Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 03–15

SPACE MAPPING FOR OPTIMAL CONTROL OF PARTIAL DIFFERENTIAL EQUATIONS

MICHAEL HINTERMÜLLER AND LUÍS N. VICENTE

ABSTRACT: Solving optimal control problems for nonlinear partial differential equations represents a significant numerical challenge due to the tremendous size and possible model difficulties (*e.g.*, nonlinearities) of the discretized problems. In this paper, a novel space-mapping technique for solving the aforementioned problem class is introduced, analyzed, and tested. The advantage of the space-mapping approach compared to classical multigrid techniques lies in the flexibility of not only using grid coarsening as a model reduction but also employing (perhaps less nonlinear) surrogates. The space mapping is based on a regularization approach which, in contrast to other space-mapping techniques, results in a smooth mapping and, thus, avoids certain irregular situations at kinks. A new Broyden's update formula for the sensitivities of the space map is also introduced. This quasi-Newton update is motivated by the usual secant condition combined with a secant condition resulting from differentiating the space-mapping surrogate. The overall algorithm employs a trust-region framework for global convergence. We highlight some of the issues involved in the computations and we report a few illustrative numerical tests.

KEYWORDS: Space Mapping, Optimal Control of PDEs, Simulation-Based Optimization, PDE-Constrained Optimization, Trust Regions, Quasi-Newton Methods.
AMS SUBJECT CLASSIFICATION (2000): 90C06, 90C26, 90C53.

1. Introduction

Let us assume that we are interested in optimizing some objective related to physical phenomena simulated by a system of differential equations. We might be trying to determine unknown system parameters by matching observable data, or we might want to control properties of the system so that its state matches a given desired profile. Let us further assume that our goal is to minimize a smooth function $g : X \subset \mathbb{R}^n \to \mathbb{R}$, evaluated by accurately solving the discretized system of differential equations that models the underlying physical phenomena. Given the computational complexity involved

The author acknowledges support by the Austrian science fund FWF under the grant SFB "Optimierung und Kontrolle".

Support for this author was provided by Centro de Matemática da Universidade de Coimbra, by FCT under grant POCTI/35059/MAT/2000, by the European Union under grant IST-2000-26063, and by Fundação Calouste Gulbenkian. The author would also like to thank the IBM T.J. Watson Research Center and the Institute for Mathematics and Its Applications for their local support.

in simulating the system, the model g might be expensive to evaluate, and some alternative smooth function $\hat{g} : \hat{X} \subset \mathbb{R}^{\hat{n}} \to \mathbb{R}$ is assumed available at a cheaper cost, by solving the system less accurately or by using some form of surrogate. We will call g the fine model and $\hat{g}$ the coarse model. Similarly, X and $\mathbb{R}^n$ will be called the fine domain and fine space, respectively, and $\hat{X}$ and $\mathbb{R}^{\hat{n}}$ the coarse domain and coarse space, respectively. We will assume that X and $\hat{X}$ are open domains.

The space-mapping technique provides an attractive framework to improve the use of the coarse model $\hat{g}$ as a surrogate for the optimization of the fine model g. The space-mapping surrogate is of the form $\hat{g} \circ P$ where P, the so-called space mapping, attempts to match, in the coarse space, the fine model values and/or their responses.

The space-mapping technique was introduced first by Bandler *et al.* [5] in 1994. The idea of space mapping has been developed along different directions and generalized to a number of contexts. To overcome some of its inherent difficulties, techniques from nonlinear optimization have been incorporated. One of the problems lies in the information necessary to compute the sensitivities (or the Jacobian) of the space mapping which involves, among other things, (possibly expensive) gradient information of the fine model. Bandler *et al.* [6] suggested the use of Broyden's method to construct linear approximations for the space mapping. This space-mapping Broyden's method has been then enhanced by Bakr *et al.* [2] with the application of trust regions for globalization. These and other approaches are reviewed in the papers Bakr *et al.* [3, 4]. See also [13, 26]. The reader is further referred to the special issue on surrogate modeling and space mapping that has been recently edited by Bandler and Madsen [7].

In this paper we are concerned with the application of the space-mapping technique to control problems for partial differential equations (PDEs). The problem under consideration is the following:

$$\text{minimize } \mathcal{J}(y,u) \quad \text{over } (y,u) \in W \times U, \tag{1a}$$

$$\text{subject to } A(y,u)y + C(y,u) = 0 \quad \text{in } \Omega \quad + \text{boundary conditions}, \tag{1b}$$

where $\mathcal{J} : W \times U \to \mathbb{R}$ is a sufficiently smooth objective functional, $A(y,u)$ denotes a second order partial differential operator, $C(y,u)$ is a possibly nonlinear mapping, and Ω is a bounded domain in $\mathbb{R}^n$. Above W, U are appropriate Hilbert spaces and y is referred to as the state variable. The variable u is the control variable. We assume that for every $u \in U$ the state equation in (1b) admits a (unique) solution $y = y(u) \in W$. Using $y(u)$ we can consider the reduced problem

$$\text{minimize } \mathcal{J}_{\text{red}}(u) = \mathcal{J}(y(u), u) \quad \text{over } u \in U \tag{2}$$

instead of (1).

One instance of the model problem (1) is given by

$$\begin{aligned} &\text{minimize } \tfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \tfrac{\delta}{2}\|u\|^2_{L^2(\Omega)} \quad \text{over } (y, u) \in H^1_0(\Omega) \times L^2(\Omega), \\ &\text{subject to } -\Delta y + f(y) = u \quad \text{in } \Omega = (0, 1)^2, \end{aligned}$$

where $y_d \in L^2(\Omega)$, $\delta > 0$ is fixed, and f denotes a sufficiently smooth function; see, *e.g.*, [18, 19]. Obviously, we have $A(y, u) = -\Delta$ and $C(y, u) = f(y) - u$. Here, $H^1_0(\Omega)$ and $L^2(\Omega)$ denote Sobolev and Lebesgue spaces; see, *e.g.*, [12].

Another instance fitting the general model problem (1) is given by output least squares formulations of parameter identification problems. In this case one aims at determining a quantity u, which is not directly accessible to measurements, by fitting the measured data y_d. Frequently, the relation between u and y can be described by a semilinear elliptic PDE:

$$-\operatorname{div}(e(u)\nabla y) + C(y, u) = 0 \text{ in } \Omega, \quad y \in H^1_0(\Omega).$$

Hence, $A(y, u)y = -\operatorname{div}(e(u)\nabla y)$, where e is a possibly nonlinear mapping.

In [21, 22, 24] a multigrid approach (algorithm MG/OPT) to discretizations of minimization problems of type (1) is considered. In a two grid approach, the algorithm combines a prescribed number of iterations of a minimization algorithm on the fine grid with high accuracy solves of a slightly modified problem on the coarse grid. Prolongation and restriction operators achieve the transport of coarse grid solutions to the fine grid and *vice versa*. Our space-mapping approach, however, allows more flexibility in the sense that we may not only consider a coarse grid discretization of the underlying optimization problem, but we can also use a surrogate model which is even simpler to solve than the discretized problem on the coarse grid. This is of particular importance in cases where the coarse grid approximation is still difficult to handle due to, *e.g.*, problematic nonlinearities and/or model complexities. Also, the surrogate idea applies without grid coarsening. In fact, we may want to replace a difficult problem by an approximating simpler one on the same (fine) grid. In addition, the trust-region based approach that we use is globally convergent in the general nonconvex case while MG/OPT, as outlined in [21, 22, 24], requires convexification to enforce descent.

The outline of the paper is as follows. In section 2 we introduce a new smooth space-mapping technique. The key idea is to utilize regularization techniques of Tikhonov-type. This approach is a remedy to problematic non-differentiabilities and nonuniquenesses in space mapping. Section 3 gives a comparison between a nonsmooth space mapping previously suggested by Vicente [27] and the smoothed one introduced in this paper. In section 4 the space-mapping algorithm of Bandler *et al.* [2, 6] is outlined in our space-mapping concept. It is based on a Broyden-type approximation of the sensitivities of the space mapping and on a trust-region type globalization. The focus of section 5 is on a new Broyden's update reflecting the approximation requirements induced by the sensitivities of the space mapping and by its use in the gradient of the space-mapping surrogate. The application of our new space-mapping approach to optimal control problems governed by partial differential equations is the core of section 6. In section 7 we consider computational aspects with respect to coarse and fine model derivatives. A report on numerical test runs is given in section 8. We end this paper with some conclusions and prospects of future work.

2. A new smooth space mapping

We introduce in this paper a new definition of space mapping $P : X \to \hat{X}$ as follows:

$$P(x) = \operatorname{argmin} \left\{ \frac{\alpha}{2} \|\hat{r}(\hat{x}) - r(x)\|_M^2 + \frac{1}{2} [\hat{g}(\hat{x}) - g(x)]^2 \mid \hat{x} \in \hat{X} \right\}, \quad (3)$$

where $\alpha > 0$ is a smoothing parameter whose role will become clear later. We assume that the argmin operator returns a single minimizer, in other words, that P is a point-to-point mapping. Here r and $\hat{r}$ are some operators that map X and $\hat{X}$ into some common space $\mathbb{R}^p$ where the values of fine and coarse variables can be compared against each other. An illustration of r and $\hat{r}$ is given in section 6. In the definition of the space mapping, M is a $p \times p$ symmetric positive definite matrix and $\| \cdot \|_M$ is the ellipsoidal norm defined by $\|z\|_M = \|M^{\frac{1}{2}} z\|_2$.

The space mapping P defines a surrogate $g_P = \hat{g} \circ P$ for the fine model g. One of the aims of space mapping is to minimize the surrogate g_P instead of minimizing g:

$$\text{minimize } g_P(x) = (\hat{g} \circ P)(x) = g(P(x))$$

over $x \in X$.

The effect of the smoothing parameter α becomes clearer by taking a close look at the first order necessary conditions of problem (3):

$$\alpha D_{\hat{r}}^{\top}(\hat{x})M(\hat{r}(\hat{x})-r(x))+[\hat{g}(\hat{x})-g(x)]\nabla\hat{g}(\hat{x}) \;=\; 0, \tag{4}$$

with $\hat{x}=P(x)$. Now, we differentiate (4) with respect to x:

$$\begin{aligned}&\alpha D_{\hat{r}}(\hat{x})^{\top}M\left(D_{\hat{r}}(\hat{x})D_P(x)-D_r(x)\right)+\alpha D_{\hat{r}}^2(\hat{x};M(\hat{r}(\hat{x})-r(x)))D_P(x)+\\ &\nabla\hat{g}(\hat{x})\left(\nabla\hat{g}(\hat{x})^{\top}D_P(x)-\nabla g(x)^{\top}\right)+[\hat{g}(\hat{x})-g(x)]\nabla^2\hat{g}(\hat{x})D_P(x) \;=\; 0,\end{aligned}$$

where $D_{\hat{r}}^2(\hat{x};z)$ is the derivative of $D_{\hat{r}}(\hat{x})z$ with respect to $\hat{x}$. So, one obtains

$$G(x)D_P(x) \;=\; \alpha D_{\hat{r}}(\hat{x})^{\top}MD_r(x)+\nabla\hat{g}(\hat{x})\nabla g(x)^{\top},$$

where

$$\begin{aligned}G(x) \;=\;& \alpha\Big(D_{\hat{r}}(\hat{x})^{\top}MD_{\hat{r}}(\hat{x})+D_{\hat{r}}^2(\hat{x};M(\hat{r}(\hat{x})-r(x)))\Big)+\\ &[\hat{g}(\hat{x})-g(x)]\nabla^2\hat{g}(\hat{x})+\nabla\hat{g}(\hat{x})\nabla\hat{g}(\hat{x})^{\top},\end{aligned}$$

with $\hat{x}=P(x)$. The following theorem summarizes the basic smoothness properties of the surrogate $g_P=\hat{g}\circ P$.

Theorem 2.1. *Let g, $\hat{g}$, r, and $\hat{r}$ be continuously differentiable functions in their domains. Assume that P is a well-defined point-to-point mapping from X to $\hat{X}$.*

(1) *Then g_P is regular in X (i.e., g_P has one-sided directional derivatives in X).*
(2) *In addition, let $\hat{g}$ and $\hat{r}$ be twice continuously differentiable in X. If α is such that $G(x)$ is uniformly nonsingular in X, then g_P is continuously differentiable in X.*

Proof: The fact stated in (1) comes directly from the properties of marginal or value functions (see, *e.g.*, [25]). The proof of (2) lies in the informal derivation given before the theorem. ∎

3. Comparing smooth and nonsmooth approaches

Let us further study the smoothing effect of $\alpha>0$ by comparing our new approach to the approach introduced by Vicente [27], where the space mapping $\bar{P}$ is defined as

$$\bar{P}(x)=\operatorname{argmin}\left\{\frac{1}{2}[\hat{g}(\hat{\bar{x}})-g(x)]^2 \mid \hat{\bar{x}}\in\hat{X}\right\} \tag{5}$$

if $S(x) = \{\hat{\bar{x}} \in \hat{X} \mid \hat{g}(\hat{\bar{x}}) = g(x)\}$ is empty, and as

$$\bar{P}(x) = \operatorname{argmin}\left\{\frac{1}{2}\|\hat{\bar{x}} - x\|_2^2 \quad \text{s.t.} \quad \hat{g}(\hat{\bar{x}}) = g(x) \mid \hat{\bar{x}} \in \hat{X}\right\} \tag{6}$$

if $S(x)$ is nonempty. In this setting it is considered that $n = \hat{n} = p$ and that r and $\hat{r}$ are the identity operators. It is proved in [27] that if g and $\hat{g}$ are continuously differentiable functions and if $\bar{P}$ is point-to-point then $g_{\bar{P}} = \hat{g} \circ \bar{P}$ is a regular function, *i.e.*, a function that has one-sided directional derivatives. The first order necessary conditions for (6) imply, under the constraint qualification $\nabla\hat{g}(\bar{P}(x)) \neq 0$, the existence of a Lagrange multiplier $\lambda(x)$ such that

$$\hat{\bar{x}} - x + \lambda(x)\nabla\hat{g}(\hat{\bar{x}}) \;=\; 0, \tag{7}$$

with $\hat{\bar{x}} = \bar{P}(x)$. Nondifferentiability can only occur on the boundary of the set $\{x \mid S(x) \neq \emptyset\}$. When approaching the boundary of $\{x \mid S(x) \neq \emptyset\}$ from its interior, a kink occurs when $\nabla\hat{g}(\bar{P}(x))$ is approaching zero and $\bar{P}(x)$ is not becoming close to x; in these situations $|\lambda(x)|$ tends to $+\infty$.

The analog of (3) in the setting considered in this section would be

$$P(x) = \operatorname{argmin}\left\{\frac{\alpha}{2}\|\hat{x} - x\|_2^2 + \frac{1}{2}[\hat{g}(\hat{x}) - g(x)]^2 \mid \hat{x} \in \hat{X}\right\}. \tag{8}$$

A variation of this definition has been independently analyzed in [26]. In this case $G(x)$ would reduce to

$$G(x) \;=\; \alpha I + [\hat{g}(\hat{x}) - g(x)]\nabla^2\hat{g}(\hat{x}) + \nabla\hat{g}(\hat{x})\nabla\hat{g}(\hat{x})^\top,$$

with $\hat{x} = P(x)$. The smoothing role of α becomes more evident in this context. Moreover, condition (4) in the simpler case (8) reduces to

$$\alpha(\hat{x} - x) + [\hat{g}(\hat{x}) - g(x)]\nabla\hat{g}(\hat{x}) \;=\; 0, \tag{9}$$

with $\hat{x} = P(x)$.

By comparing (7) and (9) and assuming that $\hat{x} = P(x)$ and $\nabla\hat{g}(P(x))$ are relatively close to $\hat{\bar{x}} = \bar{P}(x)$ and $\nabla\hat{g}(\bar{P}(x))$, respectively, we can gain some insight into the appropriate size for α:

$$\alpha\lambda(x) \;\approx\; \hat{g}(\hat{x}) - g(x).$$

In figure 1 a simple model example is displayed. The fine and coarse models are $g(x) = x^2$ and $\hat{g}(\hat{x}) = (\hat{x} - 1)^2 + 1$, respectively. The upper left plot shows the fine and coarse models together with the surrogate $\hat{g} \circ \bar{P}$, where $\bar{P}$ is given by (5)-(6). In the upper right plot the new smooth surrogate is displayed in dashed lines. The third plot focuses on the behavior near the critical kink at $x = 1$. From the above relation between $\lambda(x)$ and $\hat{g}(\hat{x}) - g(x)$ and the structure of (8) we infer that the smaller α becomes the closer the smooth and nonsmooth surrogates are.

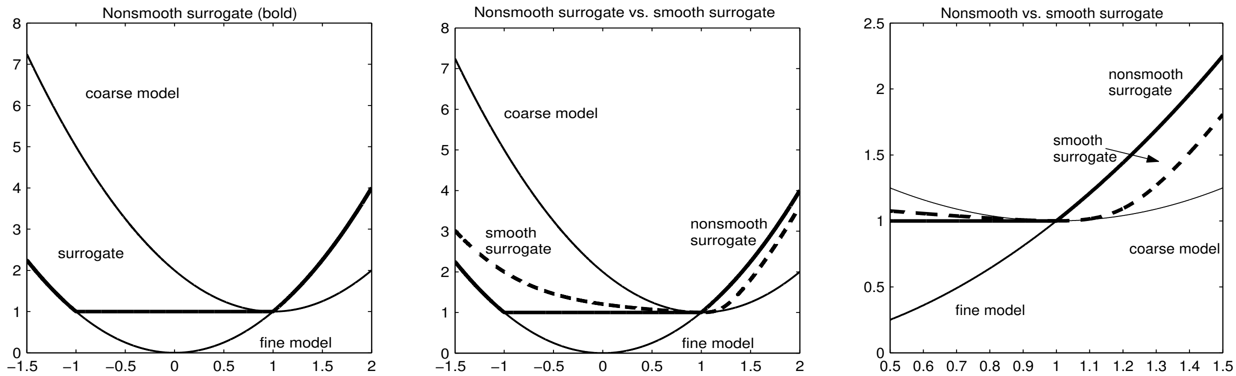

FIGURE 1. Comparison between the nonsmooth surrogate [27] and the new smooth surrogate.

4. Broyden's space-mapping method

As we have shown in section 2, the computation of the sensitivities $D_P(x)$ requires first and second order derivative information of the coarse model and, more importantly, first order derivatives of the fine model. Requiring the gradient of the fine model can pose problems in many practical situations where the evaluation of the fine model is itself very expensive. To overcome this difficulty Bakr *et al.* [6] introduced a Broyden's approach to space mapping, later globalized by Bandler *et al.* [2] with the help of the trust-region technique. This Broyden's space-mapping method using trust regions is described next for the space-mapping definition (3).

The derivative $D_P(x)$ appears both in the formula for the gradient of g_P given by

$$\nabla g_P(x) \;=\; D_P(x)^\top \nabla \hat{g}(\hat{x}) \qquad \text{with } \hat{x} = P(x), \tag{10}$$

and in the local linearization of P at x, along the increment Δx, of the form

$$P(x + \Delta x) \;\approx\; P(x) + D_P(x)\Delta x. \tag{11}$$

The Broyden's updating formula provides a matrix B which can be used to replace $D_P(x)$ in both (10) and (11).

Algorithm 4.1. *Broyden's space-mapping method*
Choose $x_0 \in \mathbb{R}^n$, $\Delta_0 > 0$, $B_0 \in \mathbb{R}^{\hat{n}\times n}$, *and* $\gamma_1, \eta_1 \in (0,1)$.

0. *Compute* $P(x_0)$ *by solving (3) with* $x = x_0$.

For $k = 0, 1, 2, \ldots$

(1) *Compute an approximated solution* Δx_k *for the trust-region subproblem*

$$\textit{minimize} \quad \hat{g}(P(x_k) + B_k \Delta x) \quad \textit{subject to} \quad \|\Delta x\| \le \Delta_k,$$

over $\Delta x \in \mathbb{R}^n$.

(2) *Compute* $P(x_k + \Delta x_k)$ *by solving (3) with* $x = x_k + \Delta x_k$.

(3) *Compute the ratio between actual and predicted reductions:*

$$\rho_k \;=\; \frac{ared(x_k, \Delta x_k)}{pred(x_k, \Delta x_k)} \;=\; \frac{\hat{g}(P(x_k)) - \hat{g}(P(x_k + \Delta x_k))}{\hat{g}(P(x_k)) - \hat{g}(P(x_k) + B_k \Delta x_k)}.$$

(4) *If* $\rho_k \ge \eta_1$ *then* $x_{k+1} = x_k + \Delta x_k$ *and* Δ_{k+1} *is chosen so that* $\Delta_{k+1} \ge \Delta_k$. *In this case, update* B_{k+1} *using Broyden's formula*

$$B_{k+1} \;=\; B_k + \frac{\Delta P_k - B_k \Delta x_k}{\|\Delta x_k\|_2^2} \Delta x_k^\top, \tag{12}$$

where $\Delta P_k = P(x_k + \Delta x_k) - P(x_k)$.

(5) *If* $\rho_k < \eta_1$ *then* $x_{k+1} = x_k$ *and* $\Delta_{k+1} = \gamma_1 \Delta_k$. *Keep* $B_{k+1} = B_k$.

end

The initial values for Δ and B can be given by the classical choices $B_0 = I$ if $\hat{n} = n$ and

$$\Delta_0 \;=\; \frac{\|B_0^\top \nabla \hat{g}(\hat{x}_0)\|_2^2}{|(B_0^\top \nabla \hat{g}(\hat{x}_0))^\top B_0^\top \nabla^2 \hat{g}(\hat{x}_0) B_0 (B_0^\top \nabla \hat{g}(\hat{x}_0))|}. \tag{13}$$

In section 6 we will introduce an appropriate choice for B_0 in a problem context where $\hat{n} \ne n$.

The norm used to define the trust region can be chosen according to practical considerations, but it is typically either the ℓ_2 or the ℓ_∞ norm. The mechanism given in steps 4-5 to update the trust radius is quite elementary but it suffices to prove global convergence of trust-region algorithms. More sophisticated strategies can be found in [11].

The global convergence analysis is described in the next theorem, for which the classical theory of trust regions provides a proof (see [11, Section 8.4] and the references therein). It is not our goal to investigate this subject further but only to list the ingredients necessary for global convergence.

Theorem 4.1. *Let g_P be a continuously differentiable function with uniformly continuous gradient in X and bounded below on $L(x_0) = \{x \in X | g_P(x) \leq g_P(x_0)\}$.*
Consider a sequence $\{x_k\}$ generated by a trust-region method of the form of algorithm 4.1, where the step Δx_k provides a fraction of the Cauchy decrease [11, section 6.3] *and the Hessian used in the trust-region model $\hat{g}(P(x_k) + B_k \Delta x)$ is uniformly bounded. Finally, let B_k satisfy Carter's condition (*[9] *and* [11, section 8.4.1]*) for all k:*

$$\frac{\|\nabla g_P(x_k) - B_k^\top \nabla \hat{g}(P(x_k))\|}{\|B_k^\top \nabla \hat{g}(P(x_k))\|} \leq \frac{\kappa_{mdc}(1-\eta_1)}{2}, \tag{14}$$

where $\kappa_{mdc} \in (0,1)$ is the fraction of the Cauchy decrease achieved by the step Δx_k. Then

$$\lim_{k \to +\infty} \|\nabla g_P(x_k)\| = 0.$$

We remark that the use of the exact sensitivities of P, in other words the use of $B_k = D_P(x_k)$, trivially satisfies (14).

5. A new update for the Broyden's space-mapping method

The Broyden's update (12) is the good Broyden's update for solving systems of nonlinear equations [14]. However, the goal in space mapping is not to solve the system $P(x) = 0$, but rather to exploit the minimization of the surrogate $g_P = \hat{g} \circ P$. Note that the derivative $D_P(x)$ appears in the formula for the gradient of g_P given in (10). Our goal is to modify Broyden's formula to better reflect the use of $D_P(x)$ in the formula for $\nabla g_P(x)$.

The good Broyden's formula is a rank one update B of B_k that satisfies the secant's equation

$$B \Delta x_k = \Delta P_k. \tag{15}$$

The matrix B replaces the role of $D_P(x_k)$ in

$$P(x_k) + D_P(x_k)\Delta x_k \approx P(x_k + \Delta x_k).$$

Now, we also want to use B to approximate the role of $D_P(x_k)$ in (10), from iteration k to $k+1$:

$$\hat{g}(P(x_k)) + \left(D_P(x_k)^\top \nabla \hat{g}(P(x_k))\right)^\top \Delta x_k \approx \hat{g}(P(x_k + \Delta x_k)).$$

This motivation leads to the new secant's condition

$$\nabla \hat{g}_k^\top B \Delta x_k \;=\; \Delta \hat{g}_k, \tag{16}$$

where $\nabla \hat{g}_k = \nabla \hat{g}(P(x_k))$ and $\Delta \hat{g}_k = \hat{g}(P(x_k + \Delta x_k)) - \hat{g}(P(x_k))$. The simultaneous satisfaction of (15) and (16) is possible only if

$$\nabla \hat{g}_k^\top \Delta P_k \;=\; \Delta \hat{g}_k,$$

a condition that would in turn reflect

$$\nabla \hat{g}(P(x_k))^\top \left(P(x_k + \Delta x_k) - P(x_k)\right) \;\approx\; \hat{g}(P(x_k + \Delta x_k)) - \hat{g}(P(x_k)). \tag{17}$$

It is unlikely that (17) is satisfied, and therefore unreasonable to compute B based on the simultaneous satisfaction of (15) and (16).

A way to circumvent this problem is to relax (15), by determining B as the optimal solution of

$$\text{minimize} \;\; \frac{1}{2}\|B\Delta x_k - \Delta P_k\|_2^2 \;\; \text{subject to} \;\; \nabla \hat{g}_k^\top B \Delta x_k \;=\; \Delta \hat{g}_k \tag{18}$$

over $B \in \mathbb{R}^{\hat{n} \times n}$. The following proposition gives a characterization of the optimal solution of problem (18).

Proposition 5.1. *Let Δx_k and $\nabla \hat{g}_k$ be nonzero vectors. The optimal solution B^* of (18) satisfies*

$$B^* \Delta x_k - \Delta P_k \;=\; \frac{\Delta \hat{g}_k - \nabla \hat{g}_k^\top \Delta P_k}{\|\nabla \hat{g}_k\|_2^2} \nabla \hat{g}_k.$$

Proof: Let us rewrite problem (18) as

$$\text{minimize} \;\; \frac{1}{2}\|V_k v - \Delta P_k\|_2^2 \;\; \text{subject to} \;\; \nabla \hat{g}_k^\top V_k v \;=\; \Delta \hat{g}_k \tag{19}$$

over $v \in \mathbb{R}^{\hat{n}n}$, with the help of the change of variables $v_{(i-1)n+j} = B_{ij}$, $i = 1, \ldots, \hat{n}$, $j = 1, \ldots, n$. The rows of the $\hat{n} \times \hat{n}n$ matrix V_k are composed by the elements of Δx_k and by $(\hat{n}-1)n$ zeros. The matrix V_k has full row rank because $\Delta x_k \neq 0$.

From the assumptions on Δx_k and $\nabla \hat{g}_k$ we know that $V_k^\top \nabla \hat{g}_k \neq 0$. The first order necessary conditions for (19) can then be stated by assuming the existence of a Lagrange multiplier λ_k such that

$$V_k^\top V_k v - V_k^\top \Delta P_k + \lambda_k V_k^\top \nabla \hat{g}_k \;=\; 0.$$

Since $V_k^\top$ has full column rank, we obtain

$$V_k v - \Delta P_k + \lambda_k \nabla \hat{g}_k \;=\; 0. \tag{20}$$

By multiplying this equation on the left by $\nabla \hat{g}_k^\top$ and using the problem's constraint in (19), we get

$$\Delta \hat{g}_k - \nabla \hat{g}_k^\top \Delta P_k + \lambda_k \|\nabla \hat{g}_k\|_2^2 \;=\; 0. \tag{21}$$

Thus, (20) and (21) together imply

$$V_k v - \Delta P_k \;=\; \frac{\Delta \hat{g}_k - \nabla \hat{g}_k^\top \Delta P_k}{\|\nabla \hat{g}_k\|_2^2} \nabla \hat{g}_k.$$

The proof is completed by returning to the formulation (18). ■

Proposition 5.1 suggests a perturbation for the right-hand side of the secant's equation (15):

$$B \Delta x_k \;=\; \Delta P_k + \frac{\Delta \hat{g}_k - \nabla \hat{g}_k^\top \Delta P_k}{\|\nabla \hat{g}_k\|_2^2} \nabla \hat{g}_k.$$

For numerical purposes it might be advantageous to reduce the size of the new term that is added to ΔP_k:

$$\widetilde{\Delta P_k} \;=\; \Delta P_k + \sigma_k \frac{\Delta \hat{g}_k - \nabla \hat{g}_k^\top \Delta P_k}{\|\nabla \hat{g}_k\|_2^2} \nabla \hat{g}_k, \tag{22}$$

with $\sigma_k \in (0, 1]$, depending on the impact that $\hat{g}$ has in the definition of the space mapping P. The new Broyden's update is therefore given by

$$B_{k+1} \;=\; B_k + \frac{\widetilde{\Delta P_k} - B_k \Delta x_k}{\|\Delta x_k\|_2^2} \Delta x_k^\top.$$

Notice that if we allow $\sigma_k = 0$ in (22), then the new Broyden's update becomes the classical Broyden's update as discussed, *e.g.*, in [14]. In section 8 we will see that, for appropriate choices of $\sigma_k \in (0, 1)$, the new Broyden's update leads to better numerical results than the classical one for an instance problem of optimal control of PDEs.

6. Application of the space-mapping method for optimal control of PDEs

In this section, we apply the space-mapping approach introduced in section 2 to the reduced problem (2). Let h and H with $H \geq h$ denote mesh sizes of discretizations of (2) yielding the fine model space $U_h = \mathbb{R}^{n_h}$ and the coarse model space $U_H = \mathbb{R}^{n_H}$. We have $n = n_h$, $X = U_h$, $\hat{n} = n_H$, and $\hat{X} = U_H$. For the ease of exposition we only argue for an L^2-setting with standard inner product. Thus, by rescaling on the discrete level we essentially have to deal with ℓ_2 inner products only.

We introduce now discretized versions of the reduced problem (2). Let $y^h(u^h)$ denote the solution of the discretized PDE in (1b) with mesh size h. Moreover, let J^h be an appropriate discretization of the cost functional $\mathcal{J}$. Then

$$J^h_{\mathrm{red}}(u^h) = J^h(y^h(u^h), u^h).$$

In an analogous way one obtains the coarse model J^H_{red}:

$$J^H_{\mathrm{red}}(u^H) = J^H(y^H(u^H), u^H).$$

In order to simplify the notation and to make it similar to the one used in section 2, we will use J, J_{red}, y, and u for fine model quantities, and $\hat{J}$, $\hat{J}_{\mathrm{red}}$, $\hat{y}$, and $\hat{u}$ for coarse model quantities.

Since $\dim U_h$ and $\dim U_H$ may differ, we define the linear restriction operator

$$I^h_H : U_h \longrightarrow U_H,$$

which maps a fine model quantity to a coarse model quantity. Typically, the definition of I^h_H depends on (infinite dimensional) regularity properties of the control variable. Here we adopt restriction operators coming from multigrid methods; see [15, 23, 28].

The introduction of I^h_H enables us to define the space mapping $P : U_h \to U_H$ by

$$\begin{aligned} P(u) = \operatorname{argmin}\{ \tfrac{\alpha_1}{2} \|\hat{y}(\hat{u}) - K^h_H y(u)\|^2_{\hat{M}_{\hat{y}}} + \tfrac{\alpha_2}{2}\|\hat{u} - I^h_H u\|^2_{\hat{M}_{\hat{u}}} + \\ \tfrac{\alpha_3}{2}|\hat{J}_{\mathrm{red}}(\hat{u}) - J_{\mathrm{red}}(u)|^2 \mid \hat{u} \in U_H\}, \end{aligned} \tag{23}$$

with fixed $\alpha_1, \alpha_2, \alpha_3 \geq 0$, and $\alpha_1 + \alpha_2 + \alpha_3 > 0$. Above, $\hat{M}_{\hat{y}}$ represent a symmetric positive definite matrix resulting from discretizing a function space norm yielding $\|\hat{y}\|^2_{\hat{M}_{\hat{y}}} = \hat{y}^T \hat{M}_{\hat{y}} \hat{y}$; analogously for $\|\cdot\|^2_{\hat{M}_{\hat{u}}}$. Moreover, K^h_H denotes a restriction operator, possibly different from I^h_H. Throughout the rest of this paper we assume that $P(u)$ is single valued for every $u \in U_h$. Instead of $\hat{y}(\hat{u}) - K^h_H y(u)$ we could have used $\hat{R}\hat{y}(\hat{u}) - K^h_H R y(u)$, restricting the matching of the coarse and fine state variables to parts of its discretized domains.

The parallel to what has been introduced in section 2 is made by setting

$$x = u, \quad \hat{x} = \hat{u}, \quad p = n^{\hat{y}}_H + n_H, \quad \alpha = \alpha_1 = \alpha_2,$$

$$r(u) = \begin{pmatrix} K_H^h y(u) \\ I_H^h u \end{pmatrix}, \quad \hat{r}(\hat{u}) = \begin{pmatrix} \hat{y}(\hat{u}) \\ \hat{u} \end{pmatrix}, \quad \text{and}$$

$$M = \begin{pmatrix} M_{\hat{y}} & 0 \\ 0 & M_{\hat{u}} \end{pmatrix},$$

where $n_H^{\hat{y}}$ is the dimension of $\hat{y}(\hat{u})$.

Following the space-mapping philosophy presented in the previous sections, we now replace the problem of finding a solution to the fine model

$$\text{minimize } J_{\text{red}}(u) \quad \text{over } u \in U_h, \tag{24}$$

by finding a solution of the problem involving the surrogate $J_{\text{red}}^P = \hat{J}_{\text{red}} \circ P$:

$$\text{minimize } J_{\text{red}}^P(u) = \hat{J}_{\text{red}}(P(u)) \quad \text{over } u \in U_h. \tag{25}$$

When solving (25) numerically, one has to evaluate J_{red}^P repeatedly which, in turn, requires repeated evaluations of the fine model and repeated solutions of the minimization problem (23). As we have seen before, given a fixed fine model point u, the computational effort can be reduced by considering the following first order approximation of the space mapping

$$P(u+s) \approx P_\ell(u;s) = P(u) + D_P(u)s \in U_H, \tag{26}$$

with $D_P : U_h \mapsto \mathbb{R}^{n_H \times n_h}$ denoting the Jacobian of P. Consequently, J_{red}^P is approximated around u by

$$\hat{J}_{\text{red}}(P(u+s)) \approx \hat{J}_{\text{red}}(P_\ell(u;s)). \tag{27}$$

The evaluation of $\hat{J}_{\text{red}}(P_\ell(u;s))$ in (27) requires only the computation of the action of $D_P(u)$ on s.

The calculation of the gradient of $J_{\text{red}}^P(u)$ in (25) involves $D_P(u)$ in the following way

$$\nabla J_{\text{red}}^P(u) \;=\; D_P(u)^\top \nabla \hat{J}_{\text{red}}(\hat{u}) \qquad \text{with } \hat{u} = P(u). \tag{28}$$

If we use the approximation (26) for P centered at u as a way of computing a step s by minimizing $\hat{J}_{\text{red}}(P_\ell(u;s))$ in (27), one also needs the evaluation of (28). In fact, $\nabla J_{\text{red}}^P(u)$ is the gradient of $\hat{J}_{\text{red}}(P_\ell(u;s))$ with respect to the increment s. The evaluation of $\nabla J_{\text{red}}^P(u)$ requires the computation of the action of $D_P(u)^\top$ on $\nabla \hat{J}_{\text{red}}(\hat{u})$.

6.1. Computation of the sensitivities of the space mapping. In order to characterize $D_P(u)$ or $D_P(u)s$ we need to consider the first order necessary conditions of (23), given by

$$\begin{aligned} &\alpha_1 D_{\hat{y}}(\hat{u})^\top \hat{M}_{\hat{y}} \left(\hat{y}(\hat{u}) - K_H^h y(u)\right) + \alpha_2 \hat{M}_{\hat{u}} \left(\hat{u} - I_H^h u\right) + \\ &\alpha_3 [\hat{J}_{\mathrm{red}}(\hat{u}) - J_{\mathrm{red}}(u)] \nabla \hat{J}_{\mathrm{red}}(\hat{u}) \;=\; 0, \end{aligned} \tag{29}$$

with $\hat{u} = P(u)$. Above $D_{\hat{y}}(\hat{u})$ denotes the Jacobian of $\hat{y}(\hat{u})$ with respect to $\hat{u}$. We obtain the characterizing equation for the sensitivities of the space mapping P by differentiation of (29) with respect to u. This results in

$$\begin{aligned} &\alpha_1 D_{\hat{y}}(\hat{u})^\top \hat{M}_{\hat{y}} \left(D_{\hat{y}}(\hat{u}) D_P(u) - K_H^h D_y(u)\right) + \\ &\alpha_1 H_{\hat{y}}(\hat{u}; \hat{M}_{\hat{y}}(\hat{y}(\hat{u}) - K_H^h y(u))) D_P(u) + \\ &\alpha_2 \hat{M}_{\hat{u}} \left(D_P(u) - I_H^h\right) + \\ &\quad \alpha_3 \nabla \hat{J}_{\mathrm{red}}(\hat{u}) \left(\nabla \hat{J}_{\mathrm{red}}(\hat{u})^\top D_P(u) - \nabla J_{\mathrm{red}}(u)^\top\right) + \\ &\quad \alpha_3 [\hat{J}_{\mathrm{red}}(\hat{u}) - J_{\mathrm{red}}(u)] H_{\hat{J}_{\mathrm{red}}}(\hat{u}) D_P(u) \;=\; 0, \end{aligned} \tag{30}$$

with $\hat{u} = P(u)$. In the above equation $H_{\hat{y}}(\hat{u}; z)$ denotes the derivative of $D_{\hat{y}}(\hat{u})z$ with respect to $\hat{u}$, $D_y(u)$ represents the Jacobian of $y(u)$ with respect to u, and $H_{\hat{J}_{\mathrm{red}}}$ is the Hessian of $\hat{J}_{\mathrm{red}}$. Let

$$\begin{aligned} G_{P(u)} = &\,\alpha_1 D_{\hat{y}}(\hat{u})^\top \hat{M}_{\hat{y}} D_{\hat{y}}(\hat{u}) + \alpha_1 H_{\hat{y}}(\hat{u}; \hat{M}_{\hat{y}}(\hat{y}(\hat{u}) - K_H^h y(u))) + \\ &\alpha_2 \hat{M}_{\hat{u}} + \\ &\alpha_3 [\hat{J}_{\mathrm{red}}(\hat{u}) - J_{\mathrm{red}}(u)] H_{\hat{J}_{\mathrm{red}}}(\hat{u}) + \alpha_3 \nabla \hat{J}_{\mathrm{red}}(\hat{u}) \nabla \hat{J}_{\mathrm{red}}(\hat{u})^\top, \end{aligned}$$

with $\hat{u} = P(u)$, and

$$\begin{aligned} r_{P(u)} = &\,\alpha_1 D_{\hat{y}}(\hat{u})^\top \hat{M}_{\hat{y}} K_H^h D_y(u) + \alpha_2 \hat{M}_{\hat{u}} I_H^h + \\ &\alpha_3 \nabla \hat{J}_{\mathrm{red}}(\hat{u}) \nabla J_{\mathrm{red}}(u)^\top, \end{aligned}$$

also with $\hat{u} = P(u)$. This notation allows us to write (30) in a more compact way as

$$G_{P(u)} D_P(u) = r_{P(u)}.$$

For given $s \in U_h$ let us define

$$s_{P(u)} = D_P(u) s \quad \text{and} \quad r_{P(u)}^s = r_{P(u)} s.$$

Then, the action of $D_P(u)$ on s, given by $s_{P(u)} = D_P(u)s$, satisfies

$$G_{P(u)} s_{P(u)} = r^s_{P(u)} \quad \text{in } U_H.$$

We point out that in the case where the PDE on the coarse level is linear, one has $H_{\hat{y}} = 0$ and the expression for $G_{P(u)}$ simplifies considerably.

6.2. A practical Broyden's space-mapping method for optimal control of PDEs. Now we adapt the Broyden's space-mapping method, introduced in [2, 6] and described in algorithm 4.1, to optimal control of partial differential equations using the setting and notation chosen in this paper for these problems.

Before we describe the algorithm, we need to adapt some of the notation of sections 4 and 5 to the optimal control framework. In fact, let

$$\nabla \hat{J}^k_{\text{red}} \;=\; \nabla \hat{J}_{\text{red}}(P(u_k)),$$

and

$$\Delta \hat{J}^k_{\text{red}} = \hat{J}_{\text{red}}(P(u_k + \Delta u_k)) - \hat{J}_{\text{red}}(P(u_k)).$$

As before, we have $\Delta P_k = P(u_k + \Delta u_k) - P(u_k)$, and we use the new Broyden's update introduced in section 5 with $\sigma_k \in (0, 1]$:

$$\widetilde{\Delta P_k} \;=\; \Delta P_k + \sigma_k \frac{\Delta \hat{J}^k_{\text{red}} - (\nabla \hat{J}^k_{\text{red}})^\top \Delta P_k}{\|\nabla \hat{J}^k_{\text{red}}\|_2^2} \nabla \hat{J}^k_{\text{red}}. \tag{31}$$

Algorithm 6.1. *Broyden's space-mapping method for optimal control of PDEs*

Choose $u_0 \in \mathbb{R}^n = \mathbb{R}^{n_h}$, $\Delta_0 > 0$, $B_0 \in \mathbb{R}^{\hat{n} \times n} = \mathbb{R}^{n_H \times n_h}$, *and* $\gamma_1, \eta_1 \in (0, 1)$.

0. *Compute* $P(u_0)$ *by solving* (23) *with* $u = u_0$.

For $k = 0, 1, 2, \ldots$

(1) *Compute an approximated solution* Δu_k *for the trust-region subproblem*

$$\textit{minimize} \quad \hat{g}(P(u_k) + B_k \Delta u) \quad \textit{subject to} \quad \|\Delta u\| \le \Delta_k, \tag{32}$$

over $\Delta u \in \mathbb{R}^n = \mathbb{R}^{n_h}$.

(2) *Compute* $P(u_k + \Delta u_k)$ *by solving* (23) *with* $u = u_k + \Delta u_k$.

(3) *Compute the ratio between actual and predicted reductions:*

$$\rho_k \;=\; \frac{ared(u_k, \Delta u_k)}{pred(u_k, \Delta u_k)} \;=\; \frac{\hat{g}(P(u_k)) - \hat{g}(P(u_k + \Delta u_k))}{\hat{g}(P(u_k)) - \hat{g}(P(u_k) + B_k \Delta u_k)}.$$

(4) *If* $\rho_k \geq \eta_1$ *then* $u_{k+1} = u_k + \Delta u_k$ *and* Δ_{k+1} *is chosen so that* $\Delta_{k+1} \geq \Delta_k$. *In this case, update* B_{k+1} *using Broyden's formula*

$$B_{k+1} \;=\; B_k + \frac{\widetilde{\Delta P_k} - B_k \Delta u_k}{\|\Delta u_k\|_2^2} \Delta u_k^\top, \tag{33}$$

where $\widetilde{\Delta P_k}$ *is given by* (31) *with* $\Delta P_k = P(u_k + \Delta u_k) - P(u_k)$ *and* $\sigma_k \in (0, 1]$. *Put* $P(u_{k+1}) = P(u_k + \Delta u_k)$.

(5) *If* $\rho_k < \eta_1$ *then* $u_{k+1} = u_k$ *and* $\Delta_{k+1} = \gamma_1 \Delta_k$. *Keep* $B_{k+1} = B_k$ *and* $P(u_{k+1}) = P(u_k)$.

end

The comments made about the norm used to shape the trust region and about the mechanisms to manage the size of the trust radius remain pertinent here. When $H = h$ the initial value for B can be given by the classical choice $B_0 = I_{n_h}$, with I_{n_h} the $n_h \times n_h$ identity matrix. When $H > h$ we can choose $B_0 = I_H^h$. The initial trust radius might be set as in (13).

In analogy to the Broyden's space-mapping method of section 4 one would expect that

$$\hat{g}(P(u_k)) = \hat{J}_{\mathrm{red}}(P(u_k)) \tag{34}$$

and

$$\hat{g}(P(u_k) + B_k \Delta u) = \hat{J}_{\mathrm{red}}(P(u_k) + B_k \Delta u). \tag{35}$$

However, in the case where $H > h$, this last choice would result in an underdetermined problem in step 1 of the algorithm, in the sense that $\Delta u \in \mathbb{R}^{n_h}$ is a fine grid quantity whereas $\hat{g}$ is defined in the coarse grid setting (yielding a singular Hessian in $\hat{J}_{\mathrm{red}}(P(u_k)+B_k\Delta u)$). There exist two immediate remedies to this situation.

(i) One possibility is to use

$$\hat{g}(P(u_k) + B_k \Delta u) = \hat{J}_{\mathrm{red}}(P(u_k) + B_k \Delta u) + \frac{\gamma}{2} \|u_k + \Delta u - u_d\|_{M_u}^2 \tag{36}$$

and

$$\hat{g}(P(u_k)) = \hat{J}_{\mathrm{red}}(P(u_k)) + \frac{\gamma}{2} \|u_k - u_d\|_{M_u}^2, \tag{37}$$

where $\gamma > 0$ and u_d denotes some reference value for the expected optimal control. For instance, u_d can be obtained by prolongating coarse grid solutions (easy to obtain) to the fine grid. The parameter γ plays the role of a regularization parameter which penalizes deviations of $u_k + \Delta u$ from u_d. In our numerical tests, γ is chosen according to the mesh sizes H and h in the following way:

$$\gamma = c_\gamma(1 - h/H) \quad \text{with} \quad 0 < c_\gamma \ll 1.$$

Note that when $H = h$ we have $\gamma = 0$ and no regularization takes place (and the coarse model in step 1 is likely not under-determined in the sense discussed above).

(ii) An alternative remedy, using the original choices (34)-(35), is given by solving an approximate problem of the type

$$\text{minimize} \quad \hat{g}(P(u_k) + B_k I_h^H \Delta\hat{u}) \quad \text{subject to} \quad \|\Delta\hat{u}\| \leq \Delta_k \tag{38}$$

instead of problem (32) in step 1. By using a restriction operator, the independent variable $\Delta\hat{u}$ is mapped to a fine grid quantity and the new problem is usually well-determined. Again, whenever $h = H$ we may choose $I_h^H = I_{n_h}$, and problem (38) becomes the original problem (32).

Both remedies have additional costs. The first one requires the computation of the reference value u_d and the second one the application of the restriction operator I_h^H. However, in the latter case only a n_H-dimensional problem has to be solved.

7. Computation of coarse and fine model derivatives

7.1. Adjoint calculation of the coarse model gradient and Hessian. The computation of the gradient $\nabla\hat{J}_{\text{red}}(\hat{u})$ can be carried out by the so-called adjoint technique. In the sequel we briefly explain some of the details.

Let $\hat{E}(\hat{y}, \hat{u}) = 0$ denote the discretized PDE on the coarse grid. Further let $\hat{E}_{\hat{y}}$, $\hat{E}_{\hat{u}}$ denote the partial Jacobians of $\hat{E}$ with respect to $\hat{y}$ and $\hat{u}$, respectively. From the assumption that the state equation admits an unique solution $y(u)$ for $u \in U$, we infer that there exists an unique $\hat{y}(\hat{u})$ such that $\hat{E}(\hat{y}(\hat{u}), \hat{u}) = 0$ and that $\hat{E}_{\hat{y}}(\hat{y}(\hat{u}), \hat{u})$ is invertible (at least for sufficiently small H). Differentiation of $\hat{E}(\hat{y}(\hat{u}), \hat{u}) = 0$ with respect to $\hat{u}$ yields

$$\hat{E}_{\hat{y}}(\hat{y}(\hat{u}), \hat{u}) D_{\hat{y}}(\hat{u}) + \hat{E}_{\hat{u}}(\hat{y}(\hat{u}), \hat{u}) = 0. \tag{39}$$

Hence, we obtain from (39)

$$D_{\hat{y}}(\hat{u}) = -\hat{E}_{\hat{y}}(\hat{y}(\hat{u}), \hat{u})^{-1} \hat{E}_{\hat{u}}(\hat{y}(\hat{u}), \hat{u}). \tag{40}$$

From the definition of $\hat{J}_{\text{red}}$ we deduce that

$$\nabla\hat{J}_{\text{red}}(\hat{u}) = \nabla_{\hat{u}}\hat{J}(\hat{y}(\hat{u}), \hat{u}) + D_{\hat{y}}(\hat{u})^\top \nabla_{\hat{y}}\hat{J}(\hat{y}(\hat{u}), \hat{u}), \tag{41}$$

where $\nabla_{\hat{y}}\hat{J}$, $\nabla_{\hat{u}}\hat{J}$ represent the partial derivatives of $\hat{J}$ with respect to the first and second argument, evaluated in (41) at $(\hat{y}(\hat{u}), \hat{u})$. Utilizing (40) in (41) yields

$$\nabla \hat{J}_{\rm red}(\hat{u}) = \nabla_{\hat{u}}\hat{J}(\hat{y}(\hat{u}), \hat{u}) + \hat{E}_{\hat{u}}(\hat{y}(\hat{u}), \hat{u})^\top \hat{p}(\hat{u}) \tag{42}$$

with

$$\hat{E}_{\hat{y}}(\hat{y}(\hat{u}), \hat{u})^\top \hat{p}(\hat{u}) = -\nabla_{\hat{y}}\hat{J}(\hat{y}(\hat{u}), \hat{u}). \tag{43}$$

Equation (43) is the so-called (discrete) adjoint equation. For computing $\nabla \hat{J}_{\rm red}(\hat{u})$ one can proceed as follows: Given $\hat{u}$ solve the state equation for $\hat{y}(\hat{u})$, then solve the adjoint equation (43) for $\hat{p}(\hat{u})$, and finally compute the gradient according to (42).

By using the definition

$$\hat{W}(\hat{y}(\hat{u}), \hat{u}) \;=\; \begin{pmatrix} D_{\hat{y}}(\hat{u}) \\ I_{n_H} \end{pmatrix},$$

it is possible to rewrite (41) as

$$\nabla \hat{J}_{\rm red}(\hat{u}) \;=\; \hat{W}(\hat{y}(\hat{u}), \hat{u})^\top \nabla \hat{J}(\hat{y}(\hat{u}), \hat{u}).$$

Also, it is possible to show (see, *e.g.*, [16]) that the Hessian of the coarse model $\hat{J}_{\rm red}(\hat{u})$ is given by

$$H_{\hat{J}_{\rm red}}(\hat{u}) \;=\; H_{\hat{y}}(\hat{u}; \nabla_{\hat{y}}\hat{J}(\hat{y}(\hat{u}), \hat{u})) + \hat{W}(\hat{y}(\hat{u}), \hat{u})^\top H_{\hat{J}}(\hat{y}(\hat{u}), \hat{u})\hat{W}(\hat{y}(\hat{u}), \hat{u}).$$

When the state equation in (1b) is linear, *i.e.*, when $\hat{E}(\hat{y}, \hat{u}) = \hat{L}\hat{y} + \hat{M}\hat{u} - \hat{f}$, where $\hat{M}$ and $\hat{L}$ are suitable matrices with $\hat{L}$ invertible and $\hat{f}$ a coarse model vector, and when the cross derivatives $\hat{J}_{\hat{y}\hat{u}}(\cdot)$ and $\hat{J}_{\hat{u}\hat{y}}(\cdot)$ are zero, one can simplify considerably the expression for the Hessian of the coarse model $\hat{J}_{\rm red}(\hat{u})$. The assumption $\hat{J}_{\hat{y}\hat{u}}(\cdot) = \hat{J}_{\hat{u}\hat{y}}(\cdot) = 0$ is satisfied for the commonly used objective functional of tracking type, *i.e.*, for

$$\mathcal{J}(y, u) = \tfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \tfrac{\delta}{2}\|u\|^2_{L^2(\Omega)},$$

with $y_d \in L^2(\Omega)$ and $\delta > 0$ fixed. Under the simplified assumptions of this paragraph, the model Hessian becomes

$$H_{\hat{J}_{\rm red}}(\hat{u}) = \hat{L}^{-T}\hat{J}_{\hat{y}\hat{y}}(\hat{L}^{-1}(\hat{f} - \hat{M}\hat{u}), \hat{u})\hat{L}^{-1} + \hat{J}_{\hat{u}\hat{u}}(\hat{L}^{-1}(\hat{f} - \hat{M}\hat{u}), \hat{u}).$$

7.2. Approximation of the fine model gradient. The gradient ∇J_{red} of the fine model can be computed also by using the adjoint technique of section 7.1. Per each gradient evaluation, this technique requires one solve for the (possibly nonlinear) state equation and one for the (linear) adjoint equation.

Since we are working on the fine model these evaluations might be extremely costly. One way to reduce or avoid fine model solves is based on restriction operators I_H^h and their analogues, prolongation operators I_h^H.

Alternatively, each fine model gradient evaluation can be calculated by a hybrid approach based on a fine model adjoint solve and a coarse model solve of the state equation.

In the sequel we describe these techniques for computing ∇J_{red} depending on whether $H > h$ (and both fine and coarse models are nonlinear) or $H = h$ (and the coarse model is linear). We present this material because of its relevance in the context of this paper despite the fact that we do not make use of any approximation to the gradient of the fine model in our numerical testing.

7.2.1. *The case* $\mathbf{H} > \mathbf{h}$ *(fine and coarse models are nonlinear).* Let I_h^H denote the (linear) prolongation operator from U_H to U_h. Analogously, one also introduces K_h^H. In the case where the fine and the coarse models of the PDE are nonlinear, a suitable approximation of the gradient is given by

$$\nabla J_{\text{red}}^{\text{app}}(u) \; = \; I_h^H \nabla \hat{J}_{\text{red}}(I_H^h u), \tag{44}$$

i.e., we restrict the fine model point $u \in U_h$ to the coarse setting U_H by using I_H^h, evaluate the gradient on the coarse level by means of the adjoint technique, and then we prolongate the coarse model gradient back to the fine model setting with the help of I_h^H.

Alternatively, one can use a hybrid approach which combines coarse and fine model solves and which is still numerically less expensive than the full fine model approach. The hybrid technique is particularly useful when the fine model involves nonlinearities. In fact, we can compute

$$\nabla J_{\text{red}}^{\text{app}}(u) \; = \; \nabla_u J(K_h^H \hat{y}(I_H^h u), u) + E_u(K_h^H \hat{y}(I_H^h u), u)^\top p(u) \tag{45}$$

with

$$E_y(K_h^H \hat{y}(I_H^h u), u)^\top p(u) \; = \; -\nabla_y J(K_h^H \hat{y}(I_H^h u), u). \tag{46}$$

The advantage of this strategy is related to the fact that the nonlinear state equation must be solved only on the coarse grid for a given $I_H^h u$. On the fine grid, one has to solve the (linear) adjoint equation. Typically, solving linear equations is significantly less expensive than computing solutions to nonlinear ones. Thus, the hybrid approach is less expensive than computing ∇J_{red} by the adjoint technique on the fine grid.

Using (44), or the hybrid approach in (45) and (46), yields the approximate sensitivity $D_P^{\mathrm{app}}(u)$ and the approximate action $s_{P(u)}^{\mathrm{app}}$. We remark that the accuracies of these approximations can by controlled by tuning the mesh size H. In fact, in the extreme case $H = h$ with $I_H^h = I_h^H = I_{n_h}$ (I_{n_h} the $n_h \times n_h$ identity matrix) only exact quantities are computed.

7.2.2. *The case* $\mathbf{H} = \mathbf{h}$ *(coarse model is linear).* In this case (45) would require the full adjoint technique on the fine grid. In order to reduce the computational burden, one may consider as the coarse model a linear approximation of the discretized PDE. If the linear equation can be solved efficiently (*e.g.*, by fast Fourier transformation techniques), then a suitable approximate gradient is given by

$$\nabla J_{\mathrm{red}}^{\mathrm{app}}(u) \;=\; \nabla_u J(\hat{y}_L(u), u) + E_u(\hat{y}_L(u), u)^\top p(u)$$

with

$$E_y(\hat{y}_L(u), u)^\top p(u) \;=\; -\nabla_y J(\hat{y}_L(u), u)$$

and $\hat{y}_L(u)$ denoting the solution of the linear coarse model $\hat{E}(\hat{y}, u) = 0$. Here we assume $\hat{u} = u$. Clearly, the approximation properties depend now on the error between the linear coarse model and the nonlinear fine model.

8. Numerical experiments

Let us now report some numerical results attained by the Broyden's space-mapping method for the optimal control of PDEs. Our test examples are of the following type:

$$\text{minimize } \tfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \tfrac{\delta}{2}\|u\|^2_{L^2(\Omega)} \quad \text{over } (y, u) \in H_0^1(\Omega) \times L^2(\Omega), \tag{47a}$$

$$\text{subject to } -\nu\Delta y + f(y) = u \quad \text{in } \Omega = (0,1)^2, \tag{47b}$$

with $y_d \in L^2(\Omega)$ and $\nu, \delta > 0$. Here f denotes some nonlinear mapping in y. Note that the parameter $\nu > 0$ allows us to emphasize the nonlinear term $f(y)$ by considering $0 < \nu \ll 1$.

We use a standard five point stencil for discretizing the Laplacian with homogeneous Dirichlet boundary conditions. The prolongation operators I_h^H, K_h^H and restriction operators I_H^h, K_H^h are chosen as follows: Motivated by an *a posteriori* analysis (in function spaces) of a solution (y, u) to our control problem, see, *e.g.*, [1], we choose $K_h^H = I_h^H$ and $K_H^h = I_H^h$. The interpolation from the coarse to the fine grid, *i.e.*, I_h^H, is achieved by a nine point prolongation. Its stencil is symbolized by

$$\begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}.$$

The restriction I_H^h is the adjoint of the nine point prolongation with symbol

$$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}.$$

For more details on prolongation and restriction operators of the above kind we refer the reader to, *e.g.*, [15].

For the numerical solution of the discretized counterpart of the nonlinear partial differential equation involved in (47) we use the Newton-CG method [20]. The discrete linearized PDE as well as the discrete adjoint equation are solved by means of the CG method.

In the Broyden's space-mapping method for optimal control, *i.e.*, algorithm 6.1, we use the following adjustment strategy for the trust radius Δ_k: Let $0 < \eta_1 \le \eta_2 < 1$, $\gamma_1 \in (0, 1)$, and $\xi_1 > 1$ be given. If $\rho_k \ge \eta_2$, then accept the current step and enlarge the trust radius by $\Delta_{k+1} = \xi_1 \Delta_k$. If $\eta_1 \le \rho_k < \eta_2$, then the current step is accepted and the trust radius is kept, *i.e.*, $\Delta_{k+1} = \Delta_k$. Finally, whenever $\rho_k < \eta_1$, then the trust radius is reduced by $\Delta_{k+1} = \gamma_1 \Delta_k$ without accepting the current step. In the examples reported below we used $\eta_1 = 10^{-5}$, $\eta_2 = 10^{-1}$, $\gamma_1 = 0.25$, and $\xi_1 = 2$. We initialize the trust radius as $\Delta_0 = 50$.

The Broyden's update procedure of algorithm 6.1 is based on a full limited memory version of Broyden's method [8]. In fact, since n_h is typically very large, and B_k tends to be a dense matrix, storing B_k is infeasible in the context of optimal control problems for PDEs. Rather we store the vectors $\{\Delta u_i\}_{i=0}^k$ and $\{\widetilde{\Delta P_i}\}_{i=0}^k$ and perform the product $B_k v$ of the Broyden's matrix B_k by a vector $v \in \mathbb{R}^{n_h}$ using vector$\times$vector-multiplications only. We initialize the Broyden's matrix as $B_0 = I_H^h$.

In the examples below the fine model consists of the fully nonlinear PDE discretized on the fine grid with mesh size h, resulting in n_h unknowns in the reduced fine model problem (24) with $U_h = \mathbb{R}^{n_h}$. The coarse model is given by the discretization of the linearized PDE on the coarse grid with mesh size H. The linearization is performed with respect to $\hat{y}_k^\dagger$ which results in

$$\nu \hat{A}\hat{y} + D_{\hat{f}}(\hat{y}_k^\dagger)\hat{y} = \hat{u} - \hat{f}(\hat{y}_k^\dagger) + D_{\hat{f}}(\hat{y}_k^\dagger)\hat{y}_k^\dagger,$$

with $\hat{u}, \hat{y}, \hat{y}_k^\dagger \in \mathbb{R}^{n_H}$. Above, the $n_H \times n_H$-matrix $\hat{A}$ represents the discretization of the operator $-\Delta$ with homogeneous Dirichlet boundary conditions on the coarse grid with mesh size H. In all test runs reported below, we chose $u_0 = 0$, $\hat{y}_0^\dagger = 0$, and $\hat{y}_k^\dagger = I_H^h y_k$, where y_k solves the discretized nonlinear PDE for $u = u_k$ on the fine grid.

We use (36)-(37) for $\hat{g}$ in the trust-region subproblem in algorithm 6.1. The reference value u_d is chosen as $u_d = 0$ in all iterations. The corresponding regularization parameter γ is specified in the examples below. The norm used to define the trust region is the ℓ_∞ one.

Algorithm 6.1 was stopped when

$$\max\{|pred(u_k, \Delta u_k)|, H\|\nabla \hat{J}_{\text{red}}(P(u_k))\|_2, h\|\Delta u_k\|_2\} \leq \texttt{tol},$$

where $\texttt{tol}= \epsilon_1 H\|\nabla \hat{J}_{\text{red}}(P(u_0))\|_2 + \epsilon_2$, with $0 < \epsilon_2 \ll \epsilon_1$. Unless otherwise specified, we chose $\epsilon_1 = 10^{-5}$ and $\epsilon_2 = 10^{-14}$.

We have used a Matlab implementation of the interior-point trust-region algorithm (`IPTR`) [17], provided by these authors, for the solution of the minimization problems subject to bounds on the variables, namely problems (23) and (32), the latter being a quadratic programming problem with simple bounds. This algorithm uses an affine-scaling approach [10] to deal with the simple bounds. The linear algebra involved in the solution of the trust-region subproblems is matrix-free and relies on the use of the linear conjugate gradient method.

8.1. Example 1. The first example is related to a simplified Ginzburg-Landau model for superconductivity [18, 19]. The data are as follows:

$$y_d = \tfrac{1}{6}\sin(2\pi x_1)\sin(2\pi x_2)\exp(2x_1), \quad f(y) = y^3 + y,$$

and $\delta = \nu = 10^{-3}$. Figure 2 shows the optimal control and the optimal state of (47), with data as specified before, computed on a 255×255 grid.

In table 1 we report the results obtained from our space-mapping algorithm 6.1 with $\alpha_1 = \alpha_2 = 100$, $\alpha_3 = 10^{-5}$, and $\gamma = 10^{-3}(1 - h/H)$. The parameter σ_k (see (31)) was set to $\sigma_k = 0.1$ for all k. By *level* we denote

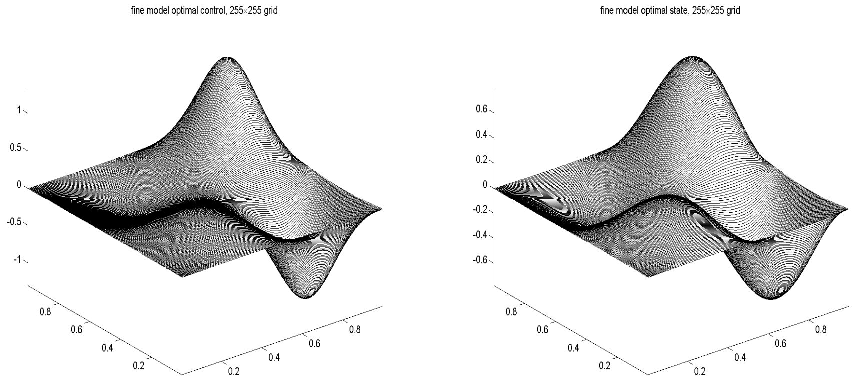

FIGURE 2. Optimal control (left) and optimal state (right) for the simplified Ginzburg-Landau model on a 255×255 grid.

the number of grid coarsenings, *i.e.*, $H = 2^{level}h$. Furthermore, #it denotes the number of iterations until successful termination, and *CPU-ratio* represents the ration between the CPU-time required by the space-mapping method *vs.* the CPU-time elapsed by `IPTR` when applied to the fine model problem. The subscript *fine* corresponds to the results obtained from `IPTR` for the fine model. From table 1 we can see that the new space-mapping

n_h	*level*	n_H	# it	*CPU-ratio*	#it$_{fine}$
127^2	3	15^2	4	0.094	6
255^2	4	15^2	4	0.066	7

TABLE 1. Results of space mapping *vs.* fine model solution for example 1.

method consumes less iterations and less CPU-time than `IPTR` for the solution of the fine model problem. We also point out that the CPU-time of the space-mapping approach increases as the levels of coarsening decrease. For instance, if $h = 1/128$ and $H = 1/32$, *i.e.*, if we have two levels of coarsening, then *CPU-ratio* $= 0.153$ compared to 0.094 for three levels of coarsening. By decreasing the levels of coarsening, as expected, the accuracy of the space-mapping solution is increased. This fact is further discussed in the following paragraph.

In figure 3 we display the controls obtained by the space-mapping method for $n_h = 127^2$ (left graph) and $n_h = 255^2$ (right graph), respectively. Furthermore, in figure 3, we plot the difference in absolute value between the optimal controls obtained from the fine model and from our space-mapping technique. The graphs in the second row of figure 3 show that the error

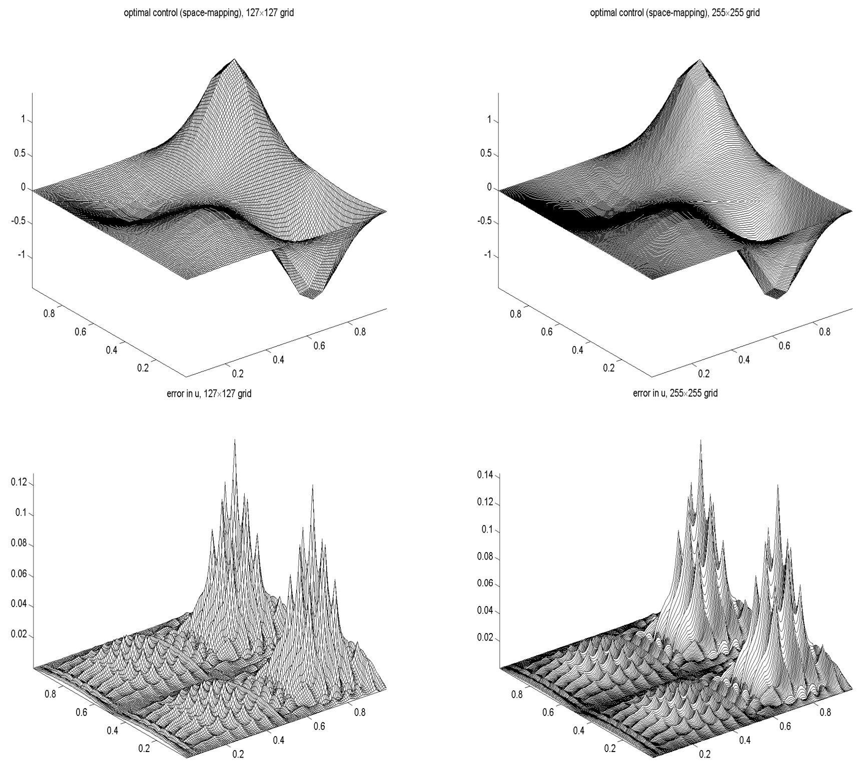

FIGURE 3. Optimal controls obtained by algorithm 6.1 (upper plots) and differences to the fine model solutions (lower plots) for $n_h = 127^2$ (left column), $n_h = 255^2$ (right column), and $n_H = 15^2$, respectively.

between the space-mapping solution and the true solution of the fine model behaves rather stably with respect to the coarse level. Indeed, the coarse level for both results is $H = 1/16$ while the fine levels are $h = 1/128$ and $h = 1/256$, respectively. In figure 4 we further investigate the dependence of the error on the levels of coarsening. Now we use $h = 1/128$ and $H = 1/32$ which corresponds to *level* $= 2$ (compared to *level* $= 3$ previously). From the graphs in figure 4 we conclude that the error is significantly reduced. Also, the graph of the space-mapping solution appears to be smoother compared to the ones in figure 3. This is related to the fact that the restriction and prolongation operators approach the unit matrix as the level of coarsening decreases.

In the following we briefly comment on the effect of the new Broyden's update (33). In table 2 we compare the new Broyden's update with $\sigma_k = 0.1$

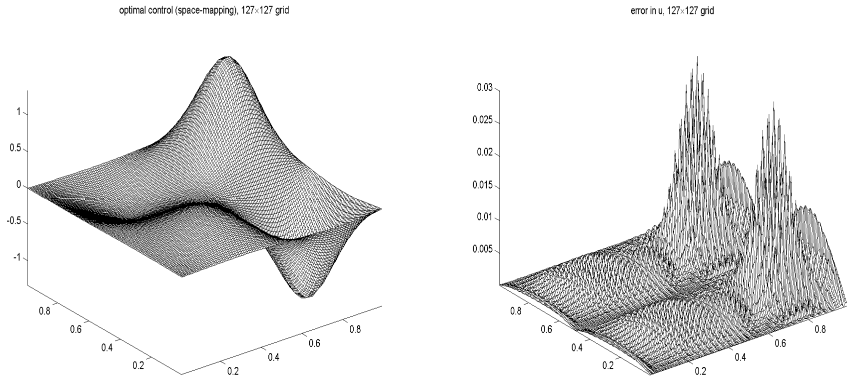

FIGURE 4. Optimal control obtained by algorithm 6.1 for $h = 1/128$ (left plot) and the difference to the fine model solution (right plot) for $level = 2$.

to the classical Broyden's update, *i.e.*, $\sigma_k = 0$ for all k. The results in

σ_k	n_h	*level*	n_H	# it	*CPU-ratio*
0.1	127^2	3	15^2	4	0.094
0.1	255^2	4	15^2	4	0.066
0.0	127^2	3	15^2	4	0.107
0.0	255^2	4	15^2	5	0.081

TABLE 2. Comparison between the new and the classical Broyden's update for example 1.

table 2 indicate that the new Broyden's update reduces the computation time and, according to our numerical experience, sometimes also the number of iterations of the new space-mapping algorithm. In general, we found that the behavior of the new method depends on the choice of σ_k. In our test runs for example 1 the choices $\sigma_k \in [0.1, 0.01)$ yielded results comparable to $\sigma_k = 0.1$ for all k. For $\sigma_k < 0.01$ there was no significant difference between the new and the classical Broyden's update. The choice $\sigma_k > 0.1$ typically degraded the performance of the method when compared to runs with $\sigma_k = 0.1$.

8.2. Example 2. The following example shows that the space-mapping method benefits from eventual evaluations of the fine model. In fact, when considering steps 2 and 4 of algorithm 6.1 we find that every computation of the space mapping P is based on the matching of fine model solves. This fact is highlighted in our definition (23) of P, where u and $y(u)$ are fine model quantities. As a consequence, we expect that the space-mapping solution yields a better approximation to the fine model solution than, *e.g.*, prolongated coarse model solutions.

The data for example 2 are like for example 1 except for f. Now we have

$$f(y) = y^3 + y + f_0, \text{ with } f_0(x) = \tfrac{1}{6}\sin(20\pi x_1)\sin(20\pi x_2)\exp(2x_1).$$

The zero order term f_0 induces oscillations to the optimal control as it can be seen from figure 5, which displays the optimal control and the corresponding optimal state for the fine model problem on a 127×127-grid. We ran algo-

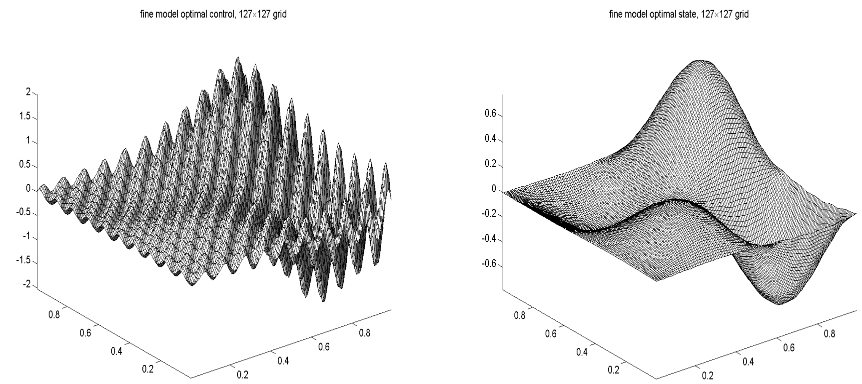

FIGURE 5. Optimal control (left) and optimal state (right) for example 2 on a 127×127 grid.

rithm 6.1 with $\alpha_1 = 0.75$, $\alpha_2 = 0.1$, $\alpha_3 = 10^{-5}$, $\gamma = 2.25 \cdot 10^{-2}(1 - h/H)$, $\sigma_k = 10^{-3}$ for all k, and $level = 3$.

Figure 6 shows (in the upper left part) the prolongated coarse grid optimal control, *i.e.*, $u_{\text{pro}} = K_h^H \bar{u}$, where $\bar{u}$ is the optimal solution of the nonlinear optimal control problem on the coarse grid, as well as the space-mapping solution (in the upper right part). The figures in the lower part show the top view of both solutions (left for prolongated and right for space mapping). First of all, we point out that there is a significant difference in the scale of both solutions, as it can be seen from the graphs in the first row of figure 6. The space-mapping scaling is significantly closer to the fine model one. The lower plots show that the space-mapping solution identified more of the fine model resolution. This is a clear indication of what we mentioned before, in the sense that fine model information has a beneficial impact on the quality of the solution obtained by space mapping.

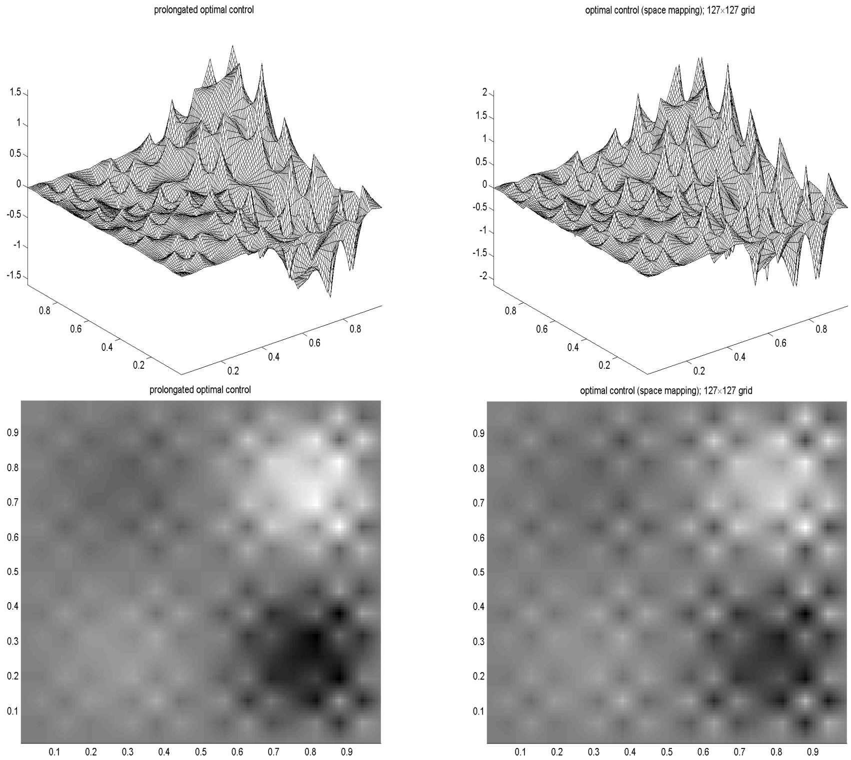

FIGURE 6. Prolongated optimal control (left column) and space-mapping solution (right column) for example 2.

9. Conclusions and future work

In this paper we have investigated the use of the space-mapping technique in the numerical solution of optimal control problems governed by partial differential equations. We have identified a space-mapping framework for this purpose that allows the integration of different coarse models, arising from linearizing and/or coarsening the fine model. The new definition for the space mapping that we introduced uses the concept of Tikhonov-type regularization as a way of finding the coarse (control and state) variables closest to some corresponding fine model values. We have also suggested a new Broyden's update to approximate the derivatives of the space mapping, with broad applicability to most of the existent space-mapping approaches.

A number of issues need to be further investigated. In this paper we have not considered, for instance, optimal control problems with constraints on the control variables like simple bounds. Adapting our approach to cover this case is relatively straightforward but it would add another layer of complexity in the numerical computations.

A topic for future research is the use of more than one coarse model in the space-mapping approach. The existence of, say, two coarse models with increasing level of accuracy and cost of evaluation is an appealing idea in some application problems. Another aspect that has not been considered in this paper is the appropriate use of different optimization algorithms for coarse and fine models along the spirit of multigrid methods.

References

[1] N. Arada, E. Casas, and F. Tröltzsch, *Error estimates for the numerical approximation of a semilinear elliptic control problem*, Comp. Optim. Appl., 23 (2002), pp. 201–229.

[2] M. H. Bakr, J. W. Bandler, R. M. Biernacki, S. H. Chen, and K. Madsen, *A trust region agressive space mapping algorithm for EM optimization*, IEEE Trans. Microwave Theory Tech., 46 (1998), pp. 2412–2425.

[3] M. H. Bakr, J. W. Bandler, K. Madsen, and J. Søndergaard, *Review of the space mapping approach to engineering optimization and modeling*, Optimization and Engineering, 1 (2000), pp. 241–276.

[4] ——, *An introduction to the space mapping technique*, Optimization and Engineering, 2 (2002), pp. 369–384.

[5] J. W. Bandler, R. M. Biernacki, S. H. Chen, P. A. Grobelny, and R. H. Hemmers, *Space mapping technique for electromagnetic optimization*, IEEE Trans. Microwave Theory Tech., 42 (1994), pp. 2536–2544.

[6] J. W. Bandler, R. M. Biernacki, S. H. Chen, R. H. Hemmers, and K. Madsen, *Electromagnetic optimization exploiting agressive space mapping*, IEEE Trans. Microwave Theory Tech., 43 (1995), pp. 2874–2882.

[7] J. W. Bandler and K. Madsen, *Editorial – Surrogate Modelling and Space Mapping for Engineering Optimization*, Optimization and Engineering, 2 (2002), pp. 367–368.

[8] R. Byrd, J. Nocedal, and R. Schnabel, *Representations of quasi-Newton matrices and their use in limited-memory methods*, Mathematical Programming A, 63 (1994), pp. 129–156.

[9] R. G. Carter, *On the global convergence of trust region algorithms using inexact gradient information*, SIAM J. Numer. Anal., 28 (1991), pp. 251–265.

[10] T. F. Coleman and Y. Li, *An interior trust region approach for nonlinear minimization subject to bounds*, SIAM J. Optim., 6 (1996), pp. 418–445.

[11] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust-Region Methods*, MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2000.

[12] R. Dautray and J.-L. Lions, *Analyse Mathématique et Calcul Numérique 3*, Masson, Paris, 1987.

[13] J. E. Dennis, *Surrogate Modelling and Space Mapping for Engineering Optimization. A summary of the Danish Technical University November 2000 Workshop*, Tech. Report TR00–35, Department of Computational and Applied Mathematics, Rice University, 2000.

[14] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice–Hall, Englewood Cliffs, (republished by SIAM, Philadelphia, in 1996, as Classics in Applied Mathematics, 16), 1983.

[15] W. Hackbusch, *Multigrid Methods and Applications*, Series in Computational Mathematics 4, Springer Verlag, Berlin, 1985.

[16] M. Heinkenschloss, *Projected sequential quadratic programming methods*, SIAM J. Optim., 6 (1996), pp. 373–417.

[17] M. Heinkenschloss, M. Ulbrich, and S. Ulbrich, *Superlinear and quadratic convergence of affine-scaling interior-point Newton methods for problems with simple bounds without strict complementarity assumption*, Math. Programming, 86 (1999), pp. 615–635.

[18] M. Hintermüller, *On a globalized augmented Lagrangian-SQP algorithm for nonlinear optimal control problems with box constraints*, in Fast Solution Methods for Discretized Optimization Problems, K. H. Hoffmann, R. H. W. Hoppe, and V. Schulz, eds., Birkhäuser Verlag, Basel, 2001, pp. 139–153.

[19] K. Ito and K. Kunisch, *Augmented Lagrangian–SQP methods for nonlinear optimal control problems of tracking type*, SIAM J. Control Optim., 34 (1996), pp. 874–891.

[20] C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, SIAM, Philadelphia, 1995.

[21] R. M. Lewis and S. G. Nash, *A multigrid approach to the optimization of systems governed by differential equations*, Tech. Report AIAA-2000-4890, American Institute of Aeronautics and Astronautics, NASA Langley Research Center, 2000.

[22] ——, *Model problems for the multigrid optimization of systems governed by differential equations*. Submitted for publication, 2002.

[23] S. F. McCormick, *Multigrid Methods*, Frontiers in Applied Mathematics, SIAM, Philadelphia, 1987.

[24] S. G. Nash, *A multigrid approach to discretized optimization problems*, Optim. Methods Softw., 14 (2000), pp. 99–116.

[25] R. T. Rockafellar, *Directional differentiability of the optimal value function in a nonlinear programming problem*, Math. Programming Stud., 21 (1984), pp. 213–226.

[26] J. Søndergaard, *Optimization Using Surrogate Models — by the Space Mapping Technique*, PhD thesis, Department of Mathematical Modelling, Technical University of Denmark, 2003.

[27] L. N. Vicente, *Space mapping: Models, sensitivities, and trust-regions methods*, Optimization and Engineering, 4 (2003), pp. 159–175.

[28] P. Wesseling, *An Introduction to Multigrid Methods*, Wiley Interscience, Chichester, 1992.

Michael Hintermüller
Department of Mathematics, University of Graz, A-8010 Graz, Austria (michael.hintermueller@kfunigraz.ac.at).

Luís N. Vicente
Departamento de Matemática, Universidade de Coimbra, 3001-454 Coimbra, Portugal (lnv@mat.uc.pt).