Technische Universität Graz

Parallel iterative solvers for discretized reduced optimality systems

U. Langer, R. Löscher, O. Steinbach, H. Yang

Berichte aus dem Institut für Angewandte Mathematik

Bericht 2023/9

Technische Universität Graz

Parallel iterative solvers for discretized reduced optimality systems

U. Langer, R. Löscher, O. Steinbach, H. Yang

Berichte aus dem Institut für Angewandte Mathematik

Bericht 2023/9

Technische Universität Graz
Institut für Angewandte Mathematik
Steyrergasse 30
A 8010 Graz

WWW: `http://www.applied.math.tugraz.at`

Parallel iterative solvers for discretized reduced optimality systems

Ulrich Langer, Richard Löscher, Olaf Steinbach, Huidong Yang

Abstract We propose, analyze, and test new iterative solvers for large-scale systems of linear algebraic equations arising from the finite element discretization of reduced optimality systems defining the finite element approximations to the solution of elliptic tracking-type distributed optimal control problems with both the standard L_2 and the more general energy regularizations. If we aim at an approximation of the given desired state y_d by the computed finite element state y_h that asymptotically differs from y_d in the order of the best L_2 approximation under acceptable costs for the control, then the optimal choice of the regularization parameter ϱ is linked to the mesh-size h by the relations $\varrho = h^4$ and $\varrho = h^2$ for the L_2 and the energy regularization, respectively. For this setting, we can construct efficient parallel iterative solvers for the reduced finite element optimality systems. These results can be generalized to variable regularization parameters adapted to the local behavior of the mesh-size that can heavily change in case of adaptive mesh refinement. Similar results can be obtained for the space-time finite element discretization of the corresponding parabolic and hyperbolic optimal control problems.

Ulrich Langer
Institute of Numerical Mathematics, JKU Linz, Austria, e-mail: ulanger@numa.uni-linz.ac.at

Richard Löscher
Institut für Angewandte Mathematik, TU Graz, Austria, e-mail: loescher@math.tugraz.at

Olaf Steinbach
Institut für Angewandte Mathematik, TU Graz, Austria, e-mail: loescher@math.tugraz.at

Huidong Yang
Faculty of Mathematics, University of Vienna, Austria, e-mail: huidong.yang@univie.ac.at

1 Introduction

Let us first consider an abstract tracking-type, distributed Optimal Control Problem (OCP) of the form: Find the state $y_\varrho \in Y$ and the control $u_\varrho \in U$ minimizing the cost functional

$$J_\varrho(y_\varrho, u_\varrho) := \frac{1}{2}\|y_\varrho - y_d\|_H^2 + \frac{\varrho}{2}\|u_\varrho\|_U^2 = \frac{1}{2}\|y_\varrho - y_d\|_H^2 + \frac{1}{2}\|\sqrt{\varrho}\, u_\varrho\|_U^2 \quad (1)$$

subject to (s.t.) the state equation

$$By_\varrho = u_\varrho \ \text{ in } \ U \subseteq P^*, \quad (2)$$

where $y_d \in H$ denotes the given desired state (target), $\varrho > 0$ is a suitably chosen regularization parameter that also affects the energy cost $\|u_\varrho\|_U^2$ for the control u_ϱ as source term in (2), and $X = Y, P, U, H$ are Hilbert spaces equipped with the corresponding norms $\|\cdot\|_X$ and scalar products $(\cdot,\cdot)_X$. We assume that $Y \subset H \subset Y^*$ and $P \subset H \subset P^*$ form Gelfand triples of Hilbert spaces, and that $B : Y \to P^*$ is an isomorphism, where X^* denotes the dual space of X with the duality product $\langle\cdot,\cdot\rangle : X^* \times X \to \mathbb{R}$ that is nothing but the extension of the scalar product $(\cdot,\cdot)_H$ in H for $X = Y$ and $X = P$. We are interested in the cases $U = H$ and $U = P^*$. Optimal control problems of the form (1)-(2) with many applications were already investigated in the classical monograph [6] by Lions and the more recent book [8] by Tröltzsch, where additional constraints of the form $u_\varrho \in U_{\rm ad} \subset U$ imposed on the control u_ϱ are permitted. The unique solvability of such kind of OCPs is based on the unique solvability of the state equation, i.e. $y_\varrho = B^{-1}u_\varrho$, the strong convexity of the quadratic cost functional and the assumption that the admissible set $U_{\rm ad}$ is a non-empty, convex, and closed subset of U; cf. Theorem 2.16 in [8]. Here we only consider the case $U_{\rm ad} = U$. Then the unique solution $(y_\varrho, u_\varrho) \in Y \times U$ of the OCP (1) - (2) can also be extracted from the unique solution $(y_\varrho, p_\varrho, u_\varrho) \in Y \times P \times U$ of the first-order optimality system (OS)

$$By_\varrho = u_\varrho,\ B^*p_\varrho = y_\varrho - y_d,\ p_\varrho + A_{1/\varrho}^{-1}u_\varrho = 0, \quad (3)$$

where the self-adjoint and elliptic regularization operator $A_{1/\varrho} : P \to P^*$ is defined by the regularization via the Riesz representation of the control. For $U = P^*$, we have $A_{1/\varrho} = \varrho^{-1}A$ and $\|u\|_{P^*}^2 = \langle u, A^{-1}u\rangle$, whereas $A = I$ (canonical embedding operator) for $U = H$. Formally, we will write $\|\sqrt{\varrho}\, u\|_U^2 = \langle u, A_{1/\varrho}^{-1}u\rangle$ that allows us to permit variable ϱ. Eliminating $u = -A_{1/\varrho}p$ from (3), we arrive at the equivalent reduced OS (ROS) written as saddle point problem: Find $(y_\varrho, p_\varrho) \in Y \times P$ such that

$$\begin{bmatrix} A_{1/\varrho} & B \\ B^* & -I \end{bmatrix} \begin{bmatrix} p_\varrho \\ y_\varrho \end{bmatrix} = \begin{bmatrix} 0 \\ -y_d \end{bmatrix} \quad \text{in } P^* \times Y^*. \quad (4)$$

Typical examples are elliptic, parabolic, and hyperbolic OCPs, where the state equation (2) is given by an elliptic boundary value problem (BVP), a parabolic

initial boundary value problem (IBVP), and a hyperbolic IBVP, respectively. In this paper, we will focus on the parallel numerical solution of elliptic OCPs, where the Dirichlet boundary value problem for the Poisson equation serves us as model problem for the state equation. We are primarily interested in efficient parallel solvers for algebraic systems arising from the finite element (FE) discretization of the ROS (4) or the corresponding primal Schur complement when the FE discretization and the regularization are balanced in an asymptotically optimal way.

2 Elliptic Optimal Control Problems

As mentioned above, we focus on elliptic OCPs of the form (1)-(2) defined by the following specifications: $Y = P = H_0^1(\Omega)$, $Y^* = P^* = H^{-1}(\Omega)$, $H = L_2(\Omega)$, and $B = -\Delta : Y = H_0^1(\Omega) \to P^* = H^{-1}(\Omega)$ is defined by the variational identity

$$\langle By, p\rangle = (\nabla y, \nabla p)_{L_2(\Omega)}, \quad \forall y \in Y = H_0^1(\Omega),\ \forall p \in P = H_0^1(\Omega), \tag{5}$$

where $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, denotes the d-dimensional computational domain that is supposed to be bounded and Lipschitz. Throughout the paper, we use the usual notations for Lebesgue and Sobolev spaces. In [7], Neumüller and Steinbach showed that $\|y_\varrho - y_d\|_{L_2(\Omega)}$ behaves like $O(\varrho^{s/r})$ provided that $y_d \in H_0^s(\Omega) := [L_2(\Omega), H_0^1(\Omega)]_s$ for some $s \in [0, 1]$, where $r = 4$ ($U = L_2(\Omega)$), and $r = 2$ ($U = H^{-1}(\Omega)$), respectively.

The FE Galerkin discretization of the ROS (4) reads as follows: Find $y_h = y_{\varrho h} \in Y_h = S_h^1(\mathcal{T}_h) \cap Y = \text{span}\{\varphi_{hk}\}_{k=1}^{n_h} \subset Y$ and $p_h = p_{\varrho h} \in P_h = S_h^1(\mathcal{T}_h) \cap P = \text{span}\{\psi_{hi}\}_{i=1}^{m_h} \subset P$ such that

$$\langle A_{1/\varrho} p_h, q_h\rangle + \langle By_h, q_h\rangle = 0 \text{ and } \langle B^* p_h, v_h\rangle - (y_h, v_h)_{L_2(\Omega)} = -(y_d, v_h)_{L_2(\Omega)} \tag{6}$$

for all $q_h \in P_h$ and $v_h \in Y_h$, where $S_h^1(\mathcal{T}_h)$ is nothing but the continuous, piecewise linear FE space defined on some shape-regular triangulation $\mathcal{T}_h$ of Ω. Here $Y_h = P_h$, $n_h = m_h$ and $\varphi_{hk} = \psi_{hk}$ since $Y = P = H_0^1(\Omega)$. Once the basis is chosen, the FE ROS leads to the symmetric and indefinite (SID) algebraic system

$$\begin{bmatrix} A_{1/\varrho,h} & B_h \\ B_h^T & -M_h \end{bmatrix} \begin{bmatrix} \mathbf{p}_h \\ \mathbf{y}_h \end{bmatrix} = \begin{bmatrix} \mathbf{0}_h \\ -\mathbf{y}_{dh} \end{bmatrix} \tag{7}$$

that can further be reduced to the symmetric and positive definite (SPD) Schur complement (SC) system

$$(B_h^T A_{1/\varrho,h}^{-1} B_h + M_h)\mathbf{y}_h = \mathbf{y}_{dh}, \tag{8}$$

where $\mathbf{y}_h = [y_k]_{k=1}^{n_h} \in \mathbb{R}^{n_h}$ and $\mathbf{p}_h = [y_k]_{k=1}^{n_h} \in \mathbb{R}^{n_h}$ are the nodal FE vectors corresponding to the FE functions $y_h \in Y_h = P_h$ and $p_h \in P_h = Y_h$ via the FE isomorphism $\mathbf{y}_h, \mathbf{p}_h \leftrightarrow y_h, p_h$. The SPD SC system (8) can be solved by the PCG

method provided that some good preconditioner for the SC is available, and the application of $A_{\varrho h}^{-1}$ to some vector performs in asymptotically optimal complexity.

Let us first briefly review some results concerning the standard L_2 regularization where $U = H = L_2(\Omega)$ yielding $A_{1/\varrho,h} = \varrho^{-1} M_h$. Then we can prove the estimate

$$\|y_{\varrho h} - y_d\|_{L_2(\Omega)} \leq c h^s \|y_d\|_{H^s(\Omega)} \tag{9}$$

provided that $\varrho = h^4$ [5]. Estimate (9) remains true for $A_{1/\varrho,h} = \varrho^{-1} D_h$, where $D_h = \mathrm{lump}(M_h)$ denotes the lumped mass matrix [3]. In both cases, the lumped mass matrix D_h is spectrally equivalent to the Schur complement $S_h = B_h^T A_{1/\varrho,h}^{-1} B_h + M_h$ provided that $\varrho = h^4$. More precisely, the spectral equivalence inequalities

$$(d+2)^{-1} D_h \leq M_h \leq S_h := B_h^T A_{1/\varrho,h}^{-1} B_h + M_h \leq (c_{\mathrm{inv}}^r + 1) M_h \leq (c_{\mathrm{inv}}^r + 1) D_h \tag{10}$$

hold with $r = 4$, where the positive constant c_{inv} is defined by the inverse inequality $\|\nabla v_h\|_{L_2(\Omega} \leq c_{\mathrm{inv}} h^{-1} \|v_h\|_{L_2(\Omega)}$ for all $v_h \in Y_h$. We refer to [3] for the proof of (10). In particular, for targets y_d not belonging to Y, e.g. discontinuous targets, it may be useful to use adaptivity and variable, mesh-adapted regularizations of the form

$$\varrho(x) = h_\tau^2, \forall x \in \tau, \forall \tau \in \mathcal{T}_h. \tag{11}$$

Now $A_{1/\varrho,h} = M_{1/\varrho}$, where $M_{1/\varrho}$ is defined by $(M_{1/\varrho} \mathbf{p}_h, \mathbf{q}_h) = ((1/\varrho) p_h, q_h)_{L_2(\Omega)}$, see [4] for details.

In this contribution, we will focus on the more non-standard H^{-1} regularization where $U = P^* = H^{-1}(\Omega)$ yielding $A_{1/\varrho,h} = K_{1/\varrho,h}$. The regularization matrix $K_{1/\varrho,h}$ is now a diffusion stiffness matrix defined by $(K_{1/\varrho,h} \mathbf{p}_h, \mathbf{q}_h) = ((1/\varrho)\nabla p_h, \nabla q_h)_{L_2(\Omega)}$, with the mesh-adapted choice (11) of the regularization ϱ. This kind of diffusion regularization was completely analyze in [2]. In particular, the spectral equivalence inequalities (10) are valid with $r = 2$. Thus, in principle, the SPD SC system (8) can be solved by the PCG with the diagonal matrix D_h as preconditioner. However, in contrast to the L_2 regularization, the matrix $A_{1/\varrho,h} = K_{1/\varrho,h}$ cannot be replace by a diagonal matrix in order to obtain a fast matrix-by-vector multiplication without disturbing the accuracy of the discretization. One loophole would be the use of inner iteration to invert $K_{1/\varrho,h}$, e.g. (algebraic) multigrid iteration as we used in some numerical experiments in [2]. In order to avoid inner iterations down to the discretization error, we can solve the larger SID system (7) by some SID solver like MINRES, Bramble-Pasciak PCG, or GMRES. Then we only need a preconditioner for $K_{1/\varrho,h}$ and S_h. In the case of a constant regularization parameter $\varrho = h^2$, we surprisingly observe that $A_{1/\varrho,h} = K_{1/\varrho,h} = (1/\varrho) K_h = (1/\varrho) B_h$, and, therefore, $S_h = B_h^T A_{1/\varrho,h}^{-1} B_h + M_h = \varrho K_h + M_h$, and the SC system (8) simplifies to a diffusion problem of the form $(\varrho K_h + M_h)\mathbf{y}_h = \mathbf{y}_{dh}$. This system can easily be solved by PCG with the diagonal preconditioner D_h.

Inspired by this observation, we now propose to solve the diffusion equation

$$(\varrho \nabla y, \nabla v)_{L_2(\Omega)} + (y, v)_{L_2(\Omega)} = (y_d, v)_{L_2(\Omega)} \; \forall v \in Y = H_0^1(\Omega) \tag{12}$$

when we want to choose variable regularization parameters $\varrho = \varrho(x)$ in connection with adaptive FE discretization. Obviously, the diffusion equation (12) is nothing but the first-order OS for minimizing the cost functional

$$\widetilde{J}(y) = 0.5[\|y - y_d\|^2 + \|\sqrt{\varrho}\nabla y\|^2] = 0.5[\|y - y_d\|^2 + \langle B^{-*}A_\varrho B^{-1}u, u\rangle] = \widetilde{J}(y,u)$$

instead of the original cost functional

$$J(y) = 0.5[\|y - y_d\|^2 + \langle B^* A_{1/\varrho}^{-1} By, y\rangle]] = 0.5[\|y - y_d\|^2 + \langle A_{1/\varrho}^{-1}u, u\rangle] = J(y,u),$$

where $B^{-*} := (B^{-1})^*$, and the subscript $L_2(\Omega)$ is here omitted from the norms. We note that $\langle A_{1/\varrho}^{-1}u, u\rangle \le \langle B^{-*}A_\varrho B^{-1}u, u\rangle$ with "=" instead of "$\le$" for constant ϱ.

The FE discretization of (12) now leads to the SPD system

$$(K_{\varrho h} + M_h)\mathbf{y}_h = \mathbf{y}_{dh}, \tag{13}$$

where the SPD diffusion stiffness matrix $K_{\varrho h}$ is defined by the identity

$$(K_{\varrho h}\mathbf{y}_h, \mathbf{v}_h) = (\varrho\nabla y_h, \nabla v_h)_{L_2(\Omega)},\ \forall\, \mathbf{y}_h, \mathbf{v}_h \leftrightarrow y_h, v_h \in Y_h \tag{14}$$

For the diffusion regularization (11), the discretization error $\|y_h - y_d\|_{L_2(\Omega)}$ can be analyzed in the similar way as was done for the original approach in [2]. Moreover, the SPD system (13) is much more simpler that the original systems (7) or (8), and can easily be solved by PCG. Indeed, the diffusion regularization (11) ensures that the system matrix $K_{\varrho h} + M_h$ of (13) fulfills the same spectral equivalence inequalities (10) as the SC $B_h^T A_{1/\varrho,h}^{-1} B_h + M_h$ with the same $r = 2$. Thus, we can solve (13) by means of the PCG preconditioned by the diagonal matrix $D_h = \mathrm{lump}(M_h)$, or by another diagonal approximation of the mass matrix M_h like $\mathrm{diag}(M_h)$. This leads to an asymptotically optimal solver for a fixed relative accuracy. This solver can be used in a nested iteration setting on a sequence of refined meshes $\mathcal{T}_\ell$, $\ell = 1, 2, \ldots, L$, in such a way that, at each level ℓ, we compute final iterates that differs from the target y_d in the order of the discretization error with respect to the L_2 norm in optimal complexity. The the whole solution process has optimal complexity. We note that, due to (10), the PCG converges in the L_2 norm, and the number of nested iterations is de facto constant over the discretization levels. It is clear that both the non-nested and the nested PCG solver can easily be parallelized. These convergence properties and parallel performance will impressively be confirmed by our numerical experiments presented in the next section.

3 Numerical experiments

We here consider the discontinuous desired state

$$y_d = 1 \text{ in } (0.25, 0.75)^3, \text{ and } y_d = 0 \text{ in } \overline{\Omega} \setminus (0.25, 0.75)^3, \tag{15}$$

where $\Omega = (0,1)^3 \subset \mathbb{R}^{d=3}$. We note that this discontinuous desired state $y_d \in L_2(\Omega)$ does not belong to the state space $H_0^1(\Omega)$, but $y_d \in H^{1/2-\varepsilon}(\Omega)$ for any $\varepsilon > 0$.

We start with a uniform decomposition of the domain $\Omega = (0,1)^3$ into 24,576 tetrahedral elements $\tau \in \mathcal{T}_1$. The coarsest mesh $\mathcal{T}_1$ contains 4,913 vertices, and the mesh size $h = h_1 = 0.0625$. A sequence of meshes $\mathcal{T}_\ell$ at the levels $\ell = 2, ..., 6$ is generated by successive uniform refinement. On the finest level $\ell = L = 6$, there are 135,005,697 vertices, $h = 1.953125\text{e}{-3}$, and $\varrho = h^2 \approx 3.81\text{e}{-6}$. In the adaptive algorithm, we have chosen the locally adapted $\varrho_\tau = h_\tau^2$ on each $\tau \in \mathcal{T}_h$. The adaptivity is driven by the localization of the error $\|\tilde{y}_\ell - y_d\|_{L_2(\Omega)}$. System (13) is solved by PCG with the preconditioner $\text{diag}(M_h)$ The PCG iteration stops when the relative preconditioned residual is reduced by a factor of 10^6. The parallel implementation is based on the opensource MFEM (`https://mfem.org/`), and is running on the HPC cluster RADON1 (`https://www.oeaw.ac.at/ricam/hpc`).

In the nested iteration, we run the PCG until the relative preconditioned residual reaches 10^{-6} at the coarsest level $\ell = 1$. At the refined levels $\ell = 2, 3, ...$, an adaptive tolerance $\alpha[n_\ell / n_{\ell-1}]^{-\beta/3}$, $\ell = 2, 3, ...$, is adopted for controlling the relative preconditioned residual, with n_ℓ being the number of degrees of freedom (#Dofs) at the mesh level ℓ. Here α is a scaling factor, e.g., 0.5 and 0.25 for the uniform and adaptive refinement, respectively. The parameter β is directly related to the convergence rate of the discretization error $\|y_\ell - y_d\|_{L_2(\Omega)}$. For our example, β is 0.5 and 0.75 for the uniform and the adaptive refinement, respectively.

The #Dofs, the error $\|y_d - y_\ell^{k_\ell}\|_{L_2(\Omega)}$, the number Its of PCG iterations, and the corresponding computational time (Time) in seconds at each level ℓ are illustrated in Table 1 for non-nested iterations starting with the initial guess $y_\ell^0 = 0$. All computations were performed using 512 cores. It is easy to observe the expected convergence rate 0.5 for the uniform refinement and 0.75 for the adaptive refinement; see also Fig. 1. The robustness of our preconditioner is confirmed by the constant PCG iteration numbers for both uniform and adaptive refinements. The efficiency of the parallelization is here demonstrated by very small computational costs at all levels. The adaptive method outperforms the uniform one in terms of computational time needed for achieving a similar accuracy.

For the nested iteration where the initial guess is interpolated from the coarser mesh, the same behavior can be observed from Table 2 which presents the same data as Table 1. The number of PCG iterations as well as the computational time are reduced by a factor of 8 approximately without loss of accuracy; cf. Fig. 2.

Finally, in Table 3 for the non-nested iterations and Table 4 for nested iterations, we compare computational time in seconds with respect to both the number of cores and number of refinement levels on uniform refinement. We clearly see the strong (rowwise) and weak (diagonal) scalability of our parallel solvers.

ℓ	Adaptive			Uniform			
	#Dofs	error	Its (Time)	#Dofs	error	eoc	Its (Time)
1	4, 913	1.61e−1	20 (6.0e−3 s)	4, 913	1.61e−1	–	20 (6.0e−3 s)
2	5, 532	1.54e−1	24 (8.3e−3 s)	35, 937	1.17e−1	0.46	23 (7.4e−3 s)
3	8, 255	1.24e−1	25 (8.5e−3 s)	274, 625	8.26e−2	0.51	23 (9.1e−3 s)
4	18, 013	9.65e−2	26 (9.9e−3 s)	2, 146, 689	5.79e−2	0.51	22 (2.0e−2 s)
5	35, 055	7.80e−2	26 (1.1e−2 s)	16, 974, 593	4.07e−2	0.51	22 (2.0e−1 s)
6	80, 381	6.28e−2	27 (1.3e−2 s)	135, 005, 697	2.87e−2	0.50	22 (1.4e−0 s)
7	167, 982	5.27e−2	27 (1.4e−2 s)				
8	316, 839	4.48e−2	27 (1.7e−2 s)				
9	410, 144	3.96e−2	27 (2.0e−2 s)				
10	1, 264, 336	3.11e−2	28 (2.8e−2 s)				
11	6, 043, 649	2.00e−2	27 (1.3e−1 s)				
12	10, 590, 586	1.83e−2	28 (2.6e−1 s)				
13	25, 217, 222	1.40e−2	27 (5.0e−1 s)				

Table 1 Non-nested iteration: Error, number Its of PCG iterations and computational time (Time) for adaptive and uniform refinements, where eoc denotes the experimental order of convergence.

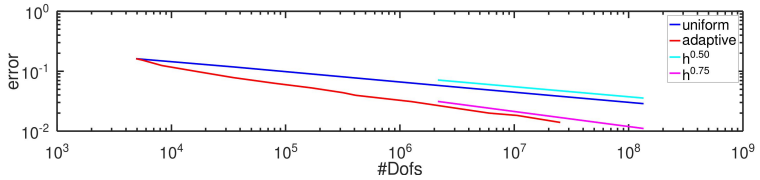

Fig. 1 Non-nested iteration: Convergence history for uniform and adaptive refinements.

ℓ	Adaptive			Uniform			
	#Dofs	error	Its (Time)	#Dofs	error	eoc	Its (Time)
1	4, 913	1.61e−1	20 (6.0e−3 s)	4, 913	1.61e−1	–	20 (6.0e−3 s)
2	5, 532	1.58e−1	2 (9.4e−4 s)	35, 937	1.18e−1	0.46	2 (8.9e−4 s)
3	8, 096	1.27e−1	3 (1.2e−3 s)	274, 625	8.22e−2	0.52	3 (1.4e−3 s)
4	17, 166	1.00e−1	3 (1.2e−3 s)	2, 146, 689	5.77e−2	0.52	3 (3.7e−3 s)
5	34, 134	9.23e−2	2 (1.4e−3 s)	16, 974, 593	4.09e−2	0.50	2 (3.0e−2 s)
6	73, 530	6.59e−2	3 (2.1e−3 s)	135, 005, 697	2.87e−2	0.52	2 (1.8e−1 s)
7	121, 987	5.68e−2	2 (1.8e−3 s)				
8	624, 691	3.80e−2	2 (2.5e−3 s)				
9	1, 260, 214	3.18e−2	2 (3.0e−3 s)				
10	6, 719, 190	1.96e−2	2 (1.6e−2 s)				
11	10, 426, 031	1.84e−2	2 (2.8e−2 s)				
12	24, 509, 387	1.40e−2	1 (3.4e−2 s)				
13	43, 437, 311	1.28e−2	3 (1.3e−1 s)				

Table 2 Nested iteration: Error, number Its of PCG iterations and computational time (Time) for adaptive ($\alpha = 0.25, \beta = 0.75$) and uniform refinements ($\alpha = 0.5, \beta = 0.5$).

4 Conclusions, Generalizations, and Outlook

We have presented a new diffusion regularization that leads to simple diffusion equation as first order optimality condition. The corresponding FE equations can

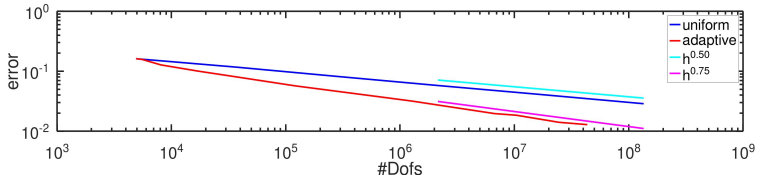

Fig. 2 Nested iteration: Convergence history for uniform and adaptive refinements.

ℓ	#Cores					
	16	32	64	128	256	512
2	23 (9.3e−3 s)	-	-	-	-	-
3	23 (8.7e−2 s)	23 (4.4e−2 s)	23 (1.8e−2 s)	23 (1.2e−2 s)	-	-
4	22 (6.1e−1 s)	22 (3.2e−1 s)	22 (1.8e−1 s)	22 (9.7e−2 s)	22 (4.9e−2 s)	22 (2.0e−2 s)
5	22 (4.9e−0 s)	22 (2.6e−0 s)	22 (1.3e−0 s)	22 (6.9e−1 s)	22 (3.7e−1 s)	22 (2.0e−1 s)
6	-	-	-	22 (5.5e−0 s)	22 (2.7e−0 s)	22 (1.4e−0 s)

Table 3 Number of non-nested PCG iterations and computational time for uniform refinement.

ℓ	#Cores					
	16	32	64	128	256	512
2	2 (1.1e−3 s)	-	-	-	-	-
3	3 (1.4e−2 s)	3 (7.1e−3 s)	3 (2.8e−3 s)	3 (1.6e−3 s)	-	-
4	3 (1.0e−1 s)	3 (5.6e−2 s)	3 (2.9e−2 s)	3 (1.7e−2 s)	3 (8.9e−3 s)	3 (3.7e−3 s)
5	2 (6.2e−1 s)	2 (3.3e−1 s)	2 (1.6e−1 s)	2 (8.8e−2 s)	2 (4.7e−2 s)	2 (3.0e−2 s)
6	-	-	-	2 (6.9e−1 s)	2 (3.4e−1 s)	2 (1.8e−1 s)

Table 4 Number of nested PCG iterations and computational time for uniform refinement.

very efficiently be solved by parallel PCG with diagonal preconditioners provided the the regularization is appropriately chosen. It is possible to include box constraints for both the control and the state [1]. These techniques can be extended to parabolic and hyperbolic OCPs.

Acknowledgements We thank the RICAM for providing us with the high performance cluster Radon1. The last author is supported by the Christian Doppler Research Association.

References

[1] P. Gangl, R. Löscher, and O. Steinbach. Regularization and finite element error estimates for elliptic distributed optimal control problems with energy regularization and state or control constraints. arXiv:2306.15316, 2023.

[2] U. Langer, R. Löscher, O. Steinbach, and H. Yang. An adaptive finite element method for distributed elliptic optimal control problems with variable energy regularization. arXiv:2209.08811, 2022.

[3] U. Langer, R. Löscher, O. Steinbach, and H. Yang. Mass-lumping discretization and solvers for distributed elliptic optimal control problems. arXiv:2304.14664, 2023.

[4] U. Langer, R. Löscher, O. Steinbach, and H. Yang. Robust iterative solvers for algebraic systems arising from elliptic optimal control problems. Berichte aus dem Institut für Angewandte Mathematik, Bericht 2023/2, TU Graz, 2023.

[5] U. Langer, R. Löscher, O. Steinbach, and H. Yang. Robust finite element discretization and solvers for distributed elliptic optimal control problems. *Comput. Methods Appl. Math.*, 23(4):989–1005, 2023.

[6] J. L. Lions. *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*. Dunod Gauthier-Villars, Paris, 1968.

[7] M. Neumüller and O. Steinbach. Regularization error estimates for distributed control problems in energy spaces. *Math. Methods Appl. Sci.*, 44(5):4176–4191, 2021.

[8] F. Tröltzsch. *Optimal control of partial differential equations: Theory, methods and applications*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, 2010.