$ is defined by

$$(\mathcal{K}_{\tau h}(p_{\tau h}), z_\tau)_{L^2(I; \mathbb{R}^{n_D})} = B(\mathcal{E}_h z_\tau, p_{\tau h}) - (y_{\tau h} - \bar{y}, \mathcal{E}_h z_\tau)_{L^2(I; L^2(\Gamma))} \quad \forall z_\tau \in V_\tau(\mathbb{R}^{n_D}),$$

with \mathcal{E}_h an arbitrary extension operator onto V_h . A straightforward calculation again gives $S_{\tau h}^* = -\mathcal{K}_{\tau h} \circ P_{\tau h}$ with $P_{\tau h}$ the solution operator of the discrete adjoint equation, which implies the last equation in (5.7).

An error estimate for the control approximation can be deduced by the arguments from [22, Theorem 4]:

THEOREM 5.5. *Let $f, \bar{y} \in L^2(I; L^2(\Gamma))$. Assume that the temporal and spatial discretization parameters are related to each other by means of $\tau = O(h^2)$. Then the solutions of (5.1) and (5.6) satisfy the error estimate*

$$\|u - u_{\tau h}\|_{L^2(I; \mathbb{R}^{n_D})} + \|y - y_{\tau h}\|_{L^2(I; L^2(\Gamma))} \leq c h^{1/2}.$$

REMARK 5.6.

- a) The result of Theorem 5.5 is sharp under the regularity assumptions on \bar{y} and f . As explained in Remark 5.4, the solutions may be more regular for, e.g., $\bar{y} \in H^{1/2}(I; L^2(\Gamma))$, and one would then expect a higher convergence rate than predicted in Theorem 5.5. This is indeed confirmed by numerical experiments, but a proof of this observation is still an open problem and will be addressed in future research.
- b) The convergence rate is limited by the low regularity of the solution. Increasing the polynomial degree for the approximations will not lead to an improvement of the convergence rate.
- c) If a stronger norm for the regularization were used, e.g., $H^s(I; \mathbb{R}^{n_D})$ with $s > 0$, then the optimal solution would be more regular leading to better approximation properties; see [46, 68].

5.3. Preconditioning of the parabolic problem. Similar to the investigations in Section 4 we use a matrix-vector representation of the optimality system (5.7), namely

$$\begin{bmatrix} \mathcal{M}_{FF} & \mathcal{M}_{FD} & \mathcal{K}_{FF}^\top \\ \mathcal{M}_{DF} & \mathcal{M}_{DD} + \beta\mathcal{I} & \mathcal{K}_{FD}^\top \\ \mathcal{K}_{FF} & \mathcal{K}_{FD} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{y}_F \\ \mathbf{u} \\ \mathbf{p}_F \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{y}}_F \\ \bar{\mathbf{y}}_D \\ \mathbf{f}_F \end{bmatrix},$$

where we abuse the notation from before to denote by \mathbf{y}_F , \mathbf{u} , and \mathbf{p}_F the vectors containing state, control, and adjoint state for all time-steps and similarly for the right-hand side. In the following, n_t denotes the number of time steps. We have

$$\begin{aligned} \mathcal{M}_{FF} &= \mathbf{I}_{n_t} \otimes \tau \mathbf{M}_{FF}, & \mathcal{M}_{FD} &= \mathbf{I}_{n_t} \otimes \tau \mathbf{M}_{FD}, \\ \mathcal{M}_{DF} &= \mathbf{I}_{n_t} \otimes \tau \mathbf{M}_{DF}, & \mathcal{M}_{DD} &= \mathbf{I}_{n_t} \otimes \tau \mathbf{M}_{DD}, \\ \mathcal{I} &= \mathbf{I}_{n_t} \otimes \tau \mathbf{I}_{n_D}, \\ \mathcal{K}_{FF} &= \mathbf{C} \otimes \mathbf{M}_{FF} + \mathbf{I}_{n_t} \otimes \tau \mathbf{K}_{FF}, & \mathcal{K}_{FD} &= \mathbf{C} \otimes \mathbf{M}_{FD} + \mathbf{I}_{n_t} \otimes \tau \mathbf{K}_{FD}. \end{aligned}$$

Here the matrix \mathbf{C} represents the time-discretization, e.g.,

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ -1 & 1 & \ddots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \dots & -1 & 1 \end{bmatrix} \in \mathbb{R}^{n_t \times n_t}.$$

The preconditioning strategy that we propose for the parabolic case follows the pattern that we presented earlier. Namely, we use

$$\begin{bmatrix} \mathcal{M}_{FF} & \mathcal{M}_{FD} \\ \mathcal{M}_{DF} & \mathcal{M}_{DD} + \beta\mathcal{I} \end{bmatrix} \approx \begin{bmatrix} \mathcal{M}_{FF} & \mathbf{0} \\ \mathbf{0} & \mathcal{M}_{DD} + \beta\mathcal{I} - \mathcal{M}_{DF} \mathcal{M}_{FF}^{-1} \mathcal{M}_{FD} \end{bmatrix},$$

and then again

$$\mathcal{S}_M = \mathcal{M}_{DD} + \beta\mathcal{I} - \mathcal{M}_{DF} \mathcal{M}_{FF}^{-1} \mathcal{M}_{FD},$$

and, utilizing the Kronecker product representation, this becomes

$$\mathcal{S}_M = \mathbf{I}_{n_t} \otimes \tau (\mathbf{M}_{DD} + \beta \mathbf{I} - \mathbf{M}_{DF} \mathbf{M}_{FF}^{-1} \mathbf{M}_{FD}).$$

We can now use the approximation $\mathbf{blkdiag}(\mathcal{M}_{FF}, \mathcal{S}_M)$ for the construction of the Schur-complement preconditioner of the parabolic problem.

We could again employ a matching argument for the time-dependent Schur complement leading to

$$\mathcal{K}_{FF} \mathcal{M}_{FF}^{-1} \mathcal{K}_{FF}^\top + \mathcal{K}_{FD} \mathcal{D}_{S_M}^{-1} \mathcal{K}_{FD}^\top \approx (\mathcal{K}_{FF} + \mathcal{N}) \mathcal{M}_{FF}^{-1} (\mathcal{K}_{FF}^\top + \mathcal{N}),$$

but again, as the eigenvalues show a rather benign dependence on β , the overall preconditioner is taken as

$$\mathcal{P} = \begin{bmatrix} \mathcal{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{S}_M & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{K}_{FF} \mathcal{M}_{FF}^{-1} \mathcal{K}_{FF}^\top \end{bmatrix},$$

where we use a direct solver for the spatial components within the preconditioner. In future research we will investigate multigrid schemes for higher efficiency. We illustrate the preconditioner's effectiveness in the next section.

6. Numerical experiments. In this section we illustrate how the methodology developed in this paper performs when applied to several challenging datasets. Our implementation is based on MATLAB. For the iterative solver we rely on the standard implementation of MINRES within MATLAB. We run the algorithm until a relative tolerance of 10^{-8} is reached.

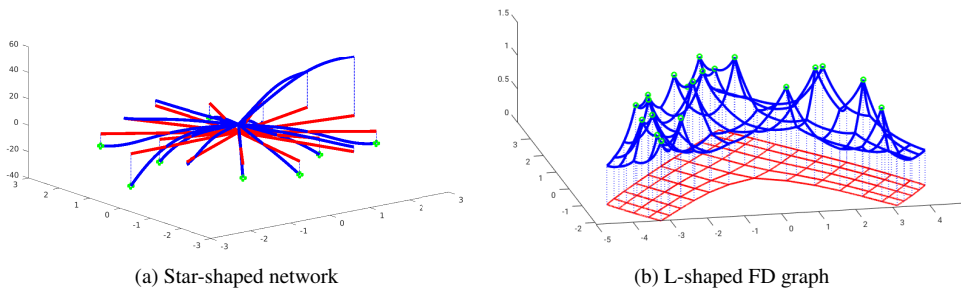


FIG. 6.1. Optimal solutions of Dirichlet control problems. The red lines represent the edges of the graph, the blue lines represent the optimal state, and the green stars are the control nodes.

6.1. Preconditioning. The purpose of the preconditioners that we designed is to provide fast and robust methods that lead to a method exploiting as much as possible the structure of the saddle point problems presented above. While for certain smaller and non-complex graphs the use of direct solvers is very efficient, the motivation for constructing preconditioners is the applicability to complex networks and also more challenging differential equations that contain time-derivatives and possibly parameter dependencies.

The first example we consider is a simple star graph, where we have one internal node and all remaining nodes are leaf nodes. The solution of a Dirichlet boundary control problem on such a graph is illustrated in Figure 6.1a. The required number of iterations of the iterative MINRES-method using the symmetric preconditioner constructed in Section 4.3 are reported in Table 6.1a. Obviously, our iterative solver is robust, both with respect to the regularization parameter β and the discretization parameter h . It also can be observed that our method outperforms the unpreconditioned iterative solver.

As a second example, we consider a finite difference graph of an L-shaped domain, also illustrated in Figure 6.1b. To generate this graph we used the Matlab function “numgrid(‘L’,N)”, where N is the number of vertices at the long edges of an L-shaped domain. In the experiment we used 12 randomly selected controllable nodes. The number of GMRES iterations are presented in Table 6.1b. As in the previous example, we observe robustness of the number of iterations with respect to changes in β and the number of finite element nodes. We see a slight mesh-dependence for an increasing number of degrees of freedom, but the number of iterations stay rather small.

TABLE 6.1

Number of iterations of MINRES required to reach a relative tolerance of 10^{-8} for different regularization parameters β and discretization parameters $N_{\text{DOF}} = n + m(n_e - 1)$. The values in parentheses are the number of iterations for the unpreconditioned method, provided that the method converged after 2000 iterations.

(a) Star graph – Iterations					
N_{DOF}	114	494	2022	8150	32694
β					
10^{-1}	8 (18)	8 (40)	8 (124)	8 (418)	8 (1437)
10^{-2}	8 (18)	8 (41)	8 (126)	8 (394)	8 (1306)
10^{-3}	10 (11)	11 (35)	8 (129)	8 (413)	8 (1375)
10^{-4}	8 (18)	8 (41)	8 (130)	8 (417)	8 (1416)
10^{-5}	8 (18)	8 (40)	8 (130)	8 (416)	8 (1400)

(b) FDM graph – Iterations				
N_{DOF}	44	564	5300	45684
β				
10^{-1}	17 (26)	23 (607)	18 (–)	19 (–)
10^{-2}	19 (27)	25 (557)	20 (–)	14 (–)
10^{-3}	19 (27)	25 (539)	19 (–)	21 (–)
10^{-4}	19 (27)	25 (605)	21 (–)	18 (–)
10^{-5}	19 (27)	26 (617)	21 (–)	14 (–)

6.2. Discretization error estimates. In this experiment we want to test whether the convergence behavior predicted by Theorem 3.6 is also observed in experiments. To this end, we use again the FDM graph of an L-shaped domain as already considered in the previous section. The input data for the optimal control problem are $\beta = 0.01$, $\bar{y} \equiv 1$, $f \equiv 0.5$, $c_0 \equiv 1$, and $u_a = -\infty$, $u_b = 1.4$. The exact solution is illustrated in Figure 6.1b. The constrained problem is solved by a primal-dual active set strategy [31] realized as follows. For a given iterate $\mathbf{u}^{(n)}$ with corresponding state $\mathbf{y}_F^{(n)}$ and adjoint state $\mathbf{p}_F^{(n)}$, we compute the residual

$$\mathbf{r}^{(n)} := \bar{\mathbf{y}}_D - \mathbf{K}_{DF}\mathbf{p}_F^{(n)} - \mathbf{M}_{DF}\mathbf{y}_F^{(n)} - \mathbf{M}_{DD}\mathbf{u}^{(n)} - \beta\mathbf{u}^{(n)}$$

and determine the set of active and inactive indices by means of

$$\mathcal{A}^{(n)} = \{i: [\mathbf{r}^{(n)} + \beta(\mathbf{u}^{(n)} - \mathbf{u}_b)]_i > 0\}, \quad \mathcal{I}^{(n)} = \{1, \dots, n_D\} \setminus \mathcal{A}^{(n)},$$

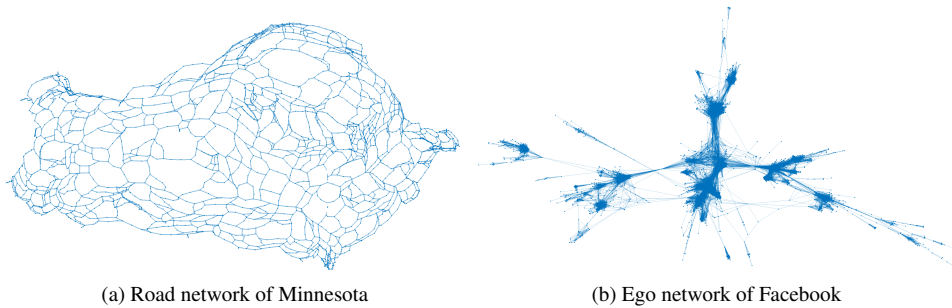


FIG. 6.2. Example networks investigated in Section 6.3.

studied is robust with respect to β and h ; see Table 6.3. The second complex network we use is the Facebook Ego Network Dataset⁴ displayed in Figure 6.2b. We here have 5,228,870 degrees of freedom, and the solver is able to obtain the solution after 38 iterations for $\beta = 10^{-3}$ and 37 iterations for $\beta = 10^{-4}$.

In Table 6.4 we illustrate the performance of our method when applied to Erdős-Renyi graphs [18]. We generate a random graph with $n = 150$ nodes and a connection probability p that is varied. We additionally increase the number of degrees of freedom on an edge n_e . It can be seen that the number of iterations remains fairly constant for changing connectivities and different degrees of freedom.

TABLE 6.4

Number of iterations of the GMRES-method for a problem on an Erdős-Renyi graph with connection probability p and $n = 150$ nodes.

$N_{\text{DOF}}(n_e)$	p	.1	.2	.3
292,992 ($n_e = 2$)	33	30	33	
3,431,904 ($n_e = 4$)	36	33	34	
12,806,640 ($n_e = 6$)	33	33	33	
31,807,104 ($n_e = 8$)	30	27	29	

6.4. Parabolic case. We here briefly illustrate the performance of the suggested preconditioner for the parabolic optimal control problem. Table 6.5 provides the number of iterations of the restarted GMRES method. We set the restart parameter to 10. It is also possible to choose MINRES, which will then lead to higher number of iterations but also a more accurate solution when compared to the solution computed using a direct method. This is likely due to the different ways in which convergence is measured in both methods.

We can see from Table 6.5 that the number of iterations is robust with respect to changes in the discretization parameters for space and time as well as changes in the regularization parameter β . We test the performance on three different networks and observe the robustness of the iterative method when applied to each network.

7. Conclusion & Outlook. In this paper we discussed a PDE-constrained optimization problem on a complex network. The steady PDE-operator and the objective function were

⁴ <https://blogs.mathworks.com/loren/2016/02/03/visualizing-facebook-networks-with-matlab/>

TABLE 6.5

Number of iterations of the GMRES-method for the solution of the parabolic optimal Dirichlet control problem for the finite difference (FDM), facebook (fb), and Minnesota (mn) networks.

	N_{DOF}	β 10^{-1}	10^{-2}	10^{-3}
FDM	18,396	33(0.29)	36 (0.21)	36 (0.19)
	173,040	33(1.20)	37 (1.28)	37 (1.29)
	1,487,808	24(9.99)	30 (15.53)	30 (17.31)
fb	7,348,530	33(52.21)	33 (121.90)	33 (97.69)
	30,904,650	33(307.82)	33 (246.46)	34 (256.96)
mn	1,584,720	36(13.68)	40 (14.35)	40 (13.19)
	3,873,920	33(22.39)	40 (31.34)	40 (26.26)

discretized using finite elements. A rigorous error analysis showed quadratic convergence for the controls, which is also confirmed by the numerical experiments. We further discussed the matrix structure following [3], and we proposed a preconditioning strategy for the saddle point system representing the first-order optimality conditions. The robustness of the preconditioned methods with respect to changes in the regularization and discretization parameters is also observed in numerical experiments. We were able to extend the preconditioning strategy to the case of a parabolic constraint leading to an efficient solution procedure also for complex networks. This is an important first step towards tackling realistic applications, as we believe that much of our methodology can be carried over to the case when nonlinear and time-dependent state equations are considered.

REFERENCES

- [1] T. APEL, M. MATEOS, J. PFEFFERER, AND A. RÖSCH, *Error estimates for Dirichlet control problems in polygonal domains: quasi-uniform meshes*, Math. Control Relat. Fields, 8 (2018), pp. 217–245.
- [2] T. APEL, S. NICAISE, AND J. PFEFFERER, *Discretization of the Poisson equation with non-smooth data and emphasis on non-convex domains*, Numer. Methods Partial Differential Equations, 32 (2016), pp. 1433–1454.
- [3] M. ARIOLI AND M. BENZI, *A finite element method for quantum graphs*, IMA J. Numer. Anal., 38 (2018), pp. 1119–1163.
- [4] A.-L. BARABÁSI, *Network Science*, Cambridge University Press, Cambridge, 2016.
- [5] M. BENZI AND P. BOITO, *Matrix functions in network analysis*, GAMM-Mitt., 43 (2020), Art. No. e202000012, 36 pages.
- [6] M. BENZI, G. H. GOLUB, AND J. LIESEN, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137.
- [7] G. BERKOLAIKO AND P. KUCHMENT, *Introduction to Quantum Graphs*, American Mathematical Society, Providence, 2013.
- [8] E. CASAS AND J.-P. RAYMOND, *Error estimates for the numerical approximation of Dirichlet boundary control for semilinear elliptic equations*, SIAM J. Control Optim., 45 (2006), pp. 1586–1611.
- [9] X. CHENG AND J. M. A. SCHERPEN, *Clustering approach to model order reduction of power networks with distributed controllers*, Adv. Comput. Math., 44 (2018), pp. 1917–1939.
- [10] F. R. K. CHUNG, *Spectral Graph Theory*, American Mathematical Society, Providence, 1997.
- [11] R. DÁGER AND E. ZUAZUA, *Wave Propagation, Observation and Control in 1-d Flexible Multi-Structures*, Springer, Berlin, 2006.
- [12] P. DOMSCHKE, A. DUA, J. J. STOLWIJK, J. LANG, AND V. MEHRMANN, *Adaptive refinement strategies for the simulation of gas flow in networks using a model hierarchy*, Electron. Trans. Numer. Anal., 48 (2018), pp. 97–113.
<http://etna.ricam.oeaw.ac.at/vol.48.2018/pp97-113.dir/pp97-113.pdf>
- [13] P. DOMSCHKE, O. KOLB, AND J. LANG, *Adjoint-based error control for the simulation and optimization of gas and water supply networks*, Appl. Math. Comput., 259 (2015), pp. 1003–1018.

- [14] I. S. DUFF, A. M. ERISMAN, AND J. K. REID, *Direct Methods for Sparse Matrices*, Oxford University Press, New York, 1989.
- [15] H. EGGER AND N. PHILIPPI, *A hybrid discontinuous galerkin method for transport equations on networks*, in International Conference on Finite Volumes for Complex Applications, R. Klöforn, E. Keilegavlen, F. A. Radu, J. Fuhrmann, eds., Springer, Cham, 2020, pp. 487–495.
- [16] H. EGGER AND L. SCHÖBEL-KRÖHN, *Chemotaxis on networks: analysis and numerical approximation*, ESAIM Math. Model. Numer. Anal., 54 (2020), pp. 1339–1372.
- [17] H. C. ELMAN, D. J. SILVESTER, AND A. J. WATHEN, *Finite Elements and Fast Iterative Solvers: With Applications in Incompressible Fluid Dynamics*, Oxford University Press, New York, 2005.
- [18] P. ERDŐS AND A. RÉNYI, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl., 5 (1960), pp. 17–61.
- [19] L. C. EVANS, *Partial Differential Equations*, 2nd ed., American Mathematical Society, Providence, 2010.
- [20] R. FERES AND M. WALLACE, *Reaction-diffusion on metric graphs and conversion probability*, Preprint on arXiv, 2015. <https://arxiv.org/abs/1501.06976>.
- [21] M. GARAVELLO AND B. PICCOLI, *Traffic Flow on Networks*, American Institute of Mathematical Sciences, Springfield, 2006.
- [22] W. GONG, M. HINZE, AND Z. ZHOU, *Finite element method and a priori error estimates for Dirichlet boundary control problems governed by parabolic PDEs*, J. Sci. Comput., 66 (2016), pp. 941–967.
- [23] L. J. GRADY AND J. R. POLIMENI, *Discrete Calculus*, Springer, London, 2010.
- [24] S. GRUNDEL AND M. HERTY, *Hyperbolic discretization via Riemann invariants*, Preprint on arXiv, 2020. <https://arxiv.org/abs/2005.12158>.
- [25] S. GRUNDEL, N. HORNUNG, AND S. ROGGENDORF, *Numerical aspects of model order reduction for gas transportation networks*, in Simulation-Driven Modeling and Optimization, S. Koziel, L. Leifsson, X.-S. Yang, eds., Springer Proceedings in Mathematics & Statistics, Springer International Publishing, Basel, 2016, pp. 1–28.
- [26] M. GUGAT, M. HERTY, AND V. SCHLEPER, *Flow control in gas networks: exact controllability to a given demand*, Math. Methods Appl. Sci., 34 (2010), pp. 745–757.
- [27] M. HERTY, J. MOHRING, AND V. SACHERS, *A new model for gas flow in pipe networks*, Math. Methods Appl. Sci., 33 (2010), pp. 845–855.
- [28] M. HERTY AND C. RINGHOFER, *Averaged kinetic models for flows on unstructured networks*, Kinet. Relat. Models, 4 (2011), pp. 1081–1096.
- [29] R. HERZOG AND E. SACHS, *Preconditioned conjugate gradient method for optimal control problems with control and state constraints*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2291–2317.
- [30] J. HILD AND G. LEUGERING, *Real-time control of urban drainage systems*, in Mathematical Optimization of Water Networks, A. Martin, K. Klamroth, J. Lang, G. Leugering, A. Morsi, M. Oberlack, M. Ostrowski, and R. Rosen, eds., vol. 162 of Internat. Ser. Numer. Math., Birkhäuser/Springer, Basel, 2012, pp. 129–150.
- [31] M. HINTERMÜLLER, K. ITO, AND K. KUNISCH, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2002), pp. 865–888.
- [32] M. HINZE, R. PINNAU, M. ULBRICH, AND S. ULBRICH, *Optimization with PDE Constraints*, Springer, New York, 2009.
- [33] K. ITO AND K. KUNISCH, *Lagrange Multiplier Approach to Variational Problems and Applications*, SIAM, Philadelphia, 2008.
- [34] P. KUCHMENT AND O. POST, *On the spectra of carbon nano-structures*, Comm. Math. Phys., 275 (2007), pp. 805–826.
- [35] K. KUNISCH AND B. VEXLER, *Constrained Dirichlet boundary control in L^2 for a class of evolution equations*, SIAM J. Control Optim., 46 (2007), pp. 1726–1753.
- [36] J. E. LAGNESE, G. LEUGERING, AND E. J. P. G. SCHMIDT, *Modeling, Analysis and Control of Dynamic Elastic Multi-Link Structures*, Birkhäuser, Boston, 1994.
- [37] G. LEUGERING, A. MARTIN, M. SCHMIDT, AND M. SIRVENT, *Nonoverlapping domain decomposition for optimal control problems governed by semilinear models for gas flow in networks*, Control Cybernet., 46 (2017), pp. 191–225.
- [38] J.-L. LIONS AND E. MAGENES, *Non-Homogeneous Boundary Value Problems and Applications. Vol. II*, Springer, New York, 1972.
- [39] K.-A. MARDAL AND R. WINTHER, *Preconditioning discretizations of systems of partial differential equations*, Numer. Linear Algebra Appl., 18 (2011), pp. 1–40.
- [40] S. MAY, R. RANNACHER, AND B. VEXLER, *Error analysis for a finite element approximation of elliptic Dirichlet boundary control problems*, SIAM J. Control Optim., 51 (2013), pp. 2585–2611.
- [41] D. MEIDNER AND B. VEXLER, *A priori error estimates for space-time finite element discretization of parabolic optimal control problems. II. Problems with control constraints*, SIAM J. Control Optim., 47 (2008), pp. 1301–1329.
- [42] P. MLINARIĆ, S. GRUNDEL, AND P. BENNER, *Efficient model order reduction for multi-agent systems using qr decomposition-based clustering*, in 2015 54th IEEE Conference on Decision and Control (CDC), IEEE

- Conference Proceedings, Los Alamitos, 2015, pp. 4794–4799.
- [43] D. MUGNOLO, *Semigroup Methods for Evolution Equations on Networks*, Springer, Cham, 2014.
- [44] M. F. MURPHY, G. H. GOLUB, AND A. J. WATHEN, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput., 21 (2000), pp. 1969–1972.
- [45] A. NORDENFELT, *Spectral analysis of discrete approximations of quantum graphs*, Tech. Rep., Lund University, Lund, 2007.
- [46] G. OF, T. X. PHAN, AND O. STEINBACH, *An energy space finite element approach for elliptic Dirichlet boundary control problems*, Numer. Math., 129 (2015), pp. 723–748.
- [47] S. F. OPPENHEIMER, *A convection-diffusion problem in a network*, Appl. Math. Comput., 112 (2000), pp. 223–240.
- [48] C. C. PAIGE AND M. A. SAUNDERS, *Solutions of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.
- [49] J. W. PEARSON, M. STOLL, AND A. J. WATHEN, *Regularization-robust preconditioners for time-dependent PDE-constrained optimization problems*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 1126–1152.
- [50] J. W. PEARSON AND A. J. WATHEN, *A new approximation of the Schur complement in preconditioners for PDE-constrained optimization*, Numer. Linear Algebra Appl., 19 (2012), pp. 816–829.
- [51] ———, *Fast iterative solvers for convection-diffusion control problems*, Electron. Trans. Numer. Anal., 40 (2013), pp. 294–310.
<http://etna.ricam.oeaw.ac.at/vol.40.2013/pp294-310.dir/pp294-310.pdf>
- [52] J. PFEFFERER AND M. WINKLER, *Finite element error estimates for normal derivatives on boundary concentrated meshes*, SIAM J. Numer. Anal., 57 (2019), pp. 2043–2073.
- [53] Y. QIU, S. GRUNDEL, M. STOLL, AND P. BENNER, *Efficient numerical methods for gas network modeling and simulation*, Netw. Heterog. Media, 15 (2020), pp. 653–679.
- [54] T. REES, H. S. DOLLAR, AND A. J. WATHEN, *Optimal solvers for PDE-constrained optimization*, SIAM J. Sci. Comput., 32 (2010), pp. 271–298.
- [55] T. REES AND M. STOLL, *Block-triangular preconditioners for PDE-constrained optimization*, Numer. Linear Algebra Appl., 17 (2010), pp. 977–996.
- [56] Y. ROMANO, M. ELAD, AND P. MILANFAR, *The little engine that could: regularization by denoising (RED)*, SIAM J. Imaging Sci., 10 (2017), pp. 1804–1844.
- [57] M. ROZŁOŹNÍK, *Saddle-Point Problems and Their Iterative Solution*, Birkhäuser/Springer, Cham, 2018.
- [58] Y. SAAD AND M. H. SCHULTZ, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
- [59] J. SCHÖBERL AND W. ZULEHNER, *Symmetric indefinite preconditioners for saddle point problems with applications to PDE-constrained optimization problems*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 752–773.
- [60] D. I. SHUMAN, S. K. NARANG, P. FROSSARD, A. ORTEGA, AND P. VANDERGHEYNST, *The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains*, IEEE Signal Proc. Mag., 30 (2013), pp. 83–98.
- [61] D. SPIELMAN, *Spectral graph theory*, in Combinatorial Scientific Computing, U. Naumann and O. Schenk, eds., Chapman & Hall/CRC Computational Science Series, CRC Press, Boca Raton, 2012, pp. 495–524.
- [62] G. STADLER, *Elliptic optimal control problems with L^1 -control cost and applications for the placement of control devices*, Comput. Optim. Appl., 44 (2009), pp. 159–181.
- [63] M. STOLL, *One-shot solution of a time-dependent time-periodic PDE-constrained optimization problem*, IMA J. Numer. Anal., 34 (2014), pp. 1554–1577.
- [64] ———, *A literature survey of matrix methods for data science*, GAMM-Mitt., 43 (2020), Art.No. e202000013, 26 pages.
- [65] M. STOLL AND A. WATHEN, *Preconditioning for partial differential equation constrained optimization with control constraints*, Numer. Linear Algebra Appl., 19 (2012), pp. 53–71.
- [66] F. TRÖLTZSCH, *Optimal Control of Partial Differential Equations: Theory, Methods, and Applications*, American Mathematical Society, Providence, 2010.
- [67] U. VON LUXBURG, *A tutorial on spectral clustering*, Stat. Comput., 17 (2007), pp. 395–416.
- [68] M. WINKLER, *Error estimates for variational normal derivatives and Dirichlet control problems with energy regularization*, Numer. Math., 144 (2020), pp. 413–445.
- [69] A. ZLOTNIK, M. CHERTKOV, AND S. BACKHAUS, *Optimal control of transient flow in natural gas networks*, in Decision and Control (CDC), 2015 IEEE 54th Annual Conference on, IEEE Conference Proceedings, Los Alamitos, 2015, pp. 4563–4570.
- [70] W. ZULEHNER, *Nonstandard norms and robust estimates for saddle point problems*, SIAM J. Matrix Analysis Applications, 32 (2011), pp. 536–560.