Fast Iterative Solution of Saddle Point Problems in Optimal Control Based on Wavelets

Angela Kunoth*

March 25, 2001

Abstract

In this paper, wavelet techniques are employed for the fast numerical solution of a control problem governed by an elliptic boundary value problem with boundary control. A quadratic cost functional involving *natural norms* of the state and the control is to be minimized. Firstly the constraint, the elliptic boundary value problem, is formulated in an appropriate weak form that allows to handle varying boundary conditions explicitly: the boundary conditions are treated by Lagrange multipliers, leading to a saddle point problem. This is combined with a fictitious domain approach in order to cover also more complicated boundaries.

Deviating from standard approaches, we then use (biorthogonal) wavelets to derive an *equivalent* infinite discretized control problem which involves only ℓ_2–norms and –operators. Classical methods from optimization yield the corresponding optimality conditions in terms of two weakly coupled (still infinite) saddle point problems for which a unique solution exists. For deriving finite–dimensional systems which are uniformly invertible, stability of the discretizations has to be ensured. This together with the ℓ_2–setting circumvents the problem of *preconditioning*: all operators have *uniformly bounded* condition numbers independent of the discretization.

In order to numerically solve the resulting (finite–dimensional) linear system of the weakly coupled saddle point problems, a fully iterative method is proposed which can be viewed as an *inexact gradient* scheme. It consists of a gradient algorithm as an outer iteration which alternatingly picks the two saddle point problems, and an inner iteration to solve each of the saddle point problems, exemplified in terms of the Uzawa algorithm. It is proved here that this strategy converges, provided that the inner systems are solved sufficiently well. Moreover, since the system matrix is well–conditioned, it is shown that in combination with a *nested iteration strategy* this iteration is asymptotically *optimal* in the sense that it provides the solution on discretization level J with an overall amount of arithmetic operations that is proportional to the number of unknowns N_J on that level.

Finally, numerical results are provided.

*Institut für Angewandte Mathematik, Universität Bonn, Wegelerstr. 6, 53115 Bonn, Germany, kunoth@iam.uni-bonn.de, www.iam.uni-bonn.de/~kunoth.

Key words: Optimal control, elliptic boundary value problem, Lagrange multiplier, saddle point problem, wavelets, infinite ℓ_2–system, preconditioning, coupled system, convergence of iterative procedure, inexact gradient method, Uzawa algorithm, nested iteration.
AMS subject classification (MSC2000): 65K10, 65N99, 93B40.

1 Introduction

In order to illustrate the main ideas, consider the following example of an optimal control problem with PDE constraints: minimize a quadratic functional of the form

$$\mathcal{J}(y,u) = \tfrac{\omega}{2}\|y - y_\square\|^2_{V_1} + \|u\|^2_{V_2}, \tag{1.1}$$

where the *state* y and the *control* u are coupled by the elliptic boundary value problem

$$\begin{aligned} -\nabla\cdot(\mathbf{a}\nabla y) + ky &= f && \text{in } \Omega,\\ y &= u && \text{on } \Gamma,\\ (\mathbf{a}\nabla y)\cdot\mathbf{n} &= 0 && \text{on } \Gamma_N. \end{aligned} \tag{1.2}$$

Here $\Omega \subset I\!R^n$ is a domain with (Lipschitz) boundary $\partial\Omega = \Gamma \cup \Gamma_N$, where Γ is a set of positive Lebesgue measure on which the control is exerted. The term $\mathbf{n} = \mathbf{n}(\mathbf{x})$ denotes the outward normal at $\mathbf{x} \in \partial\Omega \setminus \Gamma$. Furthermore, the coefficient $\mathbf{a}(\mathbf{x}) = (a_{i,j}(\mathbf{x}))_{i,j}$ is assumed to be uniformly positive definite on Ω, and $ky := \mathbf{b}\cdot\nabla y + a_0 y$ with $b_i, a_0 \in L_\infty(\Omega)$. The norms appearing in (1.1) are usually L_2 or H^1 norms on the domain or on some part of the boundary, balanced by some weight $\omega > 0$, and $y_\square$ is some prescribed value.

One can also formulate the above problem as follows: given the right hand side f, determine the boundary control u such that (1.2) has a unique solution y and $\mathcal{J}$ in (1.1) is minimized.

1.1 The Fictitious Domain–Lagrange Multiplier Approach

When one only has to solve an elliptic boundary value problem of the form (1.2) with given right hand side and *given* boundary conditions, a standard approach to set up a corresponding weak formulation is to use subspaces of Sobolev spaces as approximation spaces. These consist of functions which already satisfy the homogeneous boundary conditions, and inhomogeneous boundary conditions are handled by an appropriate homogenization. However, in the above case of varying Dirichlet boundary conditions represented by the control, this procedure would have to be repeated whenever the control changes. To circumvent this, one can handle the essential boundary condition on Γ by employing a *Lagrange multiplier* which results in a *saddle point problem* [Ba]. In view of possible extensions to time–dependent problems with moving boundaries, the Lagrange multiplier approach seems even more attractive when connected with a *fictitious domain*.

To describe this combined approach, let $\square$ be a simple domain, preferably a cube like $(0,1)^n$, which will be the *fictitious domain* containing Ω,

$$\Omega \subseteq \square. \tag{1.3}$$

We explicitly express the restriction of a function v to Γ, usually denoted by $(\cdot)|_\Gamma$, in terms of the *trace operator* B,

$$Bv := v|_\Gamma. \tag{1.4}$$

Recall that for any $v \in H^1(\Box)$ its trace Bv is known to be in $H^{1/2}(\Gamma)$, see e.g. [Gr]. Thus, the bilinear form

$$b(v,q) := \langle Bv, q\rangle_{H^{1/2}(\Gamma)\times(H^{1/2}(\Gamma))'} \tag{1.5}$$

is well–defined on $H^1(\Box) \times (H^{1/2}(\Gamma))'$. Recall that if Γ coincides with $\partial\Omega$, the dual space of $H^{1/2}(\Gamma)$ is $H^{-1/2}(\Gamma)$. Denote the extension of $f \in (H^1(\Omega))'$ to $\Box$ again by f, and correspondingly for the coefficients $\mathbf{a}$, k.

More general than (1.2), in the sequel we will always consider elliptic boundary value problems in the following weak form: Given $(f,u) \in (H^1(\Box))' \times H^{1/2}(\Gamma)$, find $(y,p) \in H^1(\Box) \times (H^{1/2}(\Gamma))'$ such that

$$\begin{aligned} a(y,v) + b(v,p) &= \langle f, v\rangle_{(H^1(\Box))'\times H^1(\Box)} && \text{for all } v \in H^1(\Box),\\ b(y,q) &= \langle u, q\rangle_{H^{1/2}(\Gamma)\times(H^{1/2}(\Gamma))'} && \text{for all } q \in (H^{1/2}(\Gamma))' \end{aligned} \tag{1.6}$$

holds. Here the bilinear form $a(\cdot,\cdot) : H^1(\Box) \times H^1(\Box) \to I\!\!R$ is defined by

$$a(v,w) := \int_\Box (\mathbf{a}\nabla v \cdot \nabla w + (kv)w)\, d\mathbf{x}. \tag{1.7}$$

(If $\Box = \Omega$, (1.6) is the standard weak formulation of the above equations (1.2) with natural boundary conditions on Γ_N.) Note that the equations (1.6) are the optimality conditions of the *saddle point problem*

$$\inf_{v\in H^1(\Box)} \sup_{q\in H^{-1/2}(\Gamma)} \tfrac{1}{2}a(v,v) - \langle f, v\rangle_{(H^1(\Box))'\times H^1(\Box)} + b(v,q) - \langle u, q\rangle_{H^{1/2}(\Gamma)\times(H^{1/2}(\Gamma))'}. \tag{1.8}$$

For this reason, the system (1.6) is also called a *saddle point problem.*

It will be convenient for later purposes to write (1.6) in operator form. Let the linear operator A be defined by

$$\langle Av, w\rangle_{(H^1(\Box))'\times H^1(\Box)} := a(v,w), \tag{1.9}$$

and let B' be the dual of B given by

$$\langle Bv, q\rangle_{H^{1/2}(\Gamma)\times(H^{1/2}(\Gamma))'} = \langle v, B'q\rangle_{H^1(\Box)\times(H^1(\Box))'} = b(v,q). \tag{1.10}$$

Then (1.6) is rewritten as follows. Given $(f,u) \in (H^1(\Box))' \times H^{1/2}(\Gamma)$, find $(y,p) \in H^1(\Box) \times (H^{1/2}(\Gamma))'$ such that

$$\mathcal{L}\begin{pmatrix} y \\ p \end{pmatrix} := \begin{pmatrix} A & B' \\ B & 0 \end{pmatrix}\begin{pmatrix} y \\ p \end{pmatrix} = \begin{pmatrix} f \\ u \end{pmatrix} \tag{1.11}$$

is satisfied. We call this the *Fictitious Domain–Lagrange Multiplier System.*

The newly introduced variable p in (1.6), the *Lagrange multiplier*, can be interpreted as the *stress* on the boundary. Thus, the formulation provides, in addition to the separate treatment of the Dirichlet boundary conditions, more information on the solution y.

The separate treatment of the boundary condition in (1.11) makes it particularly attractive for control problems with boundary control. In addition, using the fictitious domain approach means that for the *numerical solution* the main amount of work, namely, the set–up and application of a discrete finite representation of A, can be executed on a very simple domain where fast and very efficient methods and preconditioners exist.

Fictitious domain approaches have been employed for some time in different contexts, see e.g. [As, GPP, Ri]. Lately they have attracted increasing interest since, for 3D problems, mesh generation of Ω can become quite expensive. Of course, the above mentioned advantages will only prevail if in a numerical scheme there is no need to adapt the discretization on $\square$ to the particular current domain geometry Ω. This is essential when dealing with moving boundaries. But even for fixed domains this is still very desirable when aligning the mesh with the boundary of Ω causes strong distortions. For this reason, in the discretization procedure we keep the discretization of the Lagrange multipliers *geometrically* independent of the (fixed) trial spaces on $\square$. This means *not to distort* the different meshes on $\square$ and Γ as it is done in some finite element contexts, see e.g. [Pi]. Such results for general domains in arbitrary dimensions based on (appropriately selecting refinement levels in) multiscale concepts have recently been obtained in [DK]. A wavelet method for the Fictitious Domain—Lagrange Multiplier Approach has been described in [K3].

1.2 Optimal Control Problems and Appropriate Representers

The prototype of a control problem that we will discuss here can now be posed as follows: minimize (1.1),

$$\mathcal{J}(y,u) = \tfrac{\omega}{2}\|y - y_\square\|_{V_1}^2 + \|u\|_{V_2}^2,$$

involving *natural norms* for y and u subject to the constraints (1.11). The natural norm for y would be either the $H^1(\square)$–norm or, when taking the restriction of y to some part of the boundary $\Gamma_y \subseteq \partial\Omega$, the $H^{1/2}(\Gamma_y)$–norm. We treat both cases simultaneously and introduce for this purpose the *observation space*

$$Z \in \{H^1(\square), H^{1/2}(\Gamma_y)\} \tag{1.12}$$

with corresponding norm $\|\cdot\|_Z$. The boundary control u is naturally measured in the $H^{1/2}(\Gamma)$–norm and we take the full space $H^{1/2}(\Gamma)$ as the set of *admissible* controls.

In these terms, the reference control problem is to minimize

$$\mathcal{J}(y,u) = \tfrac{\omega}{2}\|y - y_\square\|_Z^2 + \|u\|_{H^{1/2}(\Gamma)}^2 \tag{1.13}$$

subject to the PDE constraints in weak form (1.11),

$$\mathcal{L}\begin{pmatrix} y \\ p \end{pmatrix}\begin{pmatrix} y \\ p \end{pmatrix} = \begin{pmatrix} f \\ u \end{pmatrix}.$$

In any case, we will be dealing with at least one 'broken' norm. Since the evaluation of fractional norms raises difficulties in finite element or finite difference approaches, such norms are avoided by either being content with taking the L_2–norm, or by possibly assuming higher regularity when choosing the H^1–norm.

On the other hand, often in such control problems the target is not so much to compute the precise quantity of the optimization functional but rather to control the *qualitative* behaviour of the minimization. This allows us to use norms in the cost functional which are *equivalent* to the previously used natural function norms.

In taking advantage of *wavelet concepts*, we derive below a *representer* of the quadratic functional (1.13) which

- is on one hand more convenient for subsequent computations, and
- on the other hand still meets the physical effects of the minimization functional.

The core problem results then in a functional involving *only* ℓ_2–norms with constraints also in form of an (infinite–dimensional) ℓ_2–system. The specific form of the control problem will emerge from the following discussions and (appropriately scaled) wavelet representations. This entails that asympotically the need for preconditioning is completely avoided: the control problem is 'shifted' into a problem involving only operators which are well-conditioned in ℓ_2–metric (in which for the numerical solution the iteration is performed). The representer of the cost functional that we will be using will be such that the dual saddle point problem derived by the standard machinery for obtaining the optimality conditions from e.g. [Li] will just be the adjoint problem.

We hasten to add that *any* other norms in the cost functional can be treated by the above technique as long as they can be evaluated in terms of wavelets. This is the case for scales of Sobolev or even Besov spaces on domains and boundaries or parts of boundaries. Employing norms other than the natural ones would, however, also include some Riesz mappings as in [K2, DKS]. For simplicity of the presentation, we dispense with these generalizations here.

For preparing the optimality conditions for a numerical solution, the resulting system of infinite linear operator equations has to be appropriately truncated. In typical Galerkin approaches, this corresponds to picking finite–dimensional subspaces of the involved function spaces and representing the operator in terms of the basis functions. When doing so, one needs to ensure that certain stability conditions are satisfied such that the discrete finite–dimensional system is stably invertible. Sufficient conditions that ensure stable discretizations for one saddle point problem have been derived in [DK, K2]. Together with the stability of the discretization for the primal saddle point problem, we will see that these results applied to the specific form of the quadratic cost functional (involving representers of natural and possibly also trace norms) also entails stability for the whole system of weakly coupled saddle point problems.

Finally, the resulting finite–dimensional system of linear equations needs to be solved numerically. In the case of the two weakly coupled saddle point problems, the system matrix is a block matrix of the form

$$\mathbf{N}_\Lambda = \begin{pmatrix} \hat{\mathbf{E}}_\Lambda & \mathbf{L}_\Lambda^T \\ \mathbf{L}_\Lambda & \mathbf{E}_\Lambda \end{pmatrix}$$

where then the finite–dimensional system inherits the property to be well–posed in Euclidean metric, that is, the spectral condition numbers of $\mathbf{L}_\Lambda, \mathbf{N}_\Lambda$ are independent of the discretization indicated by Λ.

Since $\mathbf{L}_\Lambda$ only contains discretizations of local operators (the differential operator A and the trace mapping B) and because of the weak coupling of the two systems, $\mathbf{N}_\Lambda$ is 'sparsely' populated (in a sense to be made precise later). Thus, one would like to avoid computational and storage costs which would be needed when using a direct method like QR decomposition for the whole system or for each of the saddle point problems. In contrast, iterative methods take full advantage of the sparsity of the system and allow for realistic applications in several space variables. In addition, an iterative solution of the finite–dimensional discrete Euler equations is only required up to an *accuracy* provided by the *discretization.*

Here we propose a fully iterative method for the coupled saddle point problems which can be viewed as an *inexact gradient method.* It consists of an *outer iteration* which is a simple gradient method, and *inner iterations* in which alternatingly each of the saddle point problems are solved iteratively.

The investigation of fast iterative methods for a single saddle point problem has a long tradition, starting with the well–known Uzawa algorithm, see e.g. [Br]. Multigrid and eigenvalue shift approaches by means of a preconditioner for a finite discretization for A [BWY, BP] are for stable standard Galerkin discretizations together with wavelet–based methods [K1] among the most effective schemes for saddle point problems: They provide the solution up to discretization error accuracy with a number of iterations independent of the discretization.

Choosing as inner iteration the simplest version of Uzawa's method, we show here that such an outer–inner iteration scheme converges, provided that each of the saddle point problems is solved sufficiently well. Moreover, it is shown that this scheme is in combination with a *nested iteration strategy* asymptotically *optimal*: Starting from a coarsest discretization level, one sets up and solves repeatedly the linear system on successively finer grids, taking an interpolation of the solution on the current grid to the next finer grid as starting value. Since the system matrix on each level has a uniformly bounded condition number independent of the discretization, this whole strategy provides the solution on the finest discretization level J in a number of operations which is proportional to the number of unknowns N_J on level J.

It should be emphasized already at this point that the analysis for the inexact gradient scheme can readily be extended to more sophisticated gradient methods described in [Bs] or to inexact *conjugate* gradient schemes. Conjugate gradient methods with inexactness stemming from preconditioners has been investigated in [GoY]. We could also extend the method to more sophisticated iterative methods for solving saddle point problems. In dispensing with these generalizations here, we hope to better bring out the basic ideas and to avoid additional notation, in particular, in the complexity analysis of the algorithm.

In summary, the type of problems treated in this paper provides a promising potential for the use of *wavelets*, ranging from the evaluation of non–integer norms over a built–in efficient wavelet preconditioning to a systematic way to satisfy certain stability conditions (like the LBB condition for saddle point problems [DK]).

For general surveys on the application of wavelets to operator equations, we refer to [Co, D2, D3].

There are different points of contact of the work presented in this paper to previous investigations of optimal control problems. To mention a few, for shape optimization problems involving free boundaries, the combination with Lagrange multiplier techniques and fictitious domain methods is particularly promising, see e.g. [DH, Has, HHK, HasN, KP, NT, Toi]. Finite Element based least squares techniques for control problems involving Navier–Stokes equations have been discussed in [Bo]. In an abstract framework optimal control problems with elliptic systems as constraints are considered and discretized in terms of spectral elements in [AN]. An adaptive finite element Galerkin method for quadratic optimal control problems governed by elliptic boundary value problems is analyzed in [BKR]. Optimal control problems with parabolic PDEs as constraints, reducing the problem to Differential–Algebraic Equations by finite difference and finite element techniques, are discussed in [GJL, PRG]. The question of preconditioning saddle point problems in optimal control has been investigated in [GMPS]. Numerical optimization for control problems with semilinear elliptic equations subject to inequality constraints and discretized by finite differences can be found in [MM]. Newton–SQP methods for control problems with semilinear parabolic equations are discussed in [Tr].

The remainder of this paper is structured as follows.

In Section 2 we introduce a *general concept* how continuous problems formulated as isomorphisms on Hilbert spaces are discretized in terms of wavelets. We briefly recall the main features of biorthogonal wavelets that are needed here. The concept is then used first in Sections 3.1 and 3.2 to formulate the constraints in weak form (1.11) as an (infinite) ℓ_2–automorphism. Then in Section 3.3 we pose an equivalent class of control problems involving ℓ_2–norms and these constraints. We apply the known machinery for quadratic cost functionals with linear constraints to derive the optimality conditions which provide an (infinite) ℓ_2–automorphism. Section 3.4 briefly summarizes the conditions that ensure that corresponding finite–dimensional parts of the coupled system are stable. In Section 4.1 we describe an inexact gradient method for the numerical solution of the coupled system together with estimating the computational work in combination with a nested iteration strategy in Section 4.2. Finally, in Section 4.3 some numerical results are presented.

2 The General Concept

The systems of operator equations appearing in this context can be placed into the following abstract framework.

We consider closed subspaces $H_{i,0}$, $i = 1, \ldots, M$, of Hilbert spaces H_i whose inner products are denoted by $(\cdot,\cdot)_{H_i}$. The space $H'_{i,0}$ is the normed dual of $H_{i,0}$ endowed with the norm

$$\|v\|_{H'_{i,0}} := \sup_{0 \neq w \in H_{i,0}} \frac{\langle w, v\rangle}{\|w\|_{H_i}} \tag{2.1}$$

where $\langle\cdot,\cdot\rangle$ always denotes the (respective) dual form. (For clarity, e.g. in (1.10), the duality spaces are written explicitly.) The H_i will be Sobolev spaces on either the fictitious domain $\square$, or on the boundary manifold $\partial\Omega$ or smooth subsets Γ of $\partial\Omega$. Thus, we have

$$\text{either} \qquad H_i \subseteq L_2 \subseteq H'_{i,0} \qquad \text{or} \qquad H'_{i,0} \subseteq L_2 \subseteq H_i, \tag{2.2}$$

where L_2 stands for the space of Lebesgue integrable functions $L_2(\Box)$, or $L_2 = L_2(\partial\Omega)$, $L_2(\Gamma)$. Furthermore, the spaces $H_{i,0}$ will either coincide with H_i or will be determined by homogeneous boundary conditions.

Given $H_{i,0}$, $i = 1, \ldots, M$, we define the product space

$$\mathcal{H} := \prod_{i=1}^{M} H_{i,0}, \qquad \text{where} \quad (\cdot,\cdot)_{\mathcal{H}} := \sum_{i=1}^{M} (\cdot,\cdot)_{H_i} \tag{2.3}$$

is the canonical inner product on $\mathcal{H}$. Thus, the norms for $\mathcal{H}$ and $\mathcal{H}'$ can be defined as

$$\|V\|_{\mathcal{H}} := \left(\sum_{i=1}^{M} \|v_i\|_{H_i}^2 \right)^{1/2}, \qquad \|W\|_{\mathcal{H}'} = \sup_{\|V\|_{\mathcal{H}}=1} \langle V, W \rangle, \tag{2.4}$$

where $\langle V, W\rangle := \sum_{i=1}^{M} \langle v_i, w_i \rangle$.

Step 1 — Well–Posedness:
The starting point to apply wavelet concepts is that the continuous problem is well–posed in the following sense. It means to *identify* $\mathcal{H}$ and a linear operator $\mathcal{N}$ such that $\mathcal{N} : \mathcal{H} \to \mathcal{H}'$ is an *isomorphism*: Given $F := (f_1, \ldots, f_M)^T \in \mathcal{H}'$, find $U := (u_1, \ldots, u_M)^T \in \mathcal{H}$ such that

$$\mathcal{N} U = F, \tag{2.5}$$

where $\mathcal{N} = (\mathcal{N}_{i,j})_{i,j=1}^{M}$ and

$$\|\mathcal{N} V\|_{\mathcal{H}'} \sim \|V\|_{\mathcal{H}} \qquad \text{for all } V \in \mathcal{H}. \tag{2.6}$$

Specifically the latter relation always means that there exist constants $0 < c_{\mathcal{N}} \le C_{\mathcal{N}} < \infty$ such that

$$c_{\mathcal{N}} \|V\|_{\mathcal{H}} \le \|\mathcal{N} V\|_{\mathcal{H}'} \le C_{\mathcal{N}} \|V\|_{\mathcal{H}}, \quad V \in \mathcal{H}, \tag{2.7}$$

holds.

Step 2 — Discretization and Shifting:
Once Step 1 is established, wavelet concepts can be employed to transform (2.5) into an equivalent well-posed ℓ_2*–problem* [D2].

To this end, define $I\!I := I\!I_1 \times \cdots \times I\!I_M$ with infinite index sets $I\!I_i$. For each i, the elements λ of $I\!I_i$ consist of different types of indices such as the *level of resolution* (*refinement* or *discretization level*) denoted by $|\lambda|$ and the *spatial location*. What we call *wavelets* for $\mathcal{H} = H_{1,0} \times \cdots \times H_{M,0}$ is a (catenated) collection of functions

$$\mathbf{\Psi} := \{ {}^{1}\Psi, \ldots, {}^{M}\Psi \} \tag{2.8}$$

where for each $i = 1, \ldots, M$ the collection ${}^{i}\Psi$ defined by

$${}^{i}\Psi := \{ {}^{i}\psi_\lambda : \lambda \in I\!I_i \} \subset H_{i,0} \tag{2.9}$$

has the following properties:

(I) *Riesz basis property:* Every $v \in H_{i,0}$ can be uniquely expanded in terms of ${}^i\Psi$,

$$v = \mathbf{v}^T\, {}^i\Psi := \sum_{\lambda \in I\!I_i} v_\lambda\, {}^i\psi_\lambda, \qquad (2.10)$$

and its expansion coefficients satisfy the *norm equivalence*

$$\|v\|_{H_i} \;\sim\; \|{}^i\mathbf{D}\mathbf{v}\|_{\ell_2(I\!I_i)} \qquad (2.11)$$

where ${}^i\mathbf{D}$ is some diagonal matrix. This means that the *scaled* collection ${}^i\mathbf{D}^{-1}({}^i\Psi)$ constitutes a *Riesz basis* for $H_{i,0}$. We always use the convention that the basis is normalized in L_2, i.e., for $H_i = L_2$ it follows that ${}^i\mathbf{D} = \mathbf{I}$.

(II) *Locality:* The functions ${}^i\psi_\lambda$ are compactly supported. The widths of their support are decreasing with growing discretization level $|\lambda|$,

$$\operatorname{diam}(\operatorname{supp} {}^i\psi_\lambda) \;\sim\; 2^{-|\lambda|}. \qquad (2.12)$$

For the dual pairing $\langle\cdot,\cdot\rangle$ for $H_{i,0}$ and its dual $H'_{i,0}$ there exists by Riesz' representation theorem a collection

$${}^i\tilde{\Psi} := \{{}^i\tilde{\psi}_\lambda : \lambda \in I\!I_i\} \subset H'_{i,0} \qquad (2.13)$$

such that

$$\langle {}^i\psi_\lambda, {}^i\tilde{\psi}_\mu \rangle = \delta_{\lambda\mu}, \qquad \lambda, \mu \in I\!I_i, \qquad (2.14)$$

and ${}^i\mathbf{D}({}^i\tilde{\Psi})$ is a Riesz basis for $H'_{i,0}$. Here $\delta_{\lambda\mu}$ denotes the Kronecker delta. By a duality argument one concludes from (2.11) that the corresponding norm equivalence

$$\|\tilde{v}\|_{H'_{i,0}} \;\sim\; \|{}^i\mathbf{D}^{-1}\tilde{\mathbf{v}}\|_{\ell_2(I\!I_i)} \qquad (2.15)$$

holds for any $\tilde{v} = \tilde{\mathbf{v}}^T\, {}^i\tilde{\Psi} \in H'_{i,0}$ [D1]. The coefficients v_λ in the expansion (2.10) can then be expressed in terms of the dual basis as $v_\lambda = \langle v, {}^i\tilde{\psi}_\lambda\rangle$.

$({}^i\Psi, {}^i\tilde{\Psi})$ is called a pair of *biorthogonal wavelets*. Of particular interest are the cases when the dual wavelets ${}^i\tilde{\Psi}$ also have compact support (2.12).

It will be convenient to use the following shorthand notation. We will view ${}^i\Psi$ and $\mathbf{\Psi}$ as in (2.8), (2.9) as a *collection* of functions as well as a (possibly infinite) (column) *vector* containing all functions always assembled in some fixed unspecified order. For a countable collection of functions Θ and some single function σ, the quantities $\langle\Theta,\sigma\rangle$ and $\langle\sigma,\Theta\rangle$ are to be understood as the column, respectively row, vector with entries $\langle\theta,\sigma\rangle$, respectively $\langle\sigma,\theta\rangle$, $\theta \in \Theta$. For two collections Θ, Σ, the term $\langle\Theta,\Sigma\rangle$ is then a (possibly infinite) matrix with entries $(\langle\theta,\sigma\rangle)_{\theta\in\Theta,\ \sigma\in\Sigma}$ for which $\langle\Theta,\Sigma\rangle = \langle\Sigma,\Theta\rangle^T$. This also implies for a (possibly infinite) matrix $\mathbf{C}$ that $\langle\mathbf{C}\Theta,\Sigma\rangle = \mathbf{C}\langle\Theta,\Sigma\rangle$ and $\langle\Theta,\mathbf{C}\Sigma\rangle = \langle\Theta,\Sigma\rangle\mathbf{C}^T$.

Using this notation, the expansion coefficients in (2.10) and (2.15) can explicitly be expressed as

$$\mathbf{v}^T = \langle v, {}^i\tilde{\Psi}\rangle, \qquad \tilde{\mathbf{v}} = \langle {}^i\Psi, \tilde{v}\rangle. \qquad (2.16)$$

Furthermore, the *biorthogonality* or *duality conditions* (2.14) can be written in terms of an infinite matrix,

$$\langle {}^i\Psi, {}^i\tilde{\Psi}\rangle = \mathbf{I}, \qquad (2.17)$$

where $\mathbf{I}$ is the identity matrix.

Now we are ready to transform the continuous operator equation (2.5) into a discrete system of equations in terms of the wavelet basis $\mathbf{\Psi}$ for $\mathcal{H}$ as follows. Expansion of the solution

$$U = (u_1, \ldots, u_M)^T = \left(\mathbf{u}_1^T \left({}^1\mathbf{D}^{-1}\right) {}^1\Psi, \ldots, \mathbf{u}_M^T \left({}^M\mathbf{D}^{-1}\right) {}^M\Psi\right)^T =: \mathbf{U}^T\mathbf{D}^{-1}\mathbf{\Psi} \quad \in \mathcal{H} \tag{2.18}$$

and the right hand side $F = \langle \mathbf{\Psi}, F\rangle^T \tilde{\mathbf{\Psi}} \in \mathcal{H}'$ yields the system of equations

$$\langle \mathbf{\Psi}, \mathcal{N}\mathbf{\Psi}\rangle\, \mathbf{D}^{-1}\,\mathbf{U} = \langle \mathbf{\Psi}, F\rangle. \tag{2.19}$$

Multiplying (2.19) by the diagonal matrix $\mathbf{D}^{-1}$,

$$\mathbf{D}^{-1}\,\langle \mathbf{\Psi}, \mathcal{N}\mathbf{\Psi}\rangle\, \mathbf{D}^{-1}\,\mathbf{U} = \mathbf{D}^{-1}\,\langle \mathbf{\Psi}, F\rangle, \tag{2.20}$$

and recalling the norm equivalences (2.11) and (2.15), we infer

$$\mathbf{N}\,\mathbf{U} = \mathbf{F} \tag{2.21}$$

where $\mathbf{N} : \ell_2(I\!I) \to \ell_2(I\!I)$ is an automorphism. Here we have used the abbreviations

$$\mathbf{N} := \mathbf{D}^{-1}\,\langle \mathbf{\Psi}, \mathcal{N}\mathbf{\Psi}\rangle\, \mathbf{D}^{-1}, \quad \mathbf{F} := \mathbf{D}^{-1}\,\langle \mathbf{\Psi}, F\rangle. \tag{2.22}$$

In summary, the wavelet framework provides on account of the norm equivalences (2.11) and (2.15) a mechanism that allows to transform the original operator equation (2.5) into an equivalent infinite linear system of equations (2.21) which is well–posed in the Euclidean metric $\ell_2(I\!I)$. This implies also that the *spectral condition number* of $\mathbf{N}$ satisfies

$$\kappa(\mathbf{N}) := \|\mathbf{N}\|_{\ell_2}\, \|\mathbf{N}^{-1}\|_{\ell_2} \quad \lesssim \quad 1. \tag{2.23}$$

The constants on the right hand side are quotients of the constants in the isomorphism relation (2.6) and the norm equivalences (2.11) and (2.15), see e.g. [D2]. Therefore, the multiplication by $\mathbf{D}^{-1}$ in (2.20) can be viewed as *preconditioning* the operator $\langle \mathbf{\Psi}, \mathcal{N}\mathbf{\Psi}\rangle$ in (2.19).

Remark 2.1 *Below the representer of the optimal control problem (1.13) with constraints in weak form (1.11) will be formulated in terms of ℓ_2–norms with a saddle point problem of type (2.21) as constraints.*

In directly formulating the minimization functional in this way, we can dispense with certain constants that would otherwise be taken into account from norm equivalences like (2.7).

Step 3 — Stability of the Finite–Dimensional Discrete Systems:
The last cornerstone of the strategy is to establish conditions that guarantee that in the finite–dimensional case the discretizations are *stable*, meaning that finite analogs of

(2.21) are uniformly stable invertible *independent* of the discretization. Depending on the situation, there are different ways to ensure this.

Let $\Lambda \subset I\!I$ be the product of *finite* index sets Λ_i, $i = 1, \dots, M$, (whose specific choice will be discussed later) and define

$$\mathbf{\Psi}_\Lambda := \{\mathbf{\Psi}_\lambda : \lambda \in \Lambda\} \tag{2.24}$$

and correspondingly $\tilde{\mathbf{\Psi}}_\Lambda$. Expanding $U_\Lambda = \mathbf{U}_\Lambda^T \mathbf{D}_\Lambda^{-1} \mathbf{\Psi}_\Lambda \in S(\mathbf{\Psi}_\Lambda) := \mathrm{span}\{\mathbf{\Psi}_\Lambda\}$, the finite–dimensional analog of (2.21) is

$$\mathbf{N}_\Lambda \, \mathbf{U}_\Lambda := \mathbf{D}_\Lambda^{-1} \langle \mathbf{\Psi}_\Lambda, \mathcal{N} \, \mathbf{\Psi}_\Lambda \rangle \, \mathbf{D}_\Lambda^{-1} \, \mathbf{U}_\Lambda \;=\; \mathbf{D}_\Lambda^{-1} \langle \mathbf{\Psi}_\Lambda, F \rangle =: \mathbf{F}_\Lambda \tag{2.25}$$

where $\mathbf{N}_\Lambda : \ell_2(\Lambda) \to \ell_2(\Lambda)$.

The core of the problem arising here is to ensure that $\mathbf{N}_\Lambda$ is still an automorphism with constants *independent* of Λ,

$$\|\mathbf{N}_\Lambda \boldsymbol{V}_\Lambda\|_{\ell_2} \sim \|\boldsymbol{V}_\Lambda\|_{\ell_2} \qquad \text{for all } \boldsymbol{V} \in \ell_2(\Lambda). \tag{2.26}$$

Depending on the problem formulation (2.5), this in turn may restrict the choice of the finite–dimensional subspaces $S(\mathbf{\Psi}_\Lambda) \subset \mathcal{H}$ or, equialently, the finite index sets Λ_i, $i = 1, \dots, M$.

The discrete stability on the finite–dimensional subspace (2.26) not only ensures that the problem (2.25) is invertible uniformly independent of Λ. In view of (2.23), it also entails that the spectral condition number of $\mathbf{N}_\Lambda$ is *bounded uniformly* in Λ,

$$\kappa(\mathbf{N}_\Lambda) = \|\mathbf{N}_\Lambda\|_2 \, \|\mathbf{N}_\Lambda^{-1}\|_2 \;\; \lesssim \;\; 1 \tag{2.27}$$

with constants consisting of the quotients of the constants from (2.6) and again the norm equivalences (2.11) and (2.15) but which do not depend on Λ.

In the simplest case of a symmetric elliptic operator $\mathcal{N}$, one can choose *any* such finite sets. This is usually denoted as *Galerkin stability*. For the saddle point problem (1.11), this is, however, not the case. Here one has to require additional conditions on the discretizations, known as the *Ladyšenskaja–Babuška–Brezzi (LBB) condition*, see Section 3.4 below.

Remark 2.2 *Natural least squares formulations of (2.5) are an alternative to this situation which tends to become complicated when $M > 2$, i.e., when more than two approximation spaces need to be coupled. For an* arbitrary *system of the form (2.5) with (2.6), this has been discussed in [DKS]. The least square schemes can be interpreted as a Galerkin scheme for a* different *system of operator equations, leading to a* symmetric positive definite *system of linear equations.*

Once Step 3 is completed, it remains to solve

$$\mathbf{N}_\Lambda \, \mathbf{U}_\Lambda = \mathbf{F}_\Lambda \tag{2.28}$$

numerically. For symmetric positive definite operators one can apply a conjugate gradient method which on account of (2.27) converges with a *constant* convergence speed independent of Λ. For the saddle point systems derived from (1.11), one can apply an Uzawa type algorithm which inherits this property with a proper scaling, see Section 4.2.

The general concept has been used already for different situations in [CDD1, DK, DKS, K2].

3 Optimal Control Problems in ℓ_2

In view of the optimal control problem (1.13) with constraints (1.11), the function spaces in which the continuous problem is posed are $H^1(\Box)$, $(H^{1/2}(\Gamma))'$ and $Z \in \{H^1(\Box), H^{1/2}(\Gamma_y)\}$. For such function spaces and spaces on different domains or manifolds, including L–shaped domains and their boundaries, there are by now a number of constructions of wavelets available satisfying properties (I), (II), see e.g. [D2, K2]. Starting from biorthogonal wavelets on the interval [DKU1], one can construct corresponding wavelets on the domain or its boundary that satisfy the norm equivalences (2.11), (2.15). Depending on the smoothness of $\mathcal{H}$, one can either use domain decomposition ideas [CTU, DS1] for smoothness order $|s| < 1/2$ or more sophisticated function space characterizations for higher order smoothness [DS2]. Usually the set of *primal* wavelets $\mathbf{\Psi}$ consisting in these constructions of *piecewise polynomials* will be the ones one computes with while the *dual* ones $\tilde{\mathbf{\Psi}}$ are mainly needed for analysis purposes.

Next we will recall some theoretical facts about saddle point problems in abstract form from [BF] in order to apply the first two parts of the concept from Section 2 and derive (2.21) for the constraints in weak form (1.11).

3.1 An Abstract Saddle Point Problem

Let Y and Q be Hilbert spaces with their topological duals Y', Q' and dual forms $\langle\cdot,\cdot\rangle_{Y\times Y'}$, $\langle\cdot,\cdot\rangle_{Q\times Q'}$, respectively, which we often abbreviate by $\langle\cdot,\cdot\rangle$. We denote the norms on Y and Q and the induced inner products by $(\cdot,\cdot)_Y = \|\cdot\|_Y^2$ and $(\cdot,\cdot)_Q = \|\cdot\|_Q^2$. . Let $A : Y \to Y'$ and $B : Y \to Q'$ (with adjoint $B' : Q \to Y'$) be linear continuous operators. In view of the case considered above in Section 1.1 where B is a trace operator, it will suffice here to restrict ourselves to the case of B being *surjective*, i.e., $\operatorname{range} B = Q'$ and $\ker B' = \{0\}$.

We consider the following saddle point problem in abstract form: Given $(f, u) \in Y' \times Q'$, find $(y, p) \in Y \times Q$ such that

$$\begin{pmatrix} A & B' \\ B & 0 \end{pmatrix} \begin{pmatrix} y \\ p \end{pmatrix} = \begin{pmatrix} f \\ u \end{pmatrix} \tag{3.1}$$

has a unique solution. Concerning the existence and uniqueness of solutions of such problems, one has the following result, see e.g. [BF, GR].

Theorem 3.1 *Let the linear operator A be invertible on* $\ker B \subseteq Y$*, i.e., for some constant $\alpha_1 > 0$*

$$\inf_{v\in\ker B}\sup_{w\in\ker B}\frac{\langle Av,w\rangle_{Y'\times Y}}{\|v\|_Y\|w\|_Y}\geq\alpha_1,\qquad \inf_{v\in\ker B}\sup_{w\in\ker B}\frac{\langle A'v,w\rangle_{Y'\times Y}}{\|v\|_Y\|w\|_Y}\geq\alpha_1, \tag{3.2}$$

and let for some constant $\beta_1 > 0$ the inf–sup condition

$$\sup_{v\in Y}\frac{\langle Bv,q\rangle_{Q'\times Q}}{\|v\|_Y}\geq\beta_1\|q\|_Q,\quad q\in Q, \tag{3.3}$$

hold. Then there exists a unique solution $(y,p)\in Y\times Q$ to problem (3.1) for given $f\in Y'$ and $u\in Q'$, i.e.,

$$\mathcal{L} := \begin{pmatrix} A & B' \\ B & 0 \end{pmatrix} \text{ is an isomorphism } Y\times Q\to Y'\times Q', \tag{3.4}$$

and one has the equivalence

$$c_{\mathcal{L}}\left\|\begin{pmatrix} v\\ q\end{pmatrix}\right\|_{Y\times Q}\ \leq\ \left\|\mathcal{L}\begin{pmatrix} v\\ q\end{pmatrix}\right\|_{Y'\times Q'}\ \leq\ C_{\mathcal{L}}\left\|\begin{pmatrix} v\\ q\end{pmatrix}\right\|_{Y\times Q} \tag{3.5}$$

for any $(v,q)\in Y\times Q$, where the constants $c_{\mathcal{L}}, C_{\mathcal{L}}$ are composed of α_1,β_1 and the continuity constants for A and B.

3.2 The Saddle Point Problem (1.11) and Trace Theorems

In view of (1.11) where the bilinear form $a(\cdot,\cdot)$ defined by (1.7) is usually continuous and satisfies (2.12), the relevant condition that needs to be ensured is the *inf-sup condition* (3.3) for B for the situation involving the fictitious domain (1.11). Here we have

$$Y = H^1(\Box)\quad\text{and}\quad Q = (H^{1/2}(\Gamma))'. \tag{3.6}$$

The classical trace theorem from e.g. [Gr] for domains Ω with Lipschitz continuous boundary $\partial\Omega$ is extended in [K2] with the aid of Whitney extension estimates to the following case.

Theorem 3.2 *For any $f\in H^s(\Box)$, $1/2 < s < 3/2$, one can estimate*

$$\|Bf\|_{H^{s-1/2}(\Gamma)}\ \leq c_{T_1}\ \|f\|_{H^s(\Box)}. \tag{3.7}$$

Conversely, for every $h\in H^{s-1/2}(\Gamma)$, there exists some $f\in H^s(\Box)$ such that $Bf = h$ and

$$\|f\|_{H^s(\Box)}\ \leq\ c_{T_2}\|h\|_{H^{s-1/2}(\Gamma)}. \tag{3.8}$$

From this, the *inf–sup condition* (3.3) for B can be derived.

Remark 3.3 *The* inf-sup condition *with respect to $H^1(\Box)\times(H^{1/2}(\Gamma))'$ holds,*

$$\inf_{q\in(H^{1/2}(\Gamma))'}\sup_{v\in H^1(\Box)}\frac{\langle Bv,q\rangle_{H^{1/2}(\Gamma)\times(H^{1/2}(\Gamma))'}}{\|v\|_{H^1(\Box)}\,\|q\|_{(H^{1/2}(\Gamma))'}}\geq\frac{1}{c_{T_2}}. \tag{3.9}$$

In view of the minimization functional (1.13) in the reference optimal control problem, in addition to Y and Q, in (1.12) a third Hilbert space has been introduced, the *observation spae*, whose norm stands for the energy norm in which the state is measured. At this point it is useful for later purpose to note also that there is a linear operator called the *observation operator*

$$T : Y \to Z$$

whose specific form depends on the choice of Z.

Remark 3.4 *(i) If $Z = Y$, then T is just the identity. This would mean to measure the state in the $H^1(\Box)$–norm.*

(ii) When $Y = H^1(\Box)$ and when measurements controlling the state can only be taken on (part of) the boundary $\partial\Omega$ denoted by Γ_y, Z is the trace space $Z = H^{1/2}(\Gamma_y)$, and $T : H^1(\Box) \to H^{1/2}(\Gamma_y)$ is the standard trace operator with respect to Γ_y. Recall that then also the Trace Theorem 3.2 applies: for any $f \in H^1(\Box)$, one can estimate

$$\|Tf\|_{H^{1/2}(\Gamma_y)} \le c_{T_1,y}\ \|f\|_{H^1(\Box)}. \tag{3.10}$$

Conversely, for every $h \in H^{1/2}(\Gamma_y)$, there exists some $f \in H^1(\Box)$ such that $Tf = h$ and

$$\|f\|_{H^1(\Box)} \ \le\ c_{T_2,y}\|h\|_{H^{1/2}(\Gamma_y)}. \tag{3.11}$$

Note that the nature of the space Z and of the operator T is in either case like that of the spaces Y, Q and the operators A, B.

We have asserted already that there is a pair of biorthogonal wavelet bases each for both Y and Q so that corresponding versions of the norm equivalences (2.11) and (2.15) hold,

$$\|v\|_Y \ \sim \|\mathbf{D}_Y\mathbf{v}\|_{\ell_2(I\!I_Y)}, \qquad \|\tilde{v}\|_{Y'} \sim \|\mathbf{D}_Y^{-1}\tilde{\mathbf{v}}\|_{\ell_2(I\!I_Y)}, \tag{3.12}$$

$$\|q\|_Q \ \sim \|\mathbf{D}_Q\mathbf{q}\|_{\ell_2(I\!I_Q)}, \qquad \|\tilde{q}\|_{Q'} \sim \|\mathbf{D}_Q^{-1}\tilde{\mathbf{q}}\|_{\ell_2(I\!I_Q)}. \tag{3.13}$$

All quantities referring to Y or Q will be indexed accordingly by subscript $()_Y$ or $()_Q$, respectively, if distinction is necessary.

Now we can discretize and precondition (1.11) in the sense of Step 2 as follows. Expanding (y, p) in terms of the weighted wavelet bases,

$$U := (y,p)^T = \left(\mathbf{y}^T\mathbf{D}_Y^{-1}\Psi_Y,\ \mathbf{p}^T\mathbf{D}_Q^{-1}\Psi_Q\right)^T, =: \mathbf{U}^T\mathbf{D}^{-1}\mathbf{\Psi}, \tag{3.14}$$

and also the right hand side $F = (f, u)^T = \langle F, \mathbf{\Psi}\rangle\tilde{\mathbf{\Psi}} \in \mathcal{H}'$ yields after multiplying by $\mathbf{D}^{-1} = \mathrm{diag}(\mathbf{D}_Y^{-1}, \mathbf{D}_Q^{-1})$ as a special case of (2.21) the system of equations

$$\mathbf{L}\begin{pmatrix}\mathbf{y}\\ \mathbf{p}\end{pmatrix} := \begin{pmatrix}\mathbf{A} & \mathbf{B}^T\\ \mathbf{B} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{y}\\ \mathbf{p}\end{pmatrix} = \begin{pmatrix}\boldsymbol{f}\\ \mathbf{u}\end{pmatrix} \tag{3.15}$$

where we have used the abbreviations

$$\begin{array}{llll}\mathbf{A} & := \mathbf{D}_Y^{-1}\,\langle \Psi_Y, A\Psi_Y\rangle\,\mathbf{D}_Y^{-1}, & \boldsymbol{f} & := \mathbf{D}_Y^{-1}\,\langle \Psi_Y, f\rangle, \\ \mathbf{B} & := \mathbf{D}_Q^{-1}\,\langle \Psi_Q, B\Psi_Y\rangle\,\mathbf{D}_Y^{-1}, & \mathbf{u} & := \mathbf{D}_Q^{-1}\,\langle \Psi_Q, u\rangle.\end{array} \tag{3.16}$$

Thus, covered by the general concept from Section 2, we have in view of Remark 3.3 and Theorem 3.1 the following result.

Corollary 3.5 *The operator* $\mathbf{L}$ *defined in (3.15) is an* ℓ_2*–automorphism, i.e., for every* $(\mathbf{v},\mathbf{q}) \in \ell_2(\mathit{I\!I}) = \ell_2(\mathit{I\!I}_Y \times \mathit{I\!I}_Q)$ *one has*

$$\mathbf{c_L} \left\| \begin{pmatrix}\mathbf{v}\\ \mathbf{q}\end{pmatrix} \right\|_{\ell_2} \;\leq\; \left\| \mathbf{L}\begin{pmatrix}\mathbf{v}\\ \mathbf{q}\end{pmatrix} \right\|_{\ell_2} \;\leq\; \mathbf{C_L} \left\| \begin{pmatrix}\mathbf{v}\\ \mathbf{q}\end{pmatrix} \right\|_{\ell_2} \tag{3.17}$$

with constants $\mathbf{c_L}, \mathbf{C_L}$ *only depending on* $c_{\mathcal{L}}, C_{\mathcal{L}}$ *from (3.5) and the constants in the norm equivalences (2.11) and (2.15).*

3.3 A Representer

We have now prepared the ground for formulating an appropriate representer of the control problem (1.13). In the case of Remark 3.4(i) let $\mathbf{T} = \mathbf{I}$. In the other case (ii), denote by $\mathbf{T}$ the discrete ℓ_2–mapping representing the (properly normalized) observation operator T,

$$\mathbf{T} := \mathbf{D}_Z^{-1}\,\langle \Psi_Z, T\Psi_Y\rangle\,\mathbf{D}_Y^{-1}. \tag{3.18}$$

Although often traces of functions to the boundary are not explicitly written, we will include $\mathbf{T}$ into the functional in order to keep track of it.

The representer of the minimization functional (1.13) is now defined as

$$\widetilde{\mathbf{J}}(\mathbf{y},\mathbf{p},\mathbf{u}) \;:=\; \tfrac{\omega}{2}\|\mathbf{T}\mathbf{y} - \mathbf{y}_\Box\|_{\ell_2}^2 + \tfrac{1}{2}\|\mathbf{u}\|_{\ell_2}^2, \tag{3.19}$$

where $\mathbf{y}_\Box$ descends from expanding $y_\Box$ in terms of the (properly normalized) Ψ_Z, $y_\Box = \mathbf{y}_\Box^T \mathbf{D}_Z^{-1}\Psi_Z$.

Thus, from the previous discussion the most convenient form of the minimization problem that still captures the physical features of the quadratic cost functional, namely, to measure effects in natural norms emerges as the following *discrete* ℓ_2*–optimal control problem.*

Control Problem (P1): Find $(\mathbf{y},\mathbf{p},\mathbf{u}) \in \ell_2(\mathit{I\!I}_Y \times \mathit{I\!I}_Q \times \mathit{I\!I}_Q)$ such that $\widetilde{\mathbf{J}}(\mathbf{y},\mathbf{p},\mathbf{u})$ defined in (3.19) is minimized subject to (3.15).

Next we follow standard techniques from e.g. [Li, Z] to derive existence and uniqueness results for (P1). To this end, note that with defining

$$\mathbf{R} := \begin{pmatrix}\mathbf{0}\\ \mathbf{I}\end{pmatrix}, \quad \mathbf{G} := \begin{pmatrix}\boldsymbol{f}\\ \mathbf{0}\end{pmatrix}, \quad \mathbf{S} := (\mathbf{T}\;\;\mathbf{0}), \quad \mathbf{W} := \begin{pmatrix}\mathbf{y}\\ \mathbf{p}\end{pmatrix}, \quad \mathbf{W}_\Box := \begin{pmatrix}\mathbf{y}_\Box\\ \mathbf{0}\end{pmatrix}, \tag{3.20}$$

problem (P1) can be reformulated as follows.

Control Problem (P2): Find $(\mathbf{W}, \mathbf{u}) \in \ell_2$ which minimize the cost functional

$$\widetilde{\mathbf{J}}(\mathbf{W}, \mathbf{u}) = \tfrac{\omega}{2}\|\mathbf{S}\mathbf{W} - \mathbf{W}_\square\|_{\ell_2}^2 + \tfrac{1}{2}\|\mathbf{u}\|_{\ell_2}^2 \tag{3.21}$$

$$\text{subject to} \qquad \mathbf{L}\mathbf{W} = \mathbf{R}\mathbf{u} + \mathbf{G}. \tag{3.22}$$

Since $\mathbf{L}$ is by Corollary 3.5 invertible, we can eliminate $\mathbf{W}$ from (3.22),

$$\mathbf{W} = \mathbf{L}^{-1}\mathbf{R}\mathbf{u} + \mathbf{L}^{-1}\mathbf{G}, \tag{3.23}$$

and insert it into (3.21) to obtain a quadratic functional in terms of the control only,

$$\mathbf{J}(\mathbf{u}) := \tfrac{\omega}{2}\|\mathbf{S}\mathbf{L}^{-1}\mathbf{R}\mathbf{u} + \mathbf{S}\mathbf{L}^{-1}\mathbf{G} - \mathbf{W}_\square\|_{\ell_2}^2 + \tfrac{1}{2}\|\mathbf{u}\|_{\ell_2}^2. \tag{3.24}$$

Denoting by $D^s\mathbf{J}(\mathbf{u}; \mathbf{v}_1, \ldots, \mathbf{v}_s)$ the s-th variation of $\mathbf{J}$ at $\mathbf{u}$ in directions $\mathbf{v}_1, \ldots, \mathbf{v}_s$, where in particular

$$D\mathbf{J}(\mathbf{u}; \mathbf{v}) = \langle \delta\mathbf{J}(\mathbf{u}), \mathbf{v}\rangle = \lim_{t\to 0} \frac{\mathbf{J}(\mathbf{u} + t\mathbf{v}) - \mathbf{J}(\mathbf{u})}{t}, \tag{3.25}$$

we can now collect a number of properties of $\mathbf{J}$ for later purposes.

Proposition 3.6 *The functional* $\mathbf{J}$ *defined in (3.24) is twice differentiable on* ℓ_2 *with derivative*

$$\begin{aligned} D\mathbf{J}(\mathbf{u}; \mathbf{v}) &= \omega\langle \mathbf{S}\mathbf{W}(\mathbf{u}) - \mathbf{W}_\square,\, \mathbf{S}\mathbf{L}^{-1}\mathbf{R}\mathbf{v}\rangle + \langle \mathbf{u}, \mathbf{v}\rangle \\ &= \omega\langle \mathbf{S}\mathbf{L}^{-1}\mathbf{R}\mathbf{u} + \mathbf{S}\mathbf{L}^{-1}\mathbf{G} - \mathbf{W}_\square,\, \mathbf{S}\mathbf{L}^{-1}\mathbf{R}\mathbf{v}\rangle + \langle \mathbf{u}, \mathbf{v}\rangle \end{aligned} \tag{3.26}$$

for all $\mathbf{v} \in \ell_2$. *We infer explicitly*

$$\delta\mathbf{J}(\mathbf{u}) = \omega\mathbf{R}^T\mathbf{L}^{-T}\mathbf{S}^T(\mathbf{S}\mathbf{W}(\mathbf{u}) - \mathbf{W}_\square) + \mathbf{u} =: \mathbf{Q}\mathbf{u} + \mathbf{e}. \tag{3.27}$$

The second derivative of $\mathbf{J}$ *is for all* $\mathbf{v}, \mathbf{w} \in \ell_2$

$$D^2\mathbf{J}(\mathbf{u}; \mathbf{v}, \mathbf{w}) = \omega\langle \mathbf{S}\mathbf{L}^{-1}\mathbf{R}\mathbf{v},\, \mathbf{S}\mathbf{L}^{-1}\mathbf{R}\mathbf{w}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle \tag{3.28}$$

or equivalently

$$D^2\mathbf{J}(\mathbf{u}; \mathbf{v}, \mathbf{w}) = \omega\,\langle \mathbf{S}\tilde{\mathbf{W}}, \mathbf{S}\bar{\mathbf{W}}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle, \tag{3.29}$$

where $\tilde{\mathbf{W}} = \binom{\tilde{\mathbf{y}}}{\tilde{\mathbf{p}}}$ *and* $\bar{\mathbf{W}} = \binom{\bar{\mathbf{y}}}{\bar{\mathbf{p}}}$ *solve*

$$\mathbf{L}\,\tilde{\mathbf{W}} = \begin{pmatrix}\mathbf{0}\\ \mathbf{v}\end{pmatrix} \quad \text{and} \quad \mathbf{L}\,\bar{\mathbf{W}} = \begin{pmatrix}\mathbf{0}\\ \mathbf{w}\end{pmatrix}, \tag{3.30}$$

respectively. Moreover, $D^2\mathbf{J}$ *satisfies for all* $\mathbf{v}, \mathbf{w} \in \ell_2$ *the estimates*

$$D^2\mathbf{J}(\mathbf{u}; \mathbf{v}, \mathbf{w}) \le C_*\,\|\mathbf{v}\|_{\ell_2}\|\mathbf{w}\|_{\ell_2} \tag{3.31}$$

and

$$D^2\mathbf{J}(\mathbf{u}; \mathbf{v}, \mathbf{v}) \ge c_*\,\|\mathbf{v}\|_{\ell_2}^2 \tag{3.32}$$

with constants

$$C_* := \omega(\mathbf{c}_{T_1,y}\,\mathbf{c}_{\mathbf{L}}^{-1})^2 + 1 \quad \text{and} \quad c_* := 1. \tag{3.33}$$

Thus, $\mathbf{J}$ *is strictly convex on* ℓ_2, *implying that* $\mathbf{Q}$ *in (3.27) is symmetric positive definite.*

Proof:
The standard derivation of the assertions is included here for convenience and for fixing the particular terms.

For any $\mathbf{u}, \mathbf{v} \in \ell_2$ and $t > 0$, one has for $\mathbf{J}$ by definition (3.25)

$$\frac{\mathbf{J}(\mathbf{u}+t\mathbf{v}) - \mathbf{J}(\mathbf{u})}{t} = \tfrac{\omega}{2}\left(2\langle \mathbf{SW}(\mathbf{u}) - \mathbf{W}_\Box,\, \mathbf{SL}^{-1}\mathbf{Rv}\rangle + t\,\|\mathbf{SL}^{-1}\mathbf{Rv}\|_{\ell_2}^2\right) + \tfrac{1}{2}\left(2\langle \mathbf{u},\mathbf{v}\rangle + t\|\mathbf{v}\|_{\ell_2}^2\right)$$

yielding

$$D\mathbf{J}(\mathbf{u};\mathbf{v}) = \lim_{t\to 0}\frac{\mathbf{J}(\mathbf{u}+t\mathbf{v}) - \mathbf{J}(\mathbf{u})}{t} = \omega(\mathbf{SW}(\mathbf{u}) - \mathbf{W}_\Box,\, \mathbf{SL}^{-1}\mathbf{Rv}\rangle + \langle \mathbf{u},\mathbf{v}\rangle$$

and thus (3.26) upon inserting (3.23). Furthermore, let $\hat{\mathbf{G}}(\mathbf{u};\mathbf{v}) := \langle \delta\mathbf{J}(\mathbf{u}), \mathbf{v}\rangle$. Then

$$\frac{\hat{\mathbf{G}}(\mathbf{u}+t\mathbf{w};\mathbf{v}) - \hat{\mathbf{G}}(\mathbf{u};\mathbf{v})}{t} = \omega\,\langle \mathbf{SL}^{-1}\mathbf{Rv},\, \mathbf{SL}^{-1}\mathbf{Rw}\rangle + \langle \mathbf{v},\mathbf{w}\rangle$$

is independent of $\mathbf{u}$, yielding the identity (3.28) for all $\mathbf{v},\mathbf{w} \in \ell_2$. In view of (3.30), this is the same as (3.29) and can further be estimated by

$$D^2\mathbf{J}(\mathbf{u};\mathbf{v},\mathbf{w}) \le \omega\|\mathbf{T}\tilde{\mathbf{y}}\|_{\ell_2}\|\mathbf{T}\bar{\mathbf{y}}\|_{\ell_2} + \|\mathbf{v}\|_{\ell_2}\|\mathbf{w}\|_{\ell_2}. \tag{3.34}$$

In the case of Remark 3.4(i), $\|\mathbf{T}\tilde{\mathbf{y}}\|_{\ell_2} = \|\tilde{\mathbf{y}}\|_{\ell_2}$. In the other case (ii), one has by (3.10)

$$\|\mathbf{T}\tilde{\mathbf{y}}\|_{\ell_2} \le \mathbf{c}_{T_1,y}\|\tilde{\mathbf{y}}\|_{\ell_2} \tag{3.35}$$

with some constant $\mathbf{c}_{T_1,y}$. Furthermore, the isomorphism relation (3.17) yields

$$\|\tilde{\mathbf{y}}\|_{\ell_2} \le \mathbf{c}_{\mathbf{L}}^{-1}\|\mathbf{v}\|_{\ell_2} \tag{3.36}$$

for the solution $\tilde{\mathbf{y}}$ of (3.30), and correspondingly for $\bar{\mathbf{y}}$. Thus, one obtains from (3.34) and (3.35) followed by (3.36) the estimate (3.31) with C_* defined in (3.33),

$$D^2\mathbf{J}(\mathbf{u};\mathbf{v},\mathbf{w}) \le \left(\omega(\mathbf{c}_{T_1,y}\,\mathbf{c}_{\mathbf{L}}^{-1})^2 + 1\right)\|\mathbf{v}\|_{\ell_2}\|\mathbf{w}\|_{\ell_2} = C_*\|\mathbf{v}\|_{\ell_2}\|\mathbf{w}\|_{\ell_2}.$$

As for the lower estimate, one trivially has

$$D^2\mathbf{J}(\mathbf{u};\mathbf{v},\mathbf{v}) \ge \omega\|\mathbf{T}\tilde{\mathbf{y}}\|_{\ell_2}^2 + \|\mathbf{v}\|_{\ell_2}^2 \ \ge\ \|\mathbf{v}\|_{\ell_2}^2, \tag{3.37}$$

i.e., (3.32) is satisfied with $c_* = 1$. ∎

The following generalized Weierstrass theorem is a special case of Theorem 43.D from [Z] for the situation at hand. It gives necessary and sufficient conditions in terms of derivatives of the *Lagrangian functional*

$$\mathbf{Lagr}(\hat{\mathbf{y}},\hat{\mathbf{p}},\hat{\mathbf{u}},\hat{\mathbf{z}},\hat{\boldsymbol{\mu}}) := \mathbf{J}(\hat{\mathbf{u}}) + \left\langle (\hat{\mathbf{z}},\hat{\boldsymbol{\mu}}),\, \mathbf{L}\begin{pmatrix}\hat{\mathbf{y}}\\ \hat{\mathbf{p}}\end{pmatrix} - \begin{pmatrix}\boldsymbol{f}\\ \hat{\mathbf{u}}\end{pmatrix}\right\rangle \tag{3.38}$$

which is formed as usual by appending the conditions (3.15) by means of additional Lagrange multipliers $(\hat{\mathbf{z}},\hat{\boldsymbol{\mu}}) \in \ell_2$ to the minimization functional (3.24). Here $\hat{\mathbf{W}} = \binom{\hat{\mathbf{y}}}{\hat{\mathbf{p}}}$ is written explicitly to bring out the roles of $\hat{\mathbf{y}},\hat{\mathbf{p}}$ again.

Theorem 3.7 *Let* $\mathbf{L}$ *be the* ℓ_2*–automorphism from Corollary 3.5 and let* $\mathbf{J}$ *be the quadratic functional defined in (3.24). Then the unique solution* $(\mathbf{y}, \mathbf{p}, \mathbf{u})$ *of* **Problem (P2)** *is determined by the necessary conditions*

$$\delta\mathbf{Lagr}(\mathbf{y}, \mathbf{p}, \mathbf{u}, \mathbf{z}, \boldsymbol{\mu}; \boldsymbol{V}) = 0 \qquad \text{for } \boldsymbol{V} = \mathbf{z}, \boldsymbol{\mu}, \mathbf{u}, \mathbf{y}, \mathbf{p}. \tag{3.39}$$

Explicitly the Euler equations are for $\boldsymbol{V} = \mathbf{z}, \boldsymbol{\mu}, \mathbf{u}, \mathbf{y}, \mathbf{p}$ *given by (3.15) and additional conditions for* $\mathbf{u}$ *and* $\mathbf{z}, \boldsymbol{\mu}$*,*

$$\begin{aligned} \mathbf{L}\begin{pmatrix}\mathbf{y}\\ \mathbf{p}\end{pmatrix} &= \begin{pmatrix}\boldsymbol{f}\\ \mathbf{u}\end{pmatrix} \\ \mathbf{u} &= \boldsymbol{\mu} \qquad (3.40) \\ \mathbf{L}^T\begin{pmatrix}\mathbf{z}\\ \boldsymbol{\mu}\end{pmatrix} &= \begin{pmatrix}-\omega\mathbf{T}^T(\mathbf{T}\mathbf{y} - \mathbf{y}_\Box)\\ \mathbf{0}\end{pmatrix}. \qquad (3.41) \end{aligned}$$

Remark 3.8 *In the following, we call the constraints (3.15) or (3.22)* primal system *while the system (3.41) will be called* dual system*. The specific form of the dual system involving the* adjoint *of* $\mathbf{L}$ *(without further Riesz mappings as in [K2]) emerges from the particular formulation of the minimization functional (3.19).*

In the present framework, control problems with distributed *control would be formulated in terms of a cost functional of the form (3.19) with control* $\boldsymbol{f}$ *and with constraints (3.41). The corresponding Euler Equations would be similar to (3.15) with (3.40), (3.41). In this sense, problems with boundary and distributed control are equivalent: they both lead to mutually adjoint problems.*

The trivial equation (3.40) will be used later in Section 4 to construct an inexact gradient method involving the iterative solution of (3.15) and (3.41). To this end, it is useful to rewrite $\delta\mathbf{J}$ in (3.27) yet in another way.

Proposition 3.9 *In view of (3.15), (3.40) and (3.41), the first variation of* $\mathbf{J}$ *has the representation*

$$\delta\mathbf{J}(\mathbf{u}) \;=\; \mathbf{u} - \boldsymbol{\mu}, \tag{3.42}$$

that is, the evaluation of $\delta\mathbf{J}(\mathbf{u})$ *is equivalent to solving first (3.22) and then (3.41).*

Proof:
Recall from (3.27) that

$$\delta\mathbf{J}(\mathbf{u}) \;=\; \omega\mathbf{R}^T\mathbf{L}^{-T}\mathbf{S}^T(\mathbf{S}\mathbf{W}(\mathbf{u}) - \mathbf{W}_\Box) + \mathbf{u}$$

where $\mathbf{W}(\mathbf{u})$ is the solution of (3.22). Writing the solution $\mathbf{X} = \mathbf{X}(\mathbf{u}) = (\mathbf{z}, \boldsymbol{\mu})^T$ of (3.41) explicitly,

$$\mathbf{X} = -\omega\mathbf{L}^{-T}\begin{pmatrix}\mathbf{T}^T(\mathbf{T}\,\mathbf{y}(\mathbf{u}) - \mathbf{y}_\Box)\\ \mathbf{0}\end{pmatrix},$$

and applying $\mathbf{R}^T$ from the left we infer

$$\boldsymbol{\mu} = -\omega\mathbf{R}^T\,\mathbf{L}^{-T}\begin{pmatrix}\mathbf{T}^T(\mathbf{T}\,\mathbf{y}(\mathbf{u}) - \mathbf{y}_\Box)\\ \mathbf{0}\end{pmatrix}$$

which in view of (3.27) confirms (3.42). ∎

We summarize the results obtained in view of the general concept from Section 2. Eliminating $\boldsymbol{\mu}$ from (3.40), we can write (3.15) together with (3.41) as a *weakly coupled* system of saddle point problems. Corollary 3.5 together with Theorem 3.7 then assures that the resulting operator is an ℓ_2–automorphism in the sense of Step 2 of the general concept.

Corollary 3.10 *The operator* $\mathbf{N}$ *defined by*

$$\mathbf{N}\mathbf{U} := \begin{pmatrix} \mathbf{L} & \mathbf{E} \\ \hat{\mathbf{E}} & \mathbf{L}^T \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{p} \\ \mathbf{z} \\ \mathbf{u} \end{pmatrix} := \left(\begin{array}{cc|cc} \mathbf{A} & \mathbf{B}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{B} & \mathbf{0} & \mathbf{0} & -\mathbf{I} \\ \hline \omega \mathbf{T}^T\mathbf{T} & \mathbf{0} & \mathbf{A}^T & \mathbf{B}^T \\ \mathbf{0} & \mathbf{0} & \mathbf{B} & \mathbf{0} \end{array}\right) \begin{pmatrix} \mathbf{y} \\ \mathbf{p} \\ \mathbf{z} \\ \mathbf{u} \end{pmatrix} \tag{3.43}$$

$$= \begin{pmatrix} \boldsymbol{f} \\ \mathbf{0} \\ -\omega\,\mathbf{T}^T\mathbf{y}_{\Box} \\ \mathbf{0} \end{pmatrix} =: \mathbf{F}$$

is an ℓ_2*–automorphism,* $\ell_2 = \ell_2(I\!I) := \ell_2(I\!I_Y \times I\!I_{Q'} \times I\!I_Y \times I\!I_{Q'})$*, i.e., for any* $\boldsymbol{V} \in \ell_2$ *the equivalence*

$$\|\mathbf{N}\boldsymbol{V}\|_{\ell_2} \sim \|\boldsymbol{V}\|_{\ell_2} \tag{3.44}$$

holds.

The constants in (3.44) depend on the weight ω used in (3.24) to balance the two norms, on the constants $\mathbf{c_L}$ and $\mathbf{C_L}$ from (3.17) and on c_*, C_* defined in (3.33).

3.4 Stability of the Finite–Dimensional Systems

To ensure Step 3 of the general concept from Section 2, stability of the resulting finite–dimensional discrete system, one could in principle formulate (3.43) like in [DKS] as a least squares problem and apply the truncation techniques derived there.

However, in view of Proposition 3.6 and Theorem 3.7, the derivation of the system (3.43) reveals that it suffices to ensure that the second derivation $D^2\mathbf{J}$ is uniformly bounded from above and below independent of the discretization and that the discrete stability of the finite–dimensional analog of (3.15) is satisfied. These conditions in turn essentially reduce for the situation at hand to the estimate in the Trace Theorem 3.2 for the finite–dimensional spaces, the inf–sup condition (3.9) called the *Ladyšenskaja–Babuška–Brezzi (LBB) condition*, (3.8), and the continuity estimate (3.10) which is trivially satisfied for subspaces of $H^1(\Box)$ and $H^{1/2}(\Gamma_y)$. In the context of finite elements, there has been an extensive discussion of the LBB condition, starting with the original work [Ba]. Corresponding conditions on the discretizations for finite–dimensional subspaces of $H^1(\Box)$ and $(H^{1/2}(\Gamma))'$ for ensuring discrete stability of the finite–dimensional analog of (3.15) have been investigated in a general context of multiscale schemes and wavelet techniques in [DK] and [K2] for arbitrary dimensions using techniques from approximation theory. Roughly speaking, the LBB condition is satisfied if the discretization on $\Box$ is sufficiently fine relative to the discretization of the boundary. Nevertheless, as the numerical tests in [DK] reveal, the theoretical estimates are much too pessimistic. From a numerical point of view, one often still obtains acceptable results even if the sufficient conditions for the LBB condition derived in [DK] are violated, see also Section 4.3.

Since this subject is extensively discussed in [K2], we dispense here with further details.

Proposition 3.11 *Let finite–dimensional subspaces of $H^1(\Box)$, $(H^{1/2}(\Gamma))'$ and Z be chosen such that the conditions (3.2), (3.3) and the bounds (3.31), (3.32) are satisfied independent of the discretization. Then the resulting operator derived from (3.43) is an automorphism and satisfies (2.26), (2.27) with uniformly bounded constants.*

4 Iterative Solution of the Coupled System

For the remainder of this paper, let finite–dimensional subspaces of all appearing function spaces $H^1(\Box)$ and $(H^{1/2}(\Gamma))'$ and possibly $H^{1/2}(\Gamma_y)$ be chosen such that Proposition 3.11 applies.

Consider corresponding finite–dimensional analogs of (3.15), (3.41) and (3.40) which are for convenience indexed just by Λ,

$$\begin{aligned}
\mathbf{L}_\Lambda^T \begin{pmatrix} \mathbf{z}_\Lambda \\ \boldsymbol{\mu}_\Lambda \end{pmatrix} &= \begin{pmatrix} -\omega \mathbf{T}_\Lambda^T(\mathbf{T}_\Lambda \mathbf{y}_\Lambda - (\mathbf{y}_\Box)_\Lambda) \\ \mathbf{0} \end{pmatrix} \\
\mathbf{u}_\Lambda &= \boldsymbol{\mu}_\Lambda \\
\mathbf{L}_\Lambda \begin{pmatrix} \mathbf{y}_\Lambda \\ \mathbf{p}_\Lambda \end{pmatrix} &= \begin{pmatrix} \boldsymbol{f}_\Lambda \\ \mathbf{u}_\Lambda \end{pmatrix}.
\end{aligned} \tag{4.1}$$

We will now discuss iterative methods for the numerical solution of (4.1).

The system matrix $\mathbf{L}_\Lambda$ in (4.1) is indefinite and unsymmetric, and will typically be very large.

Remark 4.1 *Due to the locality of the operators A, B and T and the wavelet bases $\boldsymbol{\Psi}$, $\tilde{\boldsymbol{\Psi}}$, $\mathbf{L}_\Lambda$ is* sparse *in the sense that it can be* applied *(without being set up explicitly) by using the Fast Wavelet Transform in an amount of arithmetic operations which is proportional to the overall number of unknowns N, see e.g. [D2].*

Because of the particular block structure of $\mathbf{N}_\Lambda$ there are a few methods that suggest themselves, among them block Kaczmarz variants, see e.g. [Ha2]. For the situation at hand, they are discussed in [K2].

4.1 An Inexact Gradient Method

Here we consider an iterative algorithm which can be viewed as an inexact gradient algorithm constructed via the (here as trivially emerging) relation (3.40). First we recall a simple gradient algorithm for the functional $\mathbf{J}$ derived in (3.24) which will serve as a basic iteration.

ALGORITHM BASIC
STEP 1: Fix $\mathbf{u}_\Lambda^{(0)}$;
STEP 2: update $\mathbf{u}_\Lambda^{(i)}$ by computing

$$\mathbf{u}_\Lambda^{(i+1)} := \mathbf{u}_\Lambda^{(i)} - \rho_i \, \delta\mathbf{J}(\mathbf{u}_\Lambda^{(i)}) \tag{4.2}$$

where ρ_i is some step size parameter determined below;
STEP 3: set $i = i+1$ and repeat Steps 2 until a prescribed tolerance for $\mathbf{u}_\Lambda^{(i)}$ is reached.

In view of the properties of the cost functional $\mathbf{J}$ defined in (3.24) collected in Proposition 3.6, one knows from e.g. [Bs, Ci] the following result.

Lemma 4.2 ALGORITHM BASIC *converges for any initial guess* $\mathbf{u}_\Lambda^{(0)} \in \ell_2$ *for step size parameters* ρ_i *satisfying*

$$0 < \rho_* \leq \rho_i \leq \rho^* < 2\frac{c_*}{(C_*)^2}, \tag{4.3}$$

where c_* *and* C_* *are defined in (3.33) and* ρ_*, ρ^* *are some fixed chosen parameters.*

Remark 4.3 *We have chosen here deliberatey the simplest descent method to elaborate the analytic details. The subsequent analysis can readily be extended to more sophisticated versions of gradient methods from e.g. [Bs] or to conjugate gradient iterations. In particular conjugate gradient methods do not require any selection of the step size parameter* ρ_i *and converge faster. However, in view of the subsequent discussion and the complexity analysis in Section 4.2 we dispense with these generalizations here and refer to [KK].*

For the *evaluation* of $\delta\mathbf{J}(\mathbf{u}_\Lambda^{(i)})$, consider the systems

$$\mathbf{L}_\Lambda \begin{pmatrix} \mathbf{y}_\Lambda^{(i+1)} \\ \mathbf{p}_\Lambda^{(i+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}_\Lambda \\ \mathbf{u}_\Lambda^{(i)} \end{pmatrix} \tag{4.4}$$

and

$$\mathbf{L}_\Lambda^T \begin{pmatrix} \mathbf{z}_\Lambda^{(i+1)} \\ \boldsymbol{\mu}_\Lambda^{(i+1)} \end{pmatrix} = -\,\omega \begin{pmatrix} \mathbf{T}_\Lambda^T(\mathbf{T}_\Lambda \mathbf{y}_\Lambda^{(i+1)} - (\mathbf{y}_\Box)_\Lambda) \\ \mathbf{0} \end{pmatrix}. \tag{4.5}$$

In view of Proposition 3.9, the first variation of the cost functional $\mathbf{J}$ from (3.24) satisfies

$$\delta\mathbf{J}(\mathbf{u}_\Lambda^{(i)}) = \mathbf{u}_\Lambda^{(i)} - \boldsymbol{\mu}_\Lambda^{(i+1)} \tag{4.6}$$

where $\boldsymbol{\mu}_\Lambda^{(i+1)}$ solves (4.5) (exactly up to machine precision) which in turn requires the (exact) solution of (4.4). Thus, each step (4.2) requires the successive solution of two mutually adjoint saddle point problems. Of course, solving the systems in (4.1) by factoring $\mathbf{L}_\Lambda$ by e.g. QR decomposition severely constrains the range of applications to relatively small problems. Thus, one looks for fully iterative methods to be able to handle also very large problems.

Solving the systems by QR factorization, convergence of Algorithm Basic has been proved in [GL1] for ρ_i in a certain range depending on the second variation of a differently chosen $\delta\mathbf{J}$ which involves an approximation of a normal derivative.

An alternative to direct methods for saddle point problems is motivated by the fact that by now many efficient iterative methods for saddle point problems exist. Thus, in order to make Algorithm Basic computationally more efficient, we propose here a method that is based on a standard iterative method for solving each of the saddle point problems (4.4) and (4.5) and prove the convergence of the resulting fully iterative method. Based on the convergence results for Algorithm Basic, one guesses that such an iteration converges when both systems are solved iteratively within suitable dynamically chosen tolerances.

We formulate the main algorithm first without specifying the type of iterative method which is used for the solution of the saddle point problems (4.4) and (4.5). We only assume that it converges and call it 'inner iteration', abbreviated as Algorithm InnIt. Since the inexact gradient algorithm involves the iterative solution of both systems, it is termed 'outer iteration'.

Algorithm OutIt

Step 1: Choose $\mathbf{u}_\Lambda^{(0)}$, $\widetilde{\mathbf{y}}_\Lambda^{(0)}$, $\widetilde{\mathbf{p}}_\Lambda^{(0)}$, $\widetilde{\mathbf{z}}_\Lambda^{(0)}$, $\widetilde{\boldsymbol{\mu}}_\Lambda^{(0)}$, set up the blocks $\mathbf{L}_\Lambda$ in (4.4), $\hat{\mathbf{E}}_\Lambda$ in (4.5), and $\boldsymbol{f}_\Lambda$, $(\mathbf{y}_\Box)_\Lambda$; let for integer i $\varepsilon_y(i)$, $\varepsilon_\mu(i)$ be some suitable stage–dependent tolerances to be determined later; set $i = 0$;

Step 2: compute an approximate solution $(\widetilde{\mathbf{y}}_\Lambda^{(i+1)}, \widetilde{\mathbf{p}}_\Lambda^{(i+1)})^T$ of (4.4) with right hand side $(\boldsymbol{f}_\Lambda, \mathbf{u}_\Lambda^{(i)})^T$ and initial guess $(\widetilde{\mathbf{y}}_\Lambda^{(i)}, \widetilde{\mathbf{p}}_\Lambda^{(i)})^T$ that satisfies

$$\left\| \mathbf{L}_\Lambda \begin{pmatrix} \widetilde{\mathbf{y}}_\Lambda^{(i+1)} \\ \widetilde{\mathbf{p}}_\Lambda^{(i+1)} \end{pmatrix} - \begin{pmatrix} \boldsymbol{f}_\Lambda \\ \mathbf{u}_\Lambda^{(i)} \end{pmatrix} \right\|_{\ell_2} < \varepsilon_y(i+1) \tag{4.7}$$

by applying Algorithm InnIt;

Step 3: apply Algorithm InnIt to compute a solution $(\widetilde{\mathbf{z}}_\Lambda^{(i+1)}, \widetilde{\boldsymbol{\mu}}_\Lambda^{(i+1)})^T$ of (4.5) with right hand side $-\omega(\mathbf{T}_\Lambda^T(\mathbf{T}_\Lambda\widetilde{\mathbf{y}}_\Lambda^{(i+1)} - (\mathbf{y}_\Box)_\Lambda), \mathbf{0})^T$ and initial guess $(\widetilde{\mathbf{z}}_\Lambda^{(i)}, \widetilde{\boldsymbol{\mu}}_\Lambda^{(i)})^T$ until

$$\left\| \mathbf{L}_\Lambda^T \begin{pmatrix} \widetilde{\mathbf{z}}_\Lambda^{(i+1)} \\ \widetilde{\boldsymbol{\mu}}_\Lambda^{(i+1)} \end{pmatrix} + \omega \begin{pmatrix} \mathbf{T}_\Lambda^T(\mathbf{T}_\Lambda\widetilde{\mathbf{y}}_\Lambda^{(i+1)} - (\mathbf{y}_\Box)_\Lambda) \\ \mathbf{0} \end{pmatrix} \right\|_{\ell_2} < \varepsilon_\mu(i+1); \tag{4.8}$$

Step 4: update $\mathbf{u}_\Lambda^{(i)}$ by

$$\mathbf{u}_\Lambda^{(i+1)} = \mathbf{u}_\Lambda^{(i)} - \rho_i\,(\mathbf{u}_\Lambda^{(i)} - \widetilde{\boldsymbol{\mu}}_\Lambda^{(i+1)}); \tag{4.9}$$

Step 5: set $i = i + 1$ and repeat Step 2, 3, 4 until a prescribed tolerance for $\mathbf{u}_\Lambda^{(i)}$ is reached.

The convergence of ALGORITHM OUTIT is not clear beforehand since the iterative solution of (4.4) and (4.5) produces an additional error that appears again in the right hand side of the corresponding adjoint system. Thus, in the convergence analysis one needs to assure that the errors produced in the inner iterations do not accumulate and can be fully controlled.

In the following, we derive conditions under which the gradient scheme (4.9) converges, depending on the tolerances $\varepsilon_y(i+1)$ and $\varepsilon_\mu(i+1)$ from (4.7) and (4.8) up to which the two saddle point problems are solved. To this end, recall from e.g. [Bs, Br] that the convergence speed $\theta_{\rm grad}$ of the gradient method (4.2) is governed by the spectral condition number of $\mathbf{Q}_\Lambda$ defined as in (3.27) for the finite–dimensional case,

$$\theta_{\rm grad} = \frac{\kappa(\mathbf{Q}_\Lambda) - 1}{\kappa(\mathbf{Q}_\Lambda) + 1}. \tag{4.10}$$

Due to the preconditioning and scaling of the ingredients of $\mathbf{L}_\Lambda$, $\mathbf{Q}_\Lambda$ has a uniformly bounded condition number such that

$$\theta \le \theta_{\rm grad} \ < \ 1 \tag{4.11}$$

holds *independent* of the discretization. This means that in each iteration of the gradient method the error will be reduced by a fixed fraction θ, i.e.,

$$\|\mathbf{u}_\Lambda^{(i+1)} - \mathbf{u}_\Lambda\|_{\ell_2} \ \le \ \theta\,\|\mathbf{u}_\Lambda^{(i)} - \mathbf{u}_\Lambda\|_{\ell_2} \tag{4.12}$$

where $\mathbf{u}_\Lambda$ is the exact solution of minimizing $\mathbf{J}(\mathbf{u}_\Lambda)$ from (3.24) over (4.4).

Denote the exact solutions $\mathbf{y}_\Lambda^{(i+1)}$ and $\boldsymbol{\mu}_\Lambda^{(i+1)}$ of (4.4) and (4.5), respectively, by

$$\mathbf{y}_\Lambda^{(i+1)} = \mathbf{y}(\mathbf{u}_\Lambda^{(i)}), \qquad \boldsymbol{\mu}_\Lambda^{(i+1)} = \boldsymbol{\mu}(\mathbf{y}_\Lambda^{(i+1)}), \tag{4.13}$$

emphasizing only that dependence on these variables that are relevant for the subsequent analysis. In this notation, (4.9) takes on the form

$$\mathbf{u}_\Lambda^{(i+1)} = \mathbf{u}_\Lambda^{(i)} + \rho_i\left(\mathbf{u}_\Lambda^{(i)} - \boldsymbol{\mu}(\mathbf{y}(\mathbf{u}_\Lambda^{(i)}))\right). \tag{4.14}$$

On the other hand, the iterative solution of (4.4) and (4.5) yields approximations

$$\widetilde{\mathbf{y}}_\Lambda^{(i+1)} \ \approx \ \mathbf{y}(\mathbf{u}_\Lambda^{(i)}), \qquad \widetilde{\boldsymbol{\mu}}_\Lambda^{(i+1)} \ \approx \ \boldsymbol{\mu}(\widetilde{\mathbf{y}}_\Lambda^{(i+1)}), \tag{4.15}$$

respectively. The iteration (4.9) in ALGORITHM OUTIT is executed using $\widetilde{\boldsymbol{\mu}}_\Lambda^{(i+1)}$ instead of $\boldsymbol{\mu}_\Lambda^{(i+1)}$. Adding zeroes, we can rewrite (4.9) in terms of $\mathbf{J}(\mathbf{u}_\Lambda^{(i)})$ from (4.6) as

$$\begin{aligned}
\mathbf{u}_\Lambda^{(i+1)} \ &:= \ \mathbf{u}_\Lambda^{(i)} - \rho_i\left(\mathbf{u}_\Lambda^{(i)} - \widetilde{\boldsymbol{\mu}}_\Lambda^{(i+1)}\right) \\
&= \ \mathbf{u}_\Lambda^{(i)} - \rho_i\Big(\mathbf{u}_\Lambda^{(i)} - \boldsymbol{\mu}(\mathbf{y}(\mathbf{u}_\Lambda^{(i)})) \\
&\qquad\qquad + \ \boldsymbol{\mu}(\mathbf{y}(\mathbf{u}_\Lambda^{(i)})) - \boldsymbol{\mu}(\widetilde{\mathbf{y}}_\Lambda^{(i+1)}) + \boldsymbol{\mu}(\widetilde{\mathbf{y}}_\Lambda^{(i+1)}) - \widetilde{\boldsymbol{\mu}}_\Lambda^{(i+1)}\Big) \\
&= \ \mathbf{u}_\Lambda^{(i)} - \rho_i\,\mathbf{J}(\mathbf{u}_\Lambda^{(i)}) \\
&\qquad\qquad + \ \rho_i\left(\boldsymbol{\mu}(\mathbf{y}(\mathbf{u}_\Lambda^{(i)})) - \boldsymbol{\mu}(\widetilde{\mathbf{y}}_\Lambda^{(i+1)}) + \boldsymbol{\mu}(\widetilde{\mathbf{y}}_\Lambda^{(i+1)}) - \widetilde{\boldsymbol{\mu}}_\Lambda^{(i+1)}\right) \\
&= \ \widehat{\mathbf{u}}_\Lambda^{(i+1)} + \rho_i\left(\boldsymbol{\mu}(\mathbf{y}(\mathbf{u}_\Lambda^{(i)})) - \boldsymbol{\mu}(\widetilde{\mathbf{y}}_\Lambda^{(i+1)}) + \boldsymbol{\mu}(\widetilde{\mathbf{y}}_\Lambda^{(i+1)}) - \widetilde{\boldsymbol{\mu}}_\Lambda^{(i+1)}\right),
\end{aligned} \tag{4.16}$$

where $\widehat{\mathbf{u}}_\Lambda^{(i+1)}$ is the solution of the exact gradient step (4.2). Recalling that the gradient method (4.2) satisfies the error reduction estimate (4.12) with fixed $\theta < 1$, we obtain by inserting (4.16)

$$\begin{aligned} \|\mathbf{u}_\Lambda^{(i+1)} - \mathbf{u}_\Lambda\|_{\ell_2} &\leq \|\widehat{\mathbf{u}}_\Lambda^{(i+1)} - \mathbf{u}_\Lambda\|_{\ell_2} + \|\mathbf{u}_\Lambda^{(i+1)} - \widehat{\mathbf{u}}_\Lambda^{(i+1)}\|_{\ell_2} \qquad (4.17) \\ &\leq \theta\, \|\mathbf{u}_\Lambda^{(i)} - \mathbf{u}_\Lambda\|_{\ell_2} + \rho_i\, \|\boldsymbol{\mu}(\mathbf{y}(\mathbf{u}_\Lambda^{(i)})) - \boldsymbol{\mu}(\widetilde{\mathbf{y}}_\Lambda^{(i+1)}) + \boldsymbol{\mu}(\widetilde{\mathbf{y}}_\Lambda^{(i+1)}) - \widetilde{\boldsymbol{\mu}}_\Lambda^{(i+1)}\|_{\ell_2} \\ &\leq \theta\, \|\mathbf{u}_\Lambda^{(i)} - \mathbf{u}_\Lambda\|_{\ell_2} + \rho_i \Big(\mathbf{c}_{\mathbf{L}}^{-1} \|\mathbf{y}(\mathbf{u}_\Lambda^{(i)}) - \widetilde{\mathbf{y}}_\Lambda^{(i+1)}\|_{\ell_2} \\ &\qquad + \|\boldsymbol{\mu}(\widetilde{\mathbf{y}}_\Lambda^{(i+1)}) - \widetilde{\boldsymbol{\mu}}_\Lambda^{(i+1)}\|_{\ell_2} \Big) \end{aligned}$$

where we have used also that $\mathbf{L}_\Lambda$ is an isomorphism with constants given in (3.17). Recalling the upper estimate in (3.17), this yields together with (4.7) and (4.8)

$$\begin{aligned} \|\mathbf{u}_\Lambda^{(i+1)} - \mathbf{u}_\Lambda\|_{\ell_2} &\leq \theta\, \|\mathbf{u}_\Lambda^{(i)} - \mathbf{u}_\Lambda\|_{\ell_2} + \rho_i \left(\mathbf{c}_{\mathbf{L}}^{-1} \mathbf{C}_{\mathbf{L}} \varepsilon_y(i+1) + \mathbf{C}_{\mathbf{L}} \varepsilon_\mu(i+1) \right) \\ &= \theta\, \|\mathbf{u}_\Lambda^{(i)} - \mathbf{u}_\Lambda\|_{\ell_2} + e_i \qquad (4.18) \end{aligned}$$

where

$$e_i := \rho_i \left(\mathbf{c}_{\mathbf{L}}^{-1} \mathbf{C}_{\mathbf{L}} \varepsilon_y(i+1) + \mathbf{C}_{\mathbf{L}} \varepsilon_\mu(i+1) \right). \qquad (4.19)$$

Repeating this argument leads to the recursion

$$\|\mathbf{u}_\Lambda^{(i+1)} - \mathbf{u}_\Lambda\|_{\ell_2} \leq \theta^{i+1}\, \|\mathbf{u}_\Lambda^{(0)} - \mathbf{u}_\Lambda\|_{\ell_2} + \sum_{l=0}^{i} \theta^l\, e_{i-l}. \qquad (4.20)$$

Thus, when e_i satisfies e.g.

$$e_i \leq \frac{\theta^i}{(1+i)^2}\, \sigma(\mathbf{u}_\Lambda), \qquad (4.21)$$

where $\sigma(\mathbf{u}_\Lambda)$ is an estimate for the initial error with respect to Λ,

$$\|\mathbf{u}_\Lambda^{(0)} - \mathbf{u}_\Lambda\|_{\ell_2} \leq \sigma(\mathbf{u}_\Lambda), \qquad (4.22)$$

one can conclude from (4.20)

$$\begin{aligned} \|\mathbf{u}_\Lambda^{(i+1)} - \mathbf{u}_\Lambda\|_{\ell_2} &\leq \theta^{i+1}\, \sigma(\mathbf{u}_\Lambda) + \theta^i \sigma(\mathbf{u}_\Lambda) \sum_{l=0}^{i} \frac{1}{(1+l-i)^2} \\ &=: \theta^{i+1}\, \sigma(\mathbf{u}_\Lambda) + \theta^i\, \sigma(\mathbf{u}_\Lambda)\, \bar{c} \\ &\leq \theta^i\, \sigma(\mathbf{u}_\Lambda)\, (\theta + \bar{c}). \qquad (4.23) \end{aligned}$$

Summarizing, we have proved convergence of the fully iterative scheme ALGORITHM OUTIT under the condition that e_i satisfies (4.21). This is the case if for instance at the $(i+1)$th stage the tolerances $\varepsilon_y(i+1)$, $\varepsilon_\mu(i+1)$ are chosen as follows.

Theorem 4.4 *If the tolerances $\varepsilon_y(i+1)$ and $\varepsilon_\mu(i+1)$ in (4.7) and (4.8) are selected at each stage according to*

$$\begin{aligned}\varepsilon_y(i+1) &:= \frac{1}{2}\frac{\mathbf{c_L}}{\mathbf{C_L}\,\rho_i}\frac{\theta^i}{(1+i)^2}\,\sigma(\mathbf{u}_\Lambda),\\ \varepsilon_\mu(i+1) &:= \frac{1}{2}\frac{1}{\mathbf{C_L}\,\rho_i}\frac{\theta^i}{(1+i)^2}\,\sigma(\mathbf{u}_\Lambda),\end{aligned} \tag{4.24}$$

then ALGORITHM OUTIT *converges for ρ_i satisfying (4.3),*

$$0 < \rho_* \;\le\; \rho_i \;\le\; \rho^* < 2\frac{c_*}{{C_*}^2},$$

where c_* and C_* are defined in (3.33).

For practical purposes, the constants θ, C_* can be estimated and computed in terms of the condition number of $\mathbf{L}_\Lambda$, see also [K2] for numerical experiments. In view of Remark 4.3, the selection of ρ_i can be avoided by resorting to more sophisticated gradient methods for which, however, the analysis is more involved.

In the next section, a detailed complexity analysis shows that ALGORITHM OUTIT leads in combination with a nested iteration strategy to an *asymptotically optimal* method.

4.2 Computational Work for ALGORITHM OUTIT

Up to this point, we have not specified the particular iterative method ALGORITHM INNIT by which (4.4) and (4.5) are solved. A simple iterative method for saddle point problems for symmetric $\mathbf{A}_\Lambda$ is the *Uzawa algorithm.* For a system of the form

$$\mathbf{L}_\Lambda\begin{pmatrix}\mathbf{y}_\Lambda\\ \mathbf{p}_\Lambda\end{pmatrix} \equiv \begin{pmatrix}\mathbf{A}_\Lambda & \mathbf{B}_\Lambda^T\\ \mathbf{B}_\Lambda & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{y}_\Lambda\\ \mathbf{p}_\Lambda\end{pmatrix} = \begin{pmatrix}\boldsymbol{f}_\Lambda\\ \mathbf{g}_\Lambda\end{pmatrix}, \tag{4.25}$$

the Uzawa algorithm reads in its simplest form (assuming that $\mathbf{A}_\Lambda$ is invertible on the whole space) for $i = 0, 1, \ldots$ when $\mathbf{y}_\Lambda^{(i)}, \mathbf{p}_\Lambda^{(i)}$ are chosen,

$$\begin{aligned}\mathbf{y}_\Lambda^{(i+1)} &= \mathbf{A}_\Lambda^{-1}(\boldsymbol{f}_\Lambda - \mathbf{B}_\Lambda^T\mathbf{p}_\Lambda^{(i)})\\ &= \mathbf{y}_\Lambda^{(i)} + \mathbf{A}_\Lambda^{-1}(\boldsymbol{f}_\Lambda - \mathbf{A}_\Lambda\mathbf{y}_\Lambda^{(i)} - \mathbf{B}_\Lambda^T\mathbf{p}_\Lambda^{(i)})\\ \mathbf{p}_\Lambda^{(i+1)} &= \mathbf{p}_\Lambda^{(i)} + \gamma\,(\mathbf{B}_\Lambda\mathbf{y}_\Lambda^{(i+1)} - \mathbf{g}_\Lambda).\end{aligned} \tag{4.26}$$

Here γ is some sufficiently small fixed step size parameter. The first system in (4.26) is not solved exactly. Its iterative solution by e.g. the conjugate gradient method corresponds to applying some approximation $(\mathbf{A}_\Lambda)_0^{-1}$ of $\mathbf{A}_\Lambda^{-1}$ which can be viewed as a preconditioner for $\mathbf{A}_\Lambda$. One usually also includes a preconditioner $(\mathbf{S}_\Lambda)_0$ for the second equation,

$$\begin{aligned}\mathbf{y}_\Lambda^{(i+1)} &= \mathbf{y}_\Lambda^{(i)} + (\mathbf{A}_\Lambda)_0^{-1}(\boldsymbol{f}_\Lambda - \mathbf{A}_\Lambda\mathbf{y}_\Lambda^{(i)} - \mathbf{B}_\Lambda^T\mathbf{p}_\Lambda^{(i)})\\ \mathbf{p}_\Lambda^{(i+1)} &= \mathbf{p}_\Lambda^{(i)} + \gamma\,(\mathbf{S}_\Lambda)_0^{-1}(\mathbf{B}_\Lambda\mathbf{y}_\Lambda^{(i+1)} - \mathbf{g}_\Lambda).\end{aligned} \tag{4.27}$$

The role of $(\mathbf{S}_\Lambda)_0$ is explained below. Algorithm (4.27) is often called *incomplete Uzawa algorithm* since the iterative method for the first equation corresponds to multiplying by an approximation $(\mathbf{A}_\Lambda)_0^{-1}$ of $\mathbf{A}_\Lambda^{-1}$, see [BPV].

For discussing the convergence properties of (4.26), one considers the *reduced equation*

$$\mathbf{B}_\Lambda \mathbf{A}_\Lambda^{-1} \mathbf{B}_\Lambda^T \mathbf{p}_\Lambda = \mathbf{B}_\Lambda \mathbf{A}_\Lambda^{-1} \boldsymbol{f}_\Lambda - \mathbf{g}_\Lambda \tag{4.28}$$

involving the *Schur complement* of (4.25). For symmetric and positive definite $\mathbf{A}_\Lambda$, the Uzawa method (4.26) is known to converge if $\mathbf{B}_\Lambda \mathbf{A}_\Lambda^{-1} \mathbf{B}_\Lambda^T$ is symmetric positive definite and if e.g. the step size parameter γ satisfies

$$\gamma \;<\; 2\, \|\mathbf{B}_\Lambda \mathbf{A}_\Lambda^{-1} \mathbf{B}_\Lambda^T\|_2^{-1}, \tag{4.29}$$

see e.g. [DHU] for a detailed derivation. In fact, an iteration for (4.28) reads

$$\mathbf{p}_\Lambda^{(i+1)} = (\mathbf{I} - \gamma\, \mathbf{B}_\Lambda \mathbf{A}_\Lambda^{-1} \mathbf{B}_\Lambda^T)\, \mathbf{p}_\Lambda^{(i)} + \mathbf{B}_\Lambda \mathbf{A}_\Lambda^{-1} \boldsymbol{f}_\Lambda - \mathbf{g}_\Lambda \tag{4.30}$$

which converges if

$$\|\mathbf{I} - \gamma\, \mathbf{B}_\Lambda \mathbf{A}_\Lambda^{-1} \mathbf{B}_\Lambda^T\|_2 < 1 \tag{4.31}$$

which follows from (4.29). In the present situation, the preconditioner $\mathbf{S}_0$ is actually only needed for a possibly diagonal scaling since the Schur complement already has a uniformly bounded condition number, see the proof of Remark 4.5 below.

For the systems (4.4) and (4.5) which satisfy Corollary 3.10, we can say the following.

Remark 4.5 *The convergence rate of solving (4.4) or (4.5) by the incomplete Uzawa algorithm (4.27) is for suitable choices of* $(\mathbf{A}_\Lambda)_0, (\mathbf{S}_\Lambda)_0$ *independent of the discretization.*

Proof:
The convergence rate θ_{Uz} of the Uzawa algorithm (4.26) (using conjugate gradient iterations for equation (4.30) and computing $\mathbf{A}_\Lambda^{-1}$ exactly) is governed by the spectral condition number of the Schur complement,

$$\theta_{\mathrm{Uz}} = \frac{\sqrt{\kappa(\mathbf{B}_\Lambda \mathbf{A}_\Lambda^{-1} \mathbf{B}_\Lambda^T)} - 1}{\sqrt{\kappa(\mathbf{B}_\Lambda \mathbf{A}_\Lambda^{-1} \mathbf{B}_\Lambda^T)} + 1}, \tag{4.32}$$

see e.g. [Br]. A proper scaling yields that $\mathbf{B}_\Lambda \mathbf{A}_\Lambda^{-1} \mathbf{B}_\Lambda^T$ is an ℓ_2–automorphism such that θ_{Uz} is independent of the refinement level. From the analysis in [BPV] one has that the convergence rate for the incomplete Uzawa algorithm (4.27) satisfies

$$\theta_{\mathrm{iUz}} \;=\; \frac{\delta_1(1-\delta_2) + \sqrt{\delta_1^2(1-\delta_2)^2 + 4\,\delta_2}}{2}, \tag{4.33}$$

where $0 < \delta_1, \delta_2 < 1$ are the convergence rates for each of the iterations in (4.27). Since $\mathbf{A}_\Lambda$ and $\mathbf{B}_\Lambda \mathbf{A}_\Lambda^{-1} \mathbf{B}_\Lambda^T$ have uniformly bounded condition numbers, the action of $(\mathbf{A}_\Lambda)_0^{-1}, (\mathbf{S}_\Lambda)_0^{-1}$ only means a scaling by a constant independent of the discretization such that indeed the convergence rates satisfy $\delta_1, \delta_2 < 1$. Thus, the convergence rate θ_{iUz} is also independent of the discretization level. Furthermore, since θ_{iUz} can be estimated as

$$\theta_{\mathrm{iUz}} \;\le\; 1 - \tfrac{1}{2}(1-\delta_1)(1-\delta_2)$$

it follows that

$$\theta_{\mathrm{iUz}} \; < 1.$$

■

Consequently, choosing the incomplete Uzawa method (4.27) as inner iteration ALGORITHM INNIT in ALGORITHM OUTIT, in both STEP 2 and STEP 3 for any size of the systems (4.7), (4.8) only a *fixed* number of iterations is needed to reduce the error by a fixed fraction. Recall also that each iteration can be applied in an amount of work proportional to the size of the system since all operators in $\mathbf{L}_\Lambda$ can be realized by successively applying sparse matrices.

In the following, we exploit the combination of the basic iterative method ALGORITHM OUTIT and the analysis used to prove Theorem 4.4 with a *nested iteration strategy* to derive an *asymptotically optimal* algorithm. In particular, the accuracy of the iterative solution of the two systems has to be made dependent also on the discretization error for $\mathbf{u}$.

The derivation will be detailed for *uniform* refinements with discretization step size $h \sim 2^{-j}$ so that we exchange the index Λ by the parameter j denoting some *discretization level*. Denote by $\tau(\mathbf{u}_j)$ an estimate for the *discretization error* for the control on level j, i.e.,

$$\|\mathbf{u}_j - \mathbf{u}\|_{\ell_2} \;\le\; \tau(\mathbf{u}_j), \tag{4.34}$$

where $\mathbf{u}$ is the exact solution of the infinite–dimensional system (3.22) and $\mathbf{u}_j$ is the exact solution of (4.5) on level j. An estimation of $\tau(\mathbf{u}_j)$ is given in (4.41) below.

Suppose that we have determined on level j an approximation $\mathbf{u}_j^{(i)}$ such that

$$\|\mathbf{u}_j^{(i)} - \mathbf{u}_j\|_{\ell_2} \;\le\; \tfrac{1}{2}\tau(\mathbf{u}_j). \tag{4.35}$$

Now choose $\mathbf{u}_j^{(i)}$ as initial guess for the iteration (4.9) on the next higher refinement level $j+1$. Note that since $\mathbf{u}_j^{(i)}$ consists of wavelet coefficients we can formally set

$$\mathbf{u}_{j+1}^{(0)} := \mathbf{u}_j^{(i)}, \tag{4.36}$$

meaning that the array $\mathbf{u}_j^{(i)}$ is extended to the vector of wavelet coefficients on level $j+1$ by simply appending zeros. We will determine next a number $i+1 = i_{j+1}$ of gradient iterations which is needed to reduce the initial error

$$\|\mathbf{u}_{j+1}^{(0)} - \mathbf{u}_{j+1}\|_{\ell_2}$$

on level $j+1$ so that

$$\|\mathbf{u}_{j+1}^{(i_{j+1})} - \mathbf{u}_{j+1}\|_{\ell_2} \;\le\; \frac{\tau(\mathbf{u}_{j+1})}{2}, \tag{4.37}$$

which would advance (4.35). Recall from (4.20) and (4.23) that one has

$$\|\mathbf{u}_{j+1}^{(i+1)} - \mathbf{u}_{j+1}\|_{\ell_2} \;\leq\; \theta^{i+1}\,\|\mathbf{u}_{j+1}^{(0)} - \mathbf{u}_{j+1}\|_{\ell_2} + \theta^i \bar{c}\,\tau(\mathbf{u}_j), \tag{4.38}$$

provided that the bounds e_i satisfy now the following counterpart to (4.21),

$$e_i \;\leq\; \frac{\theta^i}{(1+i)^2}\,\tau(\mathbf{u}_j). \tag{4.39}$$

Moreover, from (4.35), (4.36) and (4.34) we infer

$$\begin{aligned} \|\mathbf{u}_{j+1}^{(0)} - \mathbf{u}_{j+1}\|_{\ell_2} &\leq \|\mathbf{u}_j^{(i)} - \mathbf{u}_j\|_{\ell_2} + \|\mathbf{u}_j - \mathbf{u}\|_{\ell_2} + \|\mathbf{u}_{j+1} - \mathbf{u}\|_{\ell_2} \\ &\leq \tfrac{1}{2}\tau(\mathbf{u}_j) + \tau(\mathbf{u}_j) + \tau(\mathbf{u}_{j+1}) \\ &= \tfrac{3}{2}\tau(\mathbf{u}_j) + \tau(\mathbf{u}_{j+1}). \end{aligned} \tag{4.40}$$

At this point, we need to recall some facts about the underlying framework. The discrete solutions $\mathbf{u}_j$ belong to spaces S_j which in turn form an ascending sequence with growing j. Since this sequence corresponds to *uniform* refinements, standard error estimates in the natural $H^{1/2}(\Gamma)$–norm for u, see e.g. [BF], combined with direct estimates, see e.g. [D2], yield that the minimal rate of convergence is determined by

$$\|\mathbf{u}_j - \mathbf{u}\|_{\ell_2} \;\lesssim\; 2^{-sj}\|\mathbf{u}\|_{\ell_2} \tag{4.41}$$

where for the present application one has at least $0 < s < 1/2$. Thus, either

$$\frac{\tau(\mathbf{u}_{j+1})}{\tau(\mathbf{u}_j)} \;=:\; \delta \;<\; 1 \tag{4.42}$$

holds with δ proportional to 2^{-s}, or already $\mathbf{u} \in S_j$. In the case of non–uniform refinements, a strategy like the one proposed in [DHU] assures that (4.42) holds. Now we can further estimate (4.38) by employing (4.40) and (4.42),

$$\|\mathbf{u}_{j+1}^{(i+1)} - \mathbf{u}_{j+1}\|_{\ell_2} \;\leq\; \theta^i\left(\tfrac{3\theta}{2} + \theta\delta + \bar{c}\right)\tau(\mathbf{u}_j). \tag{4.43}$$

Thus, choosing $i = i_{j+1} - 1$ so that

$$\frac{3\theta}{2} + \theta\delta + \bar{c} \;\leq\; \delta/2,$$

we have verified (4.37). In particular, the numbers i_{j+1} can be kept uniformly bounded independent of the level j. Note that in terms of the tolerances $\varepsilon_y(i+1)$ and $\varepsilon_\mu(i+1)$ that control the accuracy of the inner iterations in (4.7) and (4.8) one needs, in view of (4.39),

$$\begin{aligned} \varepsilon_y(i+1,j) &:= \frac{1}{2}\,\frac{\mathbf{c_L}}{\mathbf{C_L}\,\rho_i}\,\frac{\theta^i}{(i+1)^2}\,\tau(\mathbf{u}_j), \\ \varepsilon_\mu(i+1,j) &:= \frac{1}{2}\,\frac{1}{\mathbf{C_L}\,\rho_i}\,\frac{\theta^i}{(i+1)^2}\,\tau(\mathbf{u}_j). \end{aligned} \tag{4.44}$$

Since as mentioned above $i = i_{j+1} - 1$ remains bounded, the quotients

$$\frac{\varepsilon_y(i+1,j)}{\varepsilon_y(i,j)} \qquad \text{and} \qquad \frac{\varepsilon_\mu(i+1,j)}{\varepsilon_\mu(i,j)}$$

remain proportional to δ. Thus, choosing $\widetilde{\mathbf{y}}_j^{(i)}$ and $\widetilde{\boldsymbol{\mu}}_j^{(i)}$ as initial guesses for the computation of $\widetilde{\mathbf{y}}_{j+1}^{(i+1)}$ and $\widetilde{\boldsymbol{\mu}}_{j+1}^{(i+1)}$ in (4.7) and (4.8), only a fixed level independent error reduction is used which under the above assumptions requires only a fixed number of inner iterations independent of j.

Now we can argue as follows. The systems in (4.7) and (4.8) can be solved by e.g. the incomplete Uzawa method (4.27) with a convergent rate independent of the discretizations, see Remark 4.5. Thus, taking as initial guess the solution from the previous level, only a uniformly bounded number of Uzawa steps is required to reduce the error by a fixed fraction which is all that is needed to achieve discretization error accuracy on each level. Moreover, the operators in (4.7) and (4.8) can be applied at the expense of computational work that remains proportional to the number of unknowns on that level. Thus, with a geometric series argument it follows that the overall work stays proportional to the computational work required by a matrix/vector multiplication on the highest level J, that is, the total work is proportional to $\mathcal{O}(N_J)$ where N_J is the number of unknowns on the highest level. Summarizing, we have finally proved the following.

Theorem 4.6 *If in each iteration of* Algorithm OutIt *the systems (4.7) and (4.8) are solved up to error (4.44) and these solutions are taken as initial guesses for the next higher level, then* Algorithm OutIt *is an asymptotically optimal method in the sense that it provides the solution up to discretization error on level J in an overall amount of $\mathcal{O}(N_J)$ operations where N_J is the number of unknowns in (4.7), (4.8) and (4.9).*

It should be remarked that in the above strategy the main goal is to yield an asymptotically optimal result. To obtain a quantitavely efficient scheme for each J, more care has to be taken. Corresponding results on more sophisticated iterative methods which, however, exhibit the same qualitative asymptotics, will be reported elsewhere.

4.3 A Numerical Example

We close with the following numerical example. The elliptic boundary value problem that plays the role of the constraints for the control problem (1.13) is the problem

$$\begin{aligned} -\Delta y + y &= 1 \quad \text{in } \Omega, \\ y &= u \quad \text{on } \partial\Omega. \end{aligned}$$

Here Ω is the disc with radius $R = 0.5$ around the mid point $(0.5, 0.5)$, $\Omega = \{x \in \mathbb{R}^2 : \|x - \binom{1/2}{1/2}\|_{\ell_2}^2 < R\}$, which is embedded into the fictitious domain $\square = (0,1)^2$. In the minimization functional (1.13), we have chosen $Z = Y = H^1(\square)$, $y_\square \equiv 0$ and $\omega = 1$. Thus, the primal system and corresponding adjoint system read in discretized form according to (4.4) and (4.5)

$$\mathbf{L}_\Lambda \begin{pmatrix} \mathbf{y}_\Lambda \\ \mathbf{p}_\Lambda \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}_\Lambda \\ \mathbf{u}_\Lambda \end{pmatrix}, \qquad \mathbf{L}_\Lambda \begin{pmatrix} \mathbf{z}_\Lambda \\ \mathbf{u}_\Lambda \end{pmatrix} = -\begin{pmatrix} \mathbf{y}_\Lambda \\ \mathbf{0} \end{pmatrix}. \tag{4.45}$$

j $\square$	ℓ Γ	1st it. STEP 2	2nd it. STEP 3	3rd it. STEP 2	error	time
3	3	37(2)	21(1)	11(1)	$1.22007090e-1$	0.046
3	4	62(4)	22(1)	22(2)	$3.03893551e-2$	0.070
3	5	69(4)	23(1)	12(1)	$2.55219641e-2$	0.070
3	6	437(32)	25(1)	38(3)	$1.40530157e-2$	0.300
4	3	70(4)	28(1)	14(1)	$5.64009761e-2$	0.223
4	4	78(4)	28(1)	14(1)	$4.89317933e-2$	0.256
4	5	138(8)	30(1)	43(3)	$1.88224372e-2$	0.429
4	6	530(33)	33(1)	31(2)	$1.17742201e-2$	1.149
4	7	2055(119)	55(2)	4294(260)	$5.25188732e-3$	12.226
5	3	85(4)	33(1)	30(2)	$1.07381850e-2$	1.161
5	4	89(4)	34(1)	31(2)	$1.73216501e-2$	1.186
5	5	111(5)	35(1)	17(1)	$2.43685913e-2$	1.312
5	6	296(16)	38(1)	49(3)	$8.63815998e-3$	2.965
5	7	923(49)	64(2)	70(4)	$5.10956365e-3$	7.982
6	3	112(5)	58(2)	20(1)	$1.43552113e-2$	6.668
6	4	118(5)	39(1)	35(2)	$1.24542308e-2$	7.054
6	5	172(8)	40(1)	36(2)	$1.13766598e-2$	9.270
6	6	203(9)	42(1)	68(4)	$1.26067723e-2$	11.716
6	7	612(30)	70(2)	21(1)	$5.97800253e-3$	26.456
7	3	144(6)	84(3)	23(1)	$4.34969885e-3$	53.324
7	4	184(8)	68(2)	23(1)	$6.45925688e-3$	58.098
7	5	211(9)	69(2)	42(2)	$5.90390441e-3$	64.375
7	6	337(15)	72(2)	60(3)	$2.73841284e-3$	105.074

Table 1: Iteration numbers (total number of PCG iterations with number of Uzawa steps in parantheses) for the coupled saddle point problem (4.45) solved by ALGORITHM OUTIT.

Since the multiresolution spaces on the boundary Γ can be defined via periodization, $\ell_0 = 0$ is chosen as the coarsest level on the boundary. On $\square$, tensor products of the biorthogonal wavelets of order $d_\square = 2$ and also order $\tilde{d}_\square = 4$ for the dual multiresolution constructed in [DKU1] have been employed. The lowest level on $\square$ is therefore $j_0 = 3$. For simplicity, only hierarchies of *uniformly refined* trial spaces on $\square$ and Γ are considered. The first two columns in Table 1 display the finest discretization level j on $\square$ and ℓ on Γ. We have imposed *no geometric coupling* between the discretization spaces on $\square$ and Γ. According to the results from [DK], the LBB condition is violated when e.g. the refinement level ℓ on Γ is higher than the one on the domain.

The system (4.45) is solved by applying the following variant of ALGORITHM OUTIT. In STEP 3 the system (4.8) is solved for $\mathbf{z}_\Lambda^{(i+1)}$ and $\mathbf{u}_\Lambda^{(i+1)}$ instead of $\boldsymbol{\mu}_\Lambda^{(i+1)}$, and STEP 4 is discarded since the update of $\mathbf{u}_\Lambda^{(i+1)}$ is already performed in STEP 3. As inner iteration, the Uzawa algorithm is used with a preconditioned conjugate gradient method (PCG method) to solve the first equation in (4.26) iteratively. Our stopping criterion is based on the ℓ_2 norm of the residual which is proportional to the error of y in H^1. The first numbers in the third column reveal the total number of CG iterations necessary to force the ℓ_2 error of the residual to be smaller than `tol`$= \min\{2^{-j}, 2^{-\ell}\}$. The inner iterations are terminated when the error is smaller than 0.01$*$`tol`. The numbers in parantheses show the number of Uzawa iterations.

It is observed that for these tolerances this variant of Algorithm OutIt always terminates after 1 cycle, that is, system (4.4) is solved in Step 2 up to the requested tolerance, followed by the solution of the system (4.5) in Step 3. Going back to Step 2 once again is in all cases sufficient to meet the required overall tolerance. This is perhaps due to the fact that the saddle point systems are mutually adjoint and have uniformly bounded condition numbers. For this reason, the cycle Step 2—Step 3—Step 2 is called *solution cycle* for the coupled saddle point problem (4.1). In Table 1 the iteration numbers for each step of the solution cycle are listed in columns 3, 4 and 5 and are termed *1st it.*, *2nd it.* and *3rd it.* The numbers displayed are the total number of pcg-iterations with the number of Uzawa steps following in parantheses. The total error for system (4.4) measured as before and the total time in CPU seconds is listed in the last two columns of Table 1. The iteration numbers in the 3rd iteration are always smaller than the ones from the 1st iteration since their solutions are taken as initial values. It is known that the absolute numbers in the table can be further reduced by taking more sophisticated boundary adaptations of the wavelets on $\Box$ [Bar, DKU2].

As in the experiments in [DK], the slow growth in the condition numbers when increasing ℓ relative to j accounts for the fact that the sufficient conditions for assuring the LBB condition are violated. Only when here $\ell \geq j+3$, this is negatively affecting the conditions numbers quantitatively and the computing time. Different variants on the tolerances have been tested which reveals that Algorithm OutIt turns out to be relatively robust.

There are many variants one can think of to balance the amount of iterations needed in each step of the cycle with the necessary amount of iterations. Also, since the convergence analysis employs norm equivalences, one expects that the convergence of the gradient algorithm or conjugate gradient accelerations are affected by increasing the value of the weight ω in the definition of the cost functional. Here a more sophisticated definition of norms for Lorentz spaces which seems to provide also an analytic tool for deriving robust schemes for elliptic singular problems may provide a solution [GKSch].

This paper has been dealing with linear problems only. Indeed, there are many investigations of optimal control problems involving nonlinear equations like the Navier–Stokes equations, see e.g. [Hei1, Hei2, HK1, HK2, IK, Kf, Glo, Gu, GHS1, GHS2, GL2] where the numerical solution typically involves in addition Newton–type methods for the solution of the primal system (4.4). The convergence of Algorithm OutIt with iterative methods for nonlinear problems embedded into the wavelet setting is the subject of a forthcoming paper [K4]. The necessary evaluation of nonlinear functionals in terms of wavelets has been discussed in [DSX].

Moreover, it would be interesting to combine the above ideas with *adaptive* solution strategies. Meanwhile numerous investigations of adaptive wavelet concepts have been documented in the literature. A systematic approach based on a–posteriori error estimators was considered in [Be] for a special case and in [DDHS] for a wide class of symmetric elliptic operator equations where, in particular, convergence in the energy norm could be established rigorously. While these results still leave the question of computational complexity open, a modified scheme for this class of problems has recently been developed in [CDD1]. There estimates for convergence rates and computational complexity have been obtained that are asymptotically optimal. Roughly speaking, this means that the computational work stays proportional to the number of significant wavelet coefficients that are needed to recover the solution within any desired accuracy tolerance. Besides the development of corresponding new algorithmic ingredients and data structures in [BCDU], the numerical experiments obtained there confirm the predicted behavior.

Quite recently, there have also been attempts to extend these results to indefinite problems [DHU]. There the results from [DDHS] have been carried over to saddle point problems without establishing however any complexity estimates. Extensions of the convergence and complexity results in [CDD1] to a wider scope of problems such as saddle point problems or just those for which (3.17) hold have just been obtained in [CDD2]. The special case of adaptive methods for saddle point problems based on an Uzawa algorithm is together with an implementation currently under investigation [DDU].

Acknowledgements: I wish to thank very much Wolfgang Dahmen for many encouraging discussions and the anonymous referees for their suggestions, leading to a substantial improvement of the manuscript.

References

[AN] V. Arnăutu, P. Neittaanmäki, *Discretization estimates for an elliptic control problem*, Numer. Funct. Anal. Optim. 19, 1998, 431–464.

[As] G. P. Astrakantsev, *Methods of fictitious domains for a second order elliptic equation with natural boundary conditions*, U.S.S.R. Comput. Math. Math. Phys. 18, 1978, 114–121.

[Ba] I. Babuška, *The finite element method with Lagrange multipliers*, Numer. Math. 20, 1973, 179–192.

[BWY] R.E. Bank, B.D. Welfert, H. Yserentant, *A class of iterative methods for solving saddle point problems*, Numer. Math. 56, 1990, 645-666.

[BCDU] A. Barinka, T. Barsch, Ph. Charton, A. Cohen, S. Dahlke, W. Dahmen, K. Urban, *Adaptive wavelet schemes for elliptic problems — Implementation and numerical experiments*, IGPM–Preprint #173, RWTH Aachen, June 1999, to appear in SIAM J. Sci. Comp.

[Bar] T. Barsch, *Adaptive Multiskalenverfahren für elliptische partielle Differentialgleichungen — Realisierung, Umsetzung und numerische Ergebnisse* (in German), IGPM, RWTH Aachen, PhD Thesis, 2001.

[BKR] R. Becker, H. Kapp, R. Rannacher, *Adaptive finite element methods for optimal control of partial differential equations: Basic concept*, Preprint 1998–55, University of Heidelberg, November 1998, to appear in: SIAM J. Contr. Optim.

[Be] S. Bertoluzza, *A posteriori error estimates for the wavelet Galerkin method*, Appl. Math. Lett. 8, 1995, 1–6.

[Bs] D.P. Bertsekas, *Nonlinear Programming*, 2nd ed., Athena Scientific, 1999.

[Bo] P.B. Bochev, *Least-squares methods for optimal control*, Nonlin. Anal. Theory Methods Appl. 30, 1997, 1875–1885.

[Br] D. Braess, *Finite Elements: Theory, Fast Solvers and Applications in Solid Mechanics*, Cambridge University Press, 1997.

[BP] J.H. Bramble, J. Pasciak, *A preconditioning technique for indefinite systems resulting from mixed approximations for elliptic problems*, Math. Comp. 50, 1988, 1–17.

[BPV] J.H. Bramble, J. Pasciak, A.T. Vassilev, *Analysis of the inexact Uzawa algorithm for saddle point problems*, SIAM J. Num. Anal. 34, 1997, 1072–1092.

[BF] F. Brezzi, M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer, 1991.

[CTU] C. Canuto, A. Tabacco, K. Urban, *The wavelet element method, part I: Construction and analysis*, Appl. Comput. Harm. Anal. 6, 1999, 1–52.

[Ci] P. Ciarlet, *Introduction to Numerical Linear Algebra and Optimization*, Cambridge, 1989.

[Co] A. Cohen, *Numerical Analysis of Wavelet Methods*, Handbook of Numerical Analysis II, vol. 8, P.G. Ciarlet, J.L. Lions (eds.), Elsevier Science Publishers, 1998.

[CDD1] A. Cohen, W. Dahmen, R. DeVore, *Adaptive wavelet methods for elliptic operator equations — Convergence rates*, Math. Comp. 70, 2001, 27–75.

[CDD2] A. Cohen, W. Dahmen, R. DeVore, *Adaptive wavelet methods II — Beyond the elliptic case*, IGPM–Preprint #199, RWTH Aachen, November 2000.

[DDHS] S. Dahlke, W. Dahmen, R. Hochmuth, R. Schneider, *Stable multiscale bases and local error estimation for elliptic problems*, Appl. Numer. Maths. 8, 1997, 21–47.

[DHU] S. Dahlke, R. Hochmuth, K. Urban, *Adaptive wavelet methods for saddle point problems*, M^2AN 34, 2000, 1003–1022.

[D1] W. Dahmen, *Stability of multiscale transformations*, J. Four. Anal. Appl. 2, 1996, 341–361.

[D2] W. Dahmen, *Wavelet and multiscale methods for operator equations*, Acta Numerica 1997, 55–228.

[D3] W. Dahmen, *Wavelet methods for PDEs — Some recent developments*, IGPM–Preprint #183, RWTH Aachen, December 1999.

[DDU] S. Dahlke, W. Dahmen, K. Urban, *Adaptive wavelet methods for saddle point problems — Convergence rates*, in preparation.

[DK] W. Dahmen, A. Kunoth, *Appending boundary conditions by Lagrange multipliers: Analysis the LBB condition*, revised, December 1999, to appear in Numer. Math. (Numer. Math. Online First DOI 10.1007/s002110000223).

[DKS] W. Dahmen, A. Kunoth, R. Schneider, *Wavelet least square methods for boundary value problems*, IGPM–Preprint #175, RWTH Aachen, September 1999, submitted for publication.

[DKU1] W. Dahmen, A. Kunoth, K. Urban, *Biorthogonal spline-wavelets on the interval – Stability and moment conditions*, Appl. Comput. Harm. Anal. 6, 1999, 132–196.

[DKU2] W. Dahmen, A. Kunoth, K. Urban, *Wavelets in numerical analysis and their quantitative properties*, in: Surface Fitting and Multiresolution Methods, A. Le Méhauté et al.(eds.), Vanderbilt University Press, Nashville, TN, 1997, 93–130.

[DS1] W. Dahmen, R. Schneider, *Composite wavelet bases for operator equations*, Math. Comp. 68, 1999, 1533–1567.

[DS2] W. Dahmen, R. Schneider, *Wavelets on manifolds I: Construction and domain decomposition*, SIAM J. Math. Anal. 31, 1999, 184–230.

[DSX] W. Dahmen, R. Schneider, Y. Xu, *Nonlinear functionals of wavelet expansions — Adaptive reconstruction and fast evaluation*, IGPM-Report #160, June 1998, RWTH Aachen, to appear in Numer. Math.

[DH] J. Dankova, R. Haslinger, *Fictitious domain approach used in shape optimization: Neumann boundary condition*, in: Control of Partial Differential Equations and Applications, Casas, Eduardo (ed.), Dekker, Lect. Notes Pure Appl. Math. 174, 1996, 43–49.

[GJL] Ph.E. Gill, L.O. Jay, M.W. Leonard, L.R. Petzold, V. Sharma, *An SQP method for the optimal control of large-scale dynamical systems*, J. Comput. Appl. Math. 120, 2000, 197–213.

[GMPS] Ph.E. Gill, W. Murray, D.B. Ponecon, M. Saunders, *Preconditioners for indefinite systems arising in optimization*, SIAM J. Matrix Anal. Appl. 13, 1992, 292–311.

[GR] V. Girault, P.-A. Raviart, *Finite Element Methods for Navier-Stokes Equations*, Springer, 1986.

[Glo] R. Glowinski, *Finite element methods for the numerical simulation of incompressible viscous flow. Introduction to the control of the Navier-Stokes equations*, Lect. Appl. Math. 28, 1991, 219–301.

[GPP] R. Glowinski, T.W. Pan, J. Periaux, *A fictitious domain method for Dirichlet problem and application*, Comp. Meth. Appl. Mechs. Eng. 111, 1994, 282–303.

[GoY] G.H. Golub, Q. Ye, *Inexact preconditioned conjugate gradient method with inner–outer iteration*, SIAM J. Sci. Comput. 21, 1999, 1305–1320.

[GKSch] M. Griebel, A. Kunoth, K. Scherer, private communication.

[Gr] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Pitman, 1985.

[Gu] M.D. Gunzburger (ed.), *Flow Control*, The IMA Volumes in Mathematics and its Applications 68, Springer, 1995.

[GHS1] M.D. Gunzburger, L.S. Hou, Th. P. Svobodny, *Analysis and finite element approximation of optimal control problems for the stationary Navier–Stokes equations with Dirichlet controls*, M^2AN 25, 1991, 711–748.

[GHS2] M.D. Gunzburger, L.S. Hou, Th. P. Svobodny, *Analysis and finite element approximation of optimal control problems for the stationary Navier-Stokes equations with distributed and Neumann controls*, Math. Comp. 57, 1991, 123–151.

[GL1] M. D. Gunzburger, H. C. Lee, *Analysis, approximation, and computation of a coupled solid/fluid temperature control problem*, Comp. Meth. Appl. Mech. Engrg. 118, 1994, 133–152.

[GL2] M. D. Gunzburger, H. C. Lee, *Analysis and approximation of optimal control problems for first-order elliptic systems in three dimensions*, Appl. Math. Comput. 100, 1999, 49–70.

[Ha1] W. Hackbusch, *Fast solution of elliptic control problems*, J. Optim. Theory Appl. 31, 1980, 565–581.

[Ha2] W. Hackbusch, *Iterative Solution of Large Sparse Systems of Equations*, Springer, New York, 1994.

[Has] J. Haslinger, *Fictitious domain approaches in shape optimization*, in: Borggard, Jeff (ed.) et al., Computational Methods for Optimal Design and Control, Birkhaeuser, Prog. Syst. Control Theory 24, 1998, 237–248.

[HHK] J. Haslinger, K.–H. Hoffmann, M. Kočvara, *Control fictitious domain method for solving optimal shape design problems*, RAIRO, Modelisation Math. Anal. Numer. 27, 1993, 157–182.

[HasN] P. Neittaanmäki, J. Haslinger, *Finite Element Approximation for Optimal Shape, Material and Topology Design*, Wiley, 2nd ed., 1996.

[Hei1] M. Heinkenschloss, *Formulation and analysis of a sequential quadratic programming method for the optimal Dirichlet boundary control of Navier-Stokes flow*, in: Optimal Control: Theory, Algorithms, and Applications H.W. Hager et al.(eds.), Kluwer, Appl. Optim. 15, 1998, 178–203.

[Hei2] M. Heinkenschloss, *The numerical solution of a control problem governed by a phase field model*, Optim. Methods Softw. 7, 1997, 211–263.

[HK1] M. Hinze, K. Kunisch, *On suboptimal control strategies for the Navier-Stokes equations*, ESAIM, Proc. 4, 1998, 181–198.

[HK2] M. Hinze, K. Kunisch, *Newton's method for tracking type control of the instationary Navier–Stokes equations*, Preprint, October 1999.

[IK] K. Ito, K. Kunisch, *Augmented Lagrangian-SQP methods for nonlinear optimal control problems of tracking type*, SIAM J. Contr. Optim. 34, 1996, 874–891.

[Ka] S. Kaczmarz, *Angenäherte Auflösung von Systemen linearer Gleichungen*, Bulletin de l'Academie Polonaise des Sciences et Lettres A35, 1937, 355–357.

[Kf] A. Kauffmann, *Optimal Control of the Solid Fuel Ignition Model*, Ph.D. Thesis, TU Berlin, 1998.

[KP] K. Kunisch, G. Peichl, *Shape optimization for mixed boundary value problems based on an embedding domain method*, Dyn. Contin. Discrete Impulsive Syst. 4, 1998, 439 – 478.

[K1] A. Kunoth, *Multilevel preconditioning — Appending boundary conditions by Lagrange multipliers*, Adv. Comp. Maths. 4, 1995, 145–170.

[K2] A. Kunoth, *Wavelet Methods — Boundary Value Problems And Control Problems*, Advances in Numerical Mathematics, Teubner, 2001, in Print.

[K3] A. Kunoth, *Wavelet techniques for the fictitious domain—Lagrange multiplier approach*, Manuscript, August 2000, submitted for publication.

[K4] A. Kunoth, *Iterative wavelet–based methods for nonlinear problems in optimal control*, in preparation.

[KK] A. Kunoth, A. Kurdila, *Multilevel techniques for saddle point problems in optimal control*, in preparation.

[Li] J.L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer, Berlin, 1971.

[MM] H. Mauer, H.D. Mittelmann, *Optimization techniques for solving elliptic control problems with control and state constraints: Part 1. Boundary Control*, Comput. Optim. Appl. 16, 29–55, 2000.

[NT] P. Neittaanmäki, D. Tiba, *An embedding of domains approach in free boundary problems and optimal design*, SIAM J. Contr. Optim. 33, 1995, 1587–1602.

[PRG] L. Petzold, J.B. Rosen, Ph.E. Gill, L.O. Jay, K. Park, *Numerical optimal control of parabolic PDEs using DASOPT* L.T. Biegler et al. (eds.), Large-Scale Optimization with Applications. Part 2: Optimal Design and Control, Springer, IMA Vol. Math. Appl. 93, 1997, 271–299.

[Pi] J. Pitkäranta, *Boundary subspaces for the finite element method with Lagrange multipliers*, Numer. Math. 33, 1979, 273–289.

[Ri] A. Rieder, *A domain embedding method for Dirichlet problems in arbitrary space dimensions*, RAIRO, Modelisation Math. Anal. Numer. 32, 1998, pp. 405–431.

[Toi] J. Toivanen, *Ficititious Domain Method Applied to Shape Optimization*, Ph.D. Thesis, University of Jyväskylä, 1997.

[Tr] F. Tröltzsch, *On the Lagrange–Newton–SQP method for the optimal control of semilinear parabolic equations*, SIAM J. Cont. Optim. 38, 1999, 294–312.

[Z] E. Zeidler, *Nonlinear Functional Analysis and its Applications; III: Variational Methods and Optimization*, Springer, 1985.