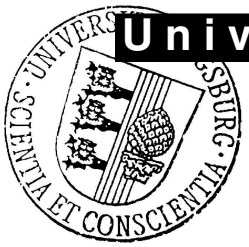

Universität Augsburg

Institut für
Mathematik

Christopher Linsenmann

On the Convergence of Right Transforming Iterations for the Numerical Solution of PDE Constrained Optimization Problems

Preprint Nr. 34/2009 — 21. Dezember 2009
Institut für Mathematik, Universitätsstraße, D-86135 Augsburg http://www.math.uni-augsburg.de/

Impressum:

Herausgeber:

Institut für Mathematik
Universität Augsburg
86135 Augsburg
`http://www.math.uni-augsburg.de/pages/de/forschung/preprints.shtml`

ViSdP:

Christopher Linsenmann
Institut für Mathematik
Universität Augsburg
86135 Augsburg

NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2000; **00**:1–16 *Prepared using* nlaauth.cls *[Version: 2002/09/18 v1.02]*

On the convergence of right transforming iterations for the numerical solution of PDE constrained optimization problems

Christopher Linsenmann*,†

Institute for Mathematics, University of Augsburg, Universitätsstr.14, D-86159 Augsburg, Germany

SUMMARY

We present an iterative solver, called right transforming iterations (or *right transformations*), for linear systems with a certain structure in the system matrix, such as they typically arise in the framework of KKT conditions for optimization problems under PDE constraints. The construction of the right transforming scheme depends on an inner approximate solver for the underlying PDE subproblems. We give a rigorous convergence proof for the right transforming iterative scheme in dependence on the convergence properties of the inner solver. Provided that a fast subsolver is available, this iterative scheme represents an efficient way of solving first order optimality conditions.
Numerical examples endorse the theoretically predicted contraction rates. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: right transforming iterations; iterative KKT solver; optimization problems with PDE constraints; perturbed splitting methods

1. INTRODUCTION

The numerical solution of PDE constrained optimization problems based on the first order optimality conditions typically involves linear systems of the form

$$Kx = \begin{pmatrix} Q & S^T & B_1^T \\ S & 0 & B_2^T \\ B_1 & B_2 & C \end{pmatrix} \begin{pmatrix} y \\ \lambda \\ u \end{pmatrix} = \begin{pmatrix} b_y \\ b_\lambda \\ b_u \end{pmatrix} = b, \tag{1}$$

where $Q, S \in \mathbb{R}^{n\times n}$, B_1, $B_2 \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{m\times m}$. In case of linear-quadratic minimization problems, such a system arises directly from the KKT conditions, whereas for nonlinear problems it stems from a Newton linearization thereof. Here, $y \in \mathbb{R}^n$ stands for the state,

*Correspondence to: Christopher Linsenmann, Institute for Mathematics, University of Augsburg, Universitätsstr.14, D-86159 Augsburg, Germany
†E-mail: linsenmann@math.uni-augsburg.de

Contract/grant sponsor: DFG SPP 1253, NSF; contract/grant number: DMS-0511611

Received 2009
Revised

$\lambda \in \mathbb{R}^n$ for the Lagrange multiplier and $u \in \mathbb{R}^m$ for the design parameter/control, and S denotes the matrix associated with the discretized partial differential operator.

The idea of right transformations is based on an appropriate 'right matrix' K_R, which is multiplied from the right to K and then gives rise to a regular splitting $KK_R = M_1 - M_2$, where due to a block-diagonal structure M_1 is relatively easy to invert. The availability of K_R depends on the properties of the submatrices of K. In practice it turns out that the computation of such a matrix K_R is too costly and solving the system $M_1 K_R^{-1} x = b$ equivalent to (1) would require the same effort as solving (1) directly by means of a Schur complement formulation. Indeed, the right transforming iterative scheme can be seen as an approximate Schur complement-based method. The benefit of the right transforming formulation, however, is the possibility of analyzing the spectral properties of the associated iteration matrix. Thus, one replaces the 'exact' matrix K_R by an approximate version $K_R(k)$, $k \in \mathbb{N}$, indicating that this matrix is constructed by means of k applications of an iterative subsolver (for details see Section 2 and Assumption 3.2). This leads to a perturbed splitting scheme with $M_1(k) \approx M_1$ and $M_2(k) \approx 0$. We will show that the thus induced iterative scheme does converge to the solution $x = K^{-1}b$, depending on the 'quality' of approximation indicated by k, and we will derive its rate of contraction and convergence.

Although the use of matrix transformations is a basic tool in numerical linear algebra (see for example [9, 10] for an application to KKT systems), the term 'right transforming iterations' goes back to Wittum, who has developed a right transforming iterative scheme serving as a smoother for the multigrid solution of Stokes and Navier-Stokes equations [22, 23]. This approach has recently been extended to transforming smoothers for PDE constrained optimization problems in [19]. In [1, 2, 3, 4, 13, 14, 15, 16], right transforming iterations have been successfully applied to the solution of shape and topology optimization problems, however, without a convergence proof.

The paper is organized as follows: In Section 2 we derive the right transforming iterative scheme and establish assumptions that guarantee its applicability. In Section 3 we present a convergence proof for the right transforming scheme and derive an estimate for the contraction rate in dependence of the spectral radii of the iteration matrices associated with the PDE subsolvers. This convergence result will be illustrated by numerical experiments in Section 4. In Section 5 we give a pseudo-algorithmic code of the right transforming scheme and, after discussing its limitations, present an iterative version of it in Section 6.

2. RIGHT TRANSFORMING ITERATIONS

2.1. Preliminaries

We consider a linear system of the form

$$\widetilde{\mathcal{K}} z = \begin{pmatrix} \mathcal{A} & \mathcal{B}^T \\ \mathcal{B} & 0 \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix}, \tag{2}$$

where $\mathcal{A} \in \mathbb{R}^{(n+m)\times(n+m)}$ and $\mathcal{B} \in \mathbb{R}^{n\times(n+m)}$ are such that $\widetilde{\mathcal{K}} \in \mathbb{R}^{(2n+m)\times(2n+m)}$ is regular, and v, $c \in \mathbb{R}^{n+m}$, w, $d \in \mathbb{R}^n$.

Among direct methods for the solution of (2) there are symmetric factorizations as well as the range space and the null space approach (see [11]). The range space approach is based on the Schur complement formulation of (2) and thus requires positive definiteness of $\mathcal{A}$. In contrast, the null space approach does not assume regularity of $\mathcal{A}$. Both methods use the special structure of the KKT matrix $\widetilde{\mathcal{K}}$.
The distinction between the range space and the null space approach can be extended to the class of iterative solvers, see [18, 20, 22]. As for the direct solver, the iterative range space approach requires regularity of $\mathcal{A}$ and works with a transforming matrix $\widetilde{K}_L$ 'from the left' (range space transformation). For the iterative analogue of the null space approach, we have a right transforming matrix which thus transforms the nullspace of the matrix. We will refer to the associated solver in the following as RT. A main ingredient thereof is to rearrange the matrix

$$\widetilde{\mathcal{K}} = \begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} & \mathcal{B}_1^T \\ \mathcal{A}_{21} & \mathcal{A}_{22} & \mathcal{B}_2^T \\ \mathcal{B}_1 & \mathcal{B}_2 & 0 \end{pmatrix} \in \mathbb{R}^{(n+m+n)\times(n+m+n)}$$

with regular $\mathcal{B}_1 \in \mathbb{R}^{n\times n}$ according to

$$\mathcal{K} := \begin{pmatrix} \mathcal{A}_{11} & \mathcal{B}_1^T & \mathcal{A}_{12} \\ \mathcal{B}_1 & 0 & \mathcal{B}_2^T \\ \mathcal{A}_{21} & \mathcal{B}_2 & \mathcal{A}_{22} \end{pmatrix}.$$

By identifying

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{pmatrix} = \begin{pmatrix} Q & B_1^T \\ B_1 & C \end{pmatrix}$$

and

$$\mathcal{B} = \begin{pmatrix} \mathcal{B}_1 & \mathcal{B}_2 \end{pmatrix} = \begin{pmatrix} S & B_2 \end{pmatrix},$$

we see that the structure of $\mathcal{K}$ fits the one of K in (1), and that the assumptions for (2) are met, if K and S are invertible. Just to keep track of all necessary assumptions, we state

Assumption 2.1. *The system matrix $K \in \mathbb{R}^{(2n+m)\times(2n+m)}$ is of the form as in (1) and regular, and the matrix $S \in \mathbb{R}^{n\times n}$ is regular as well. (Symmetry of $Q \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{m\times m}$ is not necessary, but 'helpful', cf. Section 6.)* ◇

We set

$$A := \begin{pmatrix} Q & S^T \\ S & 0 \end{pmatrix} \in \mathbb{R}^{2n\times 2n} \quad \text{and} \quad B := \begin{pmatrix} B_1 & B_2 \end{pmatrix} \in \mathbb{R}^{m\times 2n},$$

so that

$$K = \begin{pmatrix} A & B^T \\ B & C \end{pmatrix},$$

cf. (1). We note that A is indefinite but nonsingular, and relatively easy to invert if this holds true for S and S^T.

We assume that there exist fast iterative solvers for the generic subproblems

$$S\,z = f \quad \text{and} \quad S^T z = f, \quad z, f \in \mathbb{R}^n \tag{3}$$

(which correspond to a discretized PDE system and its adjoint system in the PDE-constrained optimization context). We formally set

$$A(k) := \begin{pmatrix} Q & S^T(k) \\ S(k) & 0 \end{pmatrix} \tag{4}$$

and note that its inverse is given by

$$A(k)^{-1} = \begin{pmatrix} 0 & S(k)^{-1} \\ S^T(k)^{-1} & -S^T(k)^{-1} Q\, S(k)^{-1} \end{pmatrix}. \tag{5}$$

Here and in the sequel, matrices and vectors followed by a (k) are seen as *iteratively obtained approximations*, where the quality of approximation depends on the iteration count k. Also, expressions like $S(k)$ or $S(k)^{-1}$ have to be understood in a formal sense: In practice, $S(k)^{-1} f := z(k)$ means the application of an iterative solver with k iterations to the system $Sz = f$, where $z(k)$ is the k-th iterate.

As will be pointed out below, the RT iterative scheme mainly requires application of $A(k)^{-1}$. In view of (5), we therefore need the approximate inverse of S and of its transpose, but not of Q; in fact, there are no regularity requirements on Q. This is a crucial feature of the RT scheme: The solution of system (1) is carried out by repeatedly applying PDE solvers (which are, in general, already available and well investigated). In this sense, it is a natural approach and involves only a few regularity requirements.

We further define the perturbed/approximate *right transform* $K_R(k)$ by

$$K_R(k) := \begin{pmatrix} I_{2n} & -A(k)^{-1} B^T \\ 0 & I_m \end{pmatrix}.$$

Obviously, $K_R(k)$ is nonsingular. The product $KK_R(k)$ gives rise to the regular splitting

$$\begin{aligned} K\, K_R(k) &= \begin{pmatrix} A & B^T \\ B & C \end{pmatrix} \begin{pmatrix} I_{2n} & -A(k)^{-1} B^T \\ 0 & I_m \end{pmatrix} \\ &= \underbrace{\begin{pmatrix} A(k) & 0 \\ B & C - BA(k)^{-1} B^T \end{pmatrix}}_{=: M_1(k)} - \\ &\quad \underbrace{\begin{pmatrix} (I_{2n} - AA(k)^{-1}) A(k) & -(I_{2n} - AA(k)^{-1}) B^T \\ 0 & 0 \end{pmatrix}}_{=: M_2(k)}. \end{aligned}$$

It can be easily seen that there holds $M_2(k) \approx 0$, if $AA(k)^{-1} \approx I_{2n}$.

2.2. Iterative scheme

The idea of RT is as follows: The system $Kx = b$ with $b \in \mathbb{R}^{2n+m}$ is equivalent to

$$\underbrace{KK_R(k)}_{\approx M_1(k)} \underbrace{K_R(k)^{-1} x}_{=: \tilde{x}(k)} = b.$$

Hence, $M_1(k)$ can be used as a preconditioner for the right transformed system $KK_R(k)\tilde{x}(k) = b$. Given a start iterate $\tilde{x}^{(0)} = K_R(k)^{-1}x^{(0)}$, the corresponding stationary iterative scheme reads

$$\tilde{x}^{(i+1)}(k) = \tilde{x}^{(i)}(k) + M_1(k)^{-1}\big(b - KK_R(k)\tilde{x}^{(i)}(k)\big)\ , \quad i \geq 0.$$

Multiplication from the left by $K_R(k)$ yields

$$x^{(i+1)}(k) = x^{(i)}(k) + K_R(k)M_1(k)^{-1}\big(b - Kx^{(i)}(k)\big)\ , \quad i \geq 0 \tag{6}$$

which is the *RT iteration scheme*, where $K_R(k)M_1(k)^{-1}$ serves as a preconditioner for K in a Richardson-type iteration scheme. It is easy to see that one iteration in (6) can be subdivided into three steps:

(i) Compute the residual $\xi := b - Kx^{(i)}(k)$.
(ii) Solve $M_1(k)z = \xi$. Due to the structure of $M_1(k)$, this mainly requires
   (a) the solution of the system $\xi_1 = A(k)^{-1}r_1$, where z and ξ are partitioned according to $(z_1, z_2)^T$, $(\xi_1, \xi_2)^T \in \mathbb{R}^{2n+m}$, and
   (b) computation of
   $$G(k) := A(k)^{-1}B^T \in \mathbb{R}^{n\times m}.$$
   This means that the approximate inverse has to be applied m times, namely, to each column of B^T.
   Finally, the approximate Schur complement
   $$D(k) := C - BA(k)^{-1}B^T = C - B\,G(k) \in \mathbb{R}^{m\times m}$$
   has to be computed
   (c) and thereafter, $D(k)z_2 = \xi_2 - B\xi_1$ has to be solved exactly. This can be easily done, if m is small. Altogether, step (ii) demands $(m+1)$ applications of the approximate inverse $A(k)^{-1}$.
(iii) Compute the increment $w := K_R(k)z$ and update $x^{(i+1)}(k) = x^{(i)}(k) + w$. For the computation of w, the result from step (ii)(b) can be used: $(w_1, w_2)^T = (z_1 - G(k)z_2, z_2)^T$.

Of course, in an algorithmic realization, the computationally most expensive part, step (ii)(b), will be done only once in a pre-processing step, since the matrices $G(k)$ and $D(k)$ will not change in later iterations.
Let us have a closer look at the formal expression $A(k)^{-1}\xi_1 = z_1$ from step (ii)(a): Due to the block structure of $A(k)$ (see (4)), this basically means that the iterative solvers for (3) have to be employed.

Remark 2.2. *If exact solvers for (3) are used, then* $A(\cdot)^{-1} = A^{-1}$ *and*

$$KK_R(\cdot) = M_1(\cdot) = \begin{pmatrix} A & 0 \\ B & C - BA^{-1}B^T \end{pmatrix}.$$

The regularity of $D := C - BA^{-1}B^T$ *follows from the regularity of* K *and* K_R*. For a sufficiently accurate approximation* $A(k)^{-1} \approx A^{-1}$*, we see that* $D(k)$ *is regular, whence step (ii)(c) is well-defined.* ◇

Straightforward computation reveals that the iteration matrix $I_{2n+m} - K_R(k)\ M_1(k)^{-1}K$ associated with the iterative scheme (6) is given by

$$L_{RT}(k) := \begin{pmatrix} \big(I_{2n} + A(k)^{-1}B^T D(k)^{-1}B\big)\big(I_{2n} - A(k)^{-1}A\big) & 0 \\ -D(k)^{-1}B\big(I_{2n} - A(k)^{-1}A\big) & 0 \end{pmatrix}. \tag{7}$$

For a sufficiently accurate approximation $A(k)^{-1} \approx A^{-1}$, we expect convergence of the right transforming scheme (6), since then $L_{RT}(k) \approx 0$. We will make this statement rigorous by estimating the spectral radius of $L_{RT}(k)$, which is the content of the next section.

3. CONVERGENCE RESULT FOR RIGHT TRANSFORMATIONS

In the following, we use the abbreviation

$$H(k) := \big(I_{2n} + A(k)^{-1}B^T D(k)^{-1}B\big)\big(I_{2n} - A(k)^{-1}A\big), \tag{8}$$

where $H(k)$ stands for the first upper block in $L_{RT}(k)$.

Proposition 3.1. *With $L_{RT}(k)$ from (7) and $H(k)$ from (8), there holds*

$$\varrho\big(L_{RT}(k)\big) = \varrho\big(H(k)\big),$$

where $\varrho(\cdot)$ denotes the spectral radius.

Proof. If $(\lambda, (v,w)^T)$ is an eigenpair of $L_{RT}(k)$ and $\lambda \neq 0$, then $v \neq 0$ and therefore, (λ, v) is an eigenpair of $H(k)$.
Conversely, if (λ, v) is an eigenpair of $H(k)$ with $\lambda \neq 0$, then $(\lambda, (v, -1/\lambda D(k)^{-1}B(I_{2n} - A(k)^{-1}A)v)^T)$ is an eigenpair of L_{RT}. Thus, $\varrho(L_{RT}(k)) = \varrho(H(k))$ if either $\varrho(\cdot) > 0$.
If $\varrho(L_{RT}(k)) = 0$, then by contraposition also $\varrho(H(k)) = 0$ and vice versa. □

Hence, it is sufficient to estimate the spectral radius of the matrix $H(k)$. We now specify the requirements on the iterative solver(s) for the subsystems (3):

Assumption 3.2. *There exists a consistent and convergent stationary iterative solver for the basic subproblem $Sz = f$ with associated iteration matrix $\mathfrak{S} \equiv \mathfrak{S}(k)$ (i.e., $z(k+1) = \mathfrak{S}z(k) + Nf$, $k \geq 0$, with $\mathfrak{S} = I_n - NS$) and an upper bound $r < 1$ for its spectral radius, i.e.,*

$$\varrho(\mathfrak{S}) \leq r < 1. \qquad \diamond$$

Remark 3.3. *It is clear that an iterative solver for the problem $S^T z = f$ is given by $z(k+1) = z(k) + N^T(f - S^T z(k))$, $k \geq 0$, with iteration matrix $\mathfrak{S}_T = N^T \mathfrak{S}^T N^{-T}$ (Assumptions 2.1 and 3.2 imply that N is invertible) fulfilling $\varrho(\mathfrak{S}_T) = \varrho(N^{-1}\mathfrak{S}N) = \varrho(\mathfrak{S})$, the first and second equality following from the facts that a matrix and its transpose resp. a matrix similar to it possess the same eigenvalues. Thus, under Assumptions 2.1 and 3.2 we have $\varrho(\mathfrak{S}_T) \leq r < 1$.* $\diamond$

Let us now state the main result of the paper.

Theorem 3.4. Convergence of the RT iteration scheme
Let Assumptions 2.1 and 3.2 hold true and assume that every subproblem (3) is solved by at least k iterations, $k \geq 1$, with initial guesses $z^{(0)} = 0$.
Then, for any $0 < \varepsilon < 1 - r$ and sufficiently large k, for the spectral radius of $L_{RT}(k)$ there holds

$$\varrho(L_{RT}(k)) \leq C_0(\varepsilon, k_0)\frac{(r+\varepsilon)^k}{1-r-\varepsilon} \ . \tag{9}$$

Hence, $\varrho(L_{RT}(k)) < 1$ for k large enough, i.e., scheme (6) is convergent for any right hand side b. Here, C_0 is a constant depending on ε (nondecreasing as ε decreases).

Proof. According to Proposition 3.1, we need to estimate $\varrho(H(k))$. For more clarity, we divide the proof into six steps:

(i) We recall some useful well-known facts from numerical linear algebra. Here, A stands for an arbitrary matrix in $\mathbb{R}^{N\times N}$ (or $\mathbb{C}^{N\times N}$), $N \in \mathbb{N}$.

(a) For given A and $\varepsilon > 0$, there exists a matrix norm $\|\cdot\|_*$ with $\|A\|_* \leq \varrho(A) + \varepsilon$ (see for example [17], Lemma 5.6.10).
(b) If V, $W \in \mathbb{C}^{N\times N}$ are nonsingular and $\|\cdot\|$ is a submultiplicative matrix norm, then $\|V^{-1} \cdot W\|$ defines a matrix norm on $\mathbb{C}^{N\times N}$ which is *pseudo-submultiplicative* in the sense

$$\|V^{-1}ABW\| \leq c\,\|V^{-1}AW\|\|V^{-1}BW\| \tag{10}$$

with $c = \|W^{-1}V\|$.
(c) $\varrho(A) \leq \|A\|$ for any submultiplicative matrix norm $\|\cdot\|$ (not only for those induced by vector norms), cf. [17], Theorem 5.6.9. If $\|\cdot\|$ is pseudo-submultiplicative with constant c, then $\varrho(A) \leq c\|A\|$.
(d) If $\|A\| < 1$ for some submultiplicative matrix norm, then there holds $(I_N - A)^{-1} = \sum_{i=0}^{\infty} A^i$ ([17], Corollary 5.6.16).

(ii) From Assumption 3.2 and Remark 3.3 it follows that $S(k)^{-1}f \to S^{-1}f$ and $S^T(k)^{-1}f \to S^{-T}f$, $f \in \mathbb{R}^n$, as $k \to \infty$. This implies that $A(k)$ and $D(k)^{-1}$ (cf. Remark 2.2) as well as $S(k)$, $S^T(k)$ exist for k large enough. For such k we estimate

$$\begin{aligned}\varrho(H(k)) &\leq c\,\|H(k)\|_\diamond \\ &= c\,\|A(k)^{-1}(A(k)-A) + A(k)^{-1}B^TD(k)^{-1}BA(k)^{-1}(A(k)-A)\|_\diamond \\ &\leq c^2\,\|A(k)^{-1} + A(k)^{-1}B^TD(k)^{-1}BA(k)^{-1}\|_\diamond\,\|A(k)-A\|_\diamond \end{aligned} \tag{11}$$

with a pseudo-submultiplicative matrix norm $\|\cdot\|_\diamond : \mathbb{C}^{2n\times 2n} \to \mathbb{R}$ to be specified below, likewise for the constant c. Note that

$$A(k) - A = \begin{pmatrix} 0 & S^T(k) - S^T \\ S(k) - S & 0 \end{pmatrix}$$

and that

$$\begin{aligned}&\|A(k)^{-1} + A(k)^{-1}B^TD(k)^{-1}BA(k)^{-1}\|_\diamond \leq \\ &\|A^{-1} + A^{-1}B^TD^{-1}BA^{-1}\|_\diamond + \widehat{C}(k_0) =: C_1(k_0, K, \mathfrak{S})\end{aligned} \tag{12}$$

due to the continuity of the norm. For the latter estimate it is important that the norm $\|\cdot\|_\diamond$ does not depend on k (cf. (13) and (15) below). We could stop at this point of the proof, if we were just interested in a convergence result. We just had to apply the same continuity argument to the second term in (11) and were done. However, since we are also interested in the rate of contraction, we have to estimate the expression $\|A(k)-A\|_\diamond$ appropriately to obtain a dependency on r.

(iii) Definition of appropriate norms:
Choose $0<\varepsilon<1-r$, which gives

$$R=R(\varepsilon):=\varrho(\mathfrak{S})+\varepsilon=\varrho(\mathfrak{S}_{\mathcal{T}})+\varepsilon<1$$

(with $\mathfrak{S}$, $\mathfrak{S}_{\mathcal{T}}$ as in Assumption 3.2 and Remark 3.3). Using (i)(a), there exist matrix norms $\|\cdot\|_{*,1}:\mathbb{C}^n\to\mathbb{R}$ and $\|\cdot\|_{*,2}:\mathbb{C}^n\to\mathbb{R}$ such that

$$\|\mathfrak{S}\|_{*,1}\le R<1 \quad\text{and}\quad \|\mathfrak{S}_{\mathcal{T}}\|_{*,2}\le R<1. \tag{13}$$

Note that $\|\cdot\|_{*,j}$, $1\le j\le 2$, only depend on $\mathfrak{S}$, $\mathfrak{S}_{\mathcal{T}}$, n, ε, but not on k. A closer look at Lemma 5.6.10 in [17] reveals that $\|\cdot\|_{*,j}$ is constructed according to

$$\|A\|_{*,j}:=\|V_j^{-1}AV_j\|_1\ ,\quad A\in\mathbb{C}^{n\times n} \tag{14}$$

with nonsingular matrices V_j of the form $V_j=U_jD(t_j)^{-1}\in\mathbb{C}^{n\times n}$, U_j being unitary, $D(t_j)=\operatorname{diag}(t_j,\dots,t_j^n)$ and $t_j\in\mathbb{R}_+$ chosen large enough. Now, for $\mathcal{C}=(\mathcal{C}_{ij})_{i,j=1}^2\in\mathbb{C}^{2n\times 2n}$ we define

$$\|\mathcal{C}\|_\diamond:=\left\|\begin{pmatrix} V_2^{-1} & 0\\ 0 & V_1^{-1}\end{pmatrix}\begin{pmatrix}\mathcal{C}_{11}&\mathcal{C}_{12}\\ \mathcal{C}_{21}&\mathcal{C}_{22}\end{pmatrix}\begin{pmatrix} V_1 & 0\\ 0 & V_2\end{pmatrix}\right\|_1. \tag{15}$$

Here, the matrices V_j, $1\le j\le 2$, stem from the norms $\|\cdot\|_{*,j}$ as in (13). It is straightforward to check that this defines a matrix norm on $\mathbb{C}^{2n\times 2n}$ which is pseudo-submultiplicative (cf. (10)) with constant

$$c:=\left\|\begin{pmatrix} V_1^{-1}V_2 & 0\\ 0 & V_2^{-1}V_1\end{pmatrix}\right\|_1$$

as in (11) and independent of k. It follows that

$$\|A(k)-A\|_\diamond\le\left\|\begin{pmatrix}0&0\\ S(k)-S&0\end{pmatrix}\right\|_\diamond+\left\|\begin{pmatrix}0&S^T(k)-S\\ 0&0\end{pmatrix}\right\|_\diamond.$$

By construction and due to the 1-norm used in definition (15), we get

$$\|A(k)-A\|_\diamond\le\|S(k)-S\|_{*,1}+\|S^T(k)-S^T\|_{*,2}\,. \tag{16}$$

This step is crucial to break down the problem from $A(k)-A$ to its submatrices $S(k)-S$ resp. $S^T(k)-S^T$ in appropriate norms.

(iv) Representation of $S(k)-S$:
Pick an arbitrary right hand side f for the problem $Sz=f$ with unique solution $z=S^{-1}f$. Denoting by $e(k)=z(k)-z$ the error associated with the k-th iterate, $k\ge 0$, for a stationary iterative method, we get $e(k)=\mathfrak{S}^k e(0)$ and consequently, for $z(0):=0$

$$S(k)^{-1}f\overset{\text{def}}{=}z(k)=\big(I_n-\mathfrak{S}^k\big)z=\big(I_n-\mathfrak{S}^k\big)S^{-1}f$$

for all f and, hence, for sufficiently large k there holds

$$S(k) = S\big(I_n - \mathfrak{S}^k\big)^{-1}.$$

Exploiting $\|\mathfrak{S}\|_{*,1} < 1$ from (13), observing $\|\mathfrak{S}^k\|_{*,1} < 1$ due to submultiplicativity of $\|\cdot\|_{*,1}$ and using (i)(d), we obtain

$$S(k) = S\sum_{i=0}^{\infty}(\mathfrak{S}^k)^i.$$

Consequently,

$$S(k) - S = S\big(\sum_{i=0}^{\infty}(\mathfrak{S}^k)^i - I_n\big) = S\sum_{i=1}^{\infty}(\mathfrak{S}^k)^i. \tag{17}$$

(v) Norm estimate for $S(k) - S$:
We want to estimate the first term of the right hand side in (16). We set $s := \|\mathfrak{S}^k\|_{*,1} < 1$. Due to (17), we have

$$\begin{aligned}\|S(k) - S\|_{*,1} &\le \|S\|_{*,1}\sum_{i=1}^{\infty} s^i = \|S\|_{*,1}\big(\sum_{i=0}^{\infty} s^i - 1\big)\\ &= \|S\|_{*,1}\frac{s}{1-s} \le \|S\|_{*,1}\frac{\|\mathfrak{S}\|_{*,1}^k}{1-\|\mathfrak{S}\|_{*,1}} \le \|S\|_{*,1}\frac{(r+\varepsilon)^k}{1-r-\varepsilon}\ ,\end{aligned} \tag{18}$$

where we have used (13).

(vi) Conclusion:
Steps (iv)-(v) can be repeated for $\|S^T(k) - S^T\|_{*,2}$ analogously. Combining Proposition 3.1, (11), (12), (16), and (18), we arrive at

$$\varrho(L_{RT}(k)) = \varrho(H(k)) \le 2\,c^2 C \max(\|S\|_{*,1}, \|S^T\|_{*,2})\frac{(r+\varepsilon)^k}{1-r-\varepsilon}\ .$$

For $k \ge k_1 = k_1(\varepsilon)$, we get $\varrho(L_{RT}(k)) < 1$, since $0 < r + \varepsilon < 1$. The dependence of the constant $C_0 := 2\,c^2 C_1 \max(\|S\|_{*,1}, \|S^T\|_{*,2})$ in (9) is due to the dependence of t_j (cf. (iii)) on ε, $1 \le j \le 2$. □

Definition 3.5. *(Contraction rate, convergence rate)*
Let $x^{(i)}$, $i \ge 0$, be the iterates generated by an iterative method and let $x = K^{-1}b$ be the exact solution. By $e^{(i)} := x^{(i)} - x$ we denote the error associated with the i-th iteration. Let the iteration matrix be given by L, with $\varrho(L) < 1$.
We refer to

$$R_i := \|e^{(i)}\|/\|e^{(i-1)}\|\ ,\quad i \ge 1$$

as the i-th rate of contraction *(or* contraction factor *or* factor of reduction for successive error norms*) and, following [12] or [21], to*

$$\overline{R}_i := \frac{-\ln\|L^i\|}{i}$$

as the average rate of convergence for i iterations. *From the known fact that* $\varrho(L) = \lim_{i\to\infty} \|L^i\|^{1/i}$ *for any matrix norm* $\|\cdot\|$, *it can be deduced that*

$$\lim_{i\to\infty} \overline{R}_i = -\ln(\varrho(L)),$$

which is called the asymptotic rate of convergence.
Assuming that i *is large and that* $\|L^{i-1}e^{(0)}\| \approx \|L^{i-1}\|\,\|e^{(0)}\|$, *we further obtain* $R_i \approx \varrho(L)$. *Hence,* $\varrho(L)$ *will be referred to as the* asymptotic rate of contraction. ◇

Corollary 3.6. *Let the assumptions of Theorem 3.4 hold true with a sharp bound* $r = \varrho(\mathfrak{S})$. *For large* i, *the quantity* $R_{RT}(k) = \|x^{(i)}(k) - x\|/\|x^{(i-1)}(k) - x\|$ *is called the contraction rate of the RT iterative scheme (6), with* $x^{(i)}(k)$ *being the* i*-th RT iterate constructed by means of* k *applications of the subsolver(s) as in Assumption 3.2. Then, we obtain an estimated asymptotic contraction rate*

$$R_{RT}(k) \approx \mathcal{O}(r^k) \tag{19}$$

and consequently, the asymptotic convergence rate

$$-ln(\varrho(L_{RT})) \approx k\,ln(1/r) - c.$$ □

4. NUMERICAL EXAMPLES

We consider system (1) stemming from a shape optimization problem subject to box constraints on the design parameters u_i, $1 \le i \le m$, and subject to PDE constraints representing stationary incompressible flow. The design parameters are chosen as Bézier control points in a parametrization of the walls of a channel-like domain. The inequality constraints are treated by an interior point approach with barrier parameter $\mu > 0$, and the (nonlinear) optimality conditions of the associated perturbed optimization subproblems (with associated solutions $x(\mu)$) lead, after applying Newton's method, to a system exactly of the form as in (1). For further details, the reader is referred to [4]. In this case, the matrix S is the Stokes matrix

$$S = \begin{pmatrix} S_A & S_B^T \\ S_B & 0 \end{pmatrix} = S^T$$

arising from a P2-P1 Taylor-Hood discretization of the Stokes equations with Dirichlet *and* Neumann-type boundary conditions resulting in a nonsingular matrix S. As iterative solver, we choose the augmented Lagrangian solver from [6]. It has recently been shown† that this iterative procedure can be written as a stationary preconditioned Richardson scheme. The spectral radius of its associated iteration matrix can be specified as $\varrho(\mathfrak{S}) = 1/(1 + \rho\,\lambda_1)$ (this result is also known from [6]), where $\rho > 0$ (usually $\gg 0$) is the penalty and update parameter of the augmented Lagrangian algorithm and $\lambda_1 = \min\sigma(S_B S_A^{-1} S_B^T) > 0$ is the smallest eigenvalue of the Schur complement associated with the matrix S. Hence, Assumption 3.2 is fulfilled. Moreover, it is known from [7] that under certain regularity assumptions on the local minimizer, the matrix $K = K(\tilde{x}(\mu))$ is regular for μ small enough and a sufficiently good approximtion $\tilde{x}(\mu) \approx x(\mu)$. Then also Assumption 2.1 holds true for our test cases.

†Linsenmann C. *The augmented Lagrangian method as smoother for the multigrid solution of the Stokes equations.* In preparation.

4.1. Test problem 1

As computational domain we choose a channel-like geometry and pick $h := 0.35$ as maximal mesh width, resulting in $n_1 = 938$ velocity nodes and $n_2 = 127$ pressure nodes, so that $n = n_1 + n_2 = 1065$. We further choose $m := 8$ design parameters. It turns out that here $\lambda_1 = 1.85\text{e}{-3}$. Choosing $\rho := 5.41\text{e}{+5}$ for the augmented Lagrangian algorithm, we obtain

$$r := \varrho(\mathfrak{S}) = 0.5.$$

In view of Corollary 3.6, we want to check whether the contraction factors

$$R_i(k) := \frac{\|x^{(i)}(k) - x\|_2}{\|x^{(i-1)}(k) - x\|_2}\ , \quad i \geq 1$$

from RT iteration $(i-1)$ to iteration i satisfy $R_i(k) = \mathcal{O}(r^k)$. Indeed, Figure 1 displays

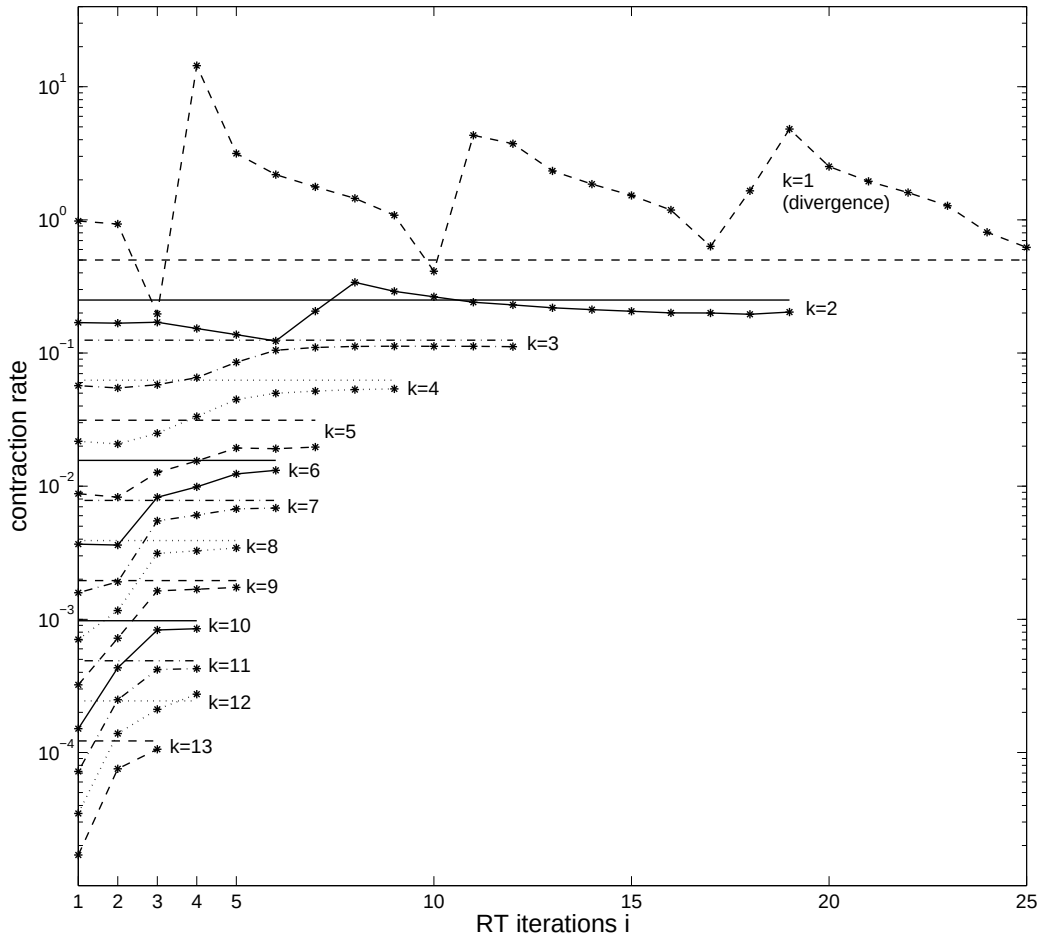

Figure 1. First problem: The observed contraction rates vs. the by r^k predicted asymptotic contraction rates (horizontal lines) for $k = 1, 2, \ldots, 13$ and $r = 0.5$.

the coincidence between the observed rates of contraction and the predicted asymptotic rate (horizontal lines). For $k = 1$ the algorithm does not converge, underlining the requirement that k must be "large enough". For $k \geq 2$ the predicted rate is confirmed impressively, and the asymptotic character of the quantity $R_i(k)$ (cf. Definition 3.5) shows up clearly. The fact that for $k = 2$ and 12 the observed rates are slightly larger than the predicted ones does not contradict estimate (19). It can also be observed that the larger k, the faster $R_i(k)$ attains its asymptotic behavior, which is what we expect: Large values of k lead to a fast damping of eigenvectors associated with small eigenvalues contributing to the iteration error. Note further that in general we *over*estimate the contraction rate.

4.2. Test problem 2

This problem differs from the first one by choosing $h_{\max} := 0.0625$ and $\rho := 1.42\text{e}{+5}$. This results in $\lambda_1 = 6.35\text{e}{-5}$, $r = 0.25$, and $n = 28\,234 + 3\,578 = 31\,812$. Hence, the overall number of unknowns (63 632) is significantly larger compared to the first problem. Also this test case confirms our contraction rate estimate $R_i(k) \approx r^k$, as can be seen in Figure 2. Note that compared to the first example, there is no change in the qualitative behavior of the 'ratio' of

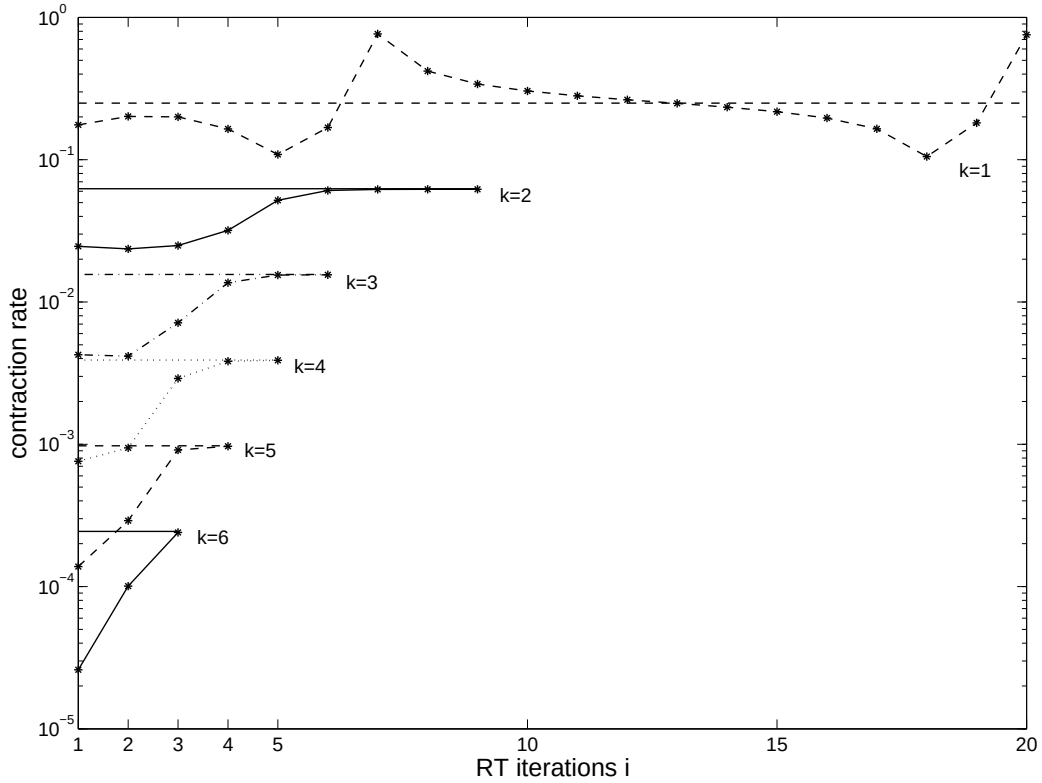

Figure 2. Second test problem: The observed contraction rates vs. the by r^k predicted asymptotic rates (horizontal lines) for $k = 1, \ldots, 6$ and $r = 0.25$.

predicted and observed contraction rates, although the dimension of the problem is significantly larger. Interestingly, for $k = 1$ the algorithm shows the same unusual periodic development of the contraction rate as in the first example (where it occurs for $k = 1$ *and* 2 due to the larger value of $r = 0.5$ compared to 0.25 in test problem two). The typical asymptotic development emerges for $k \geq 2$ and again reveals the behavior 'the larger k, the faster the approximation of the asymptotic contraction rate'.

5. RIGHT TRANSFORMATIONS ALGORITHM (I)

For the following algorithm in pseudo-form, we use the notation known from MATLAB or the standard monograph [8], where exemplarily $A(:, j)$ denotes the j-th column of a matrix A. For simplicity, we drop the (k)-notation.

Note that the computation of G_y, G_λ does not rely on information of previous iterations j, whence the pre-processing step qualifies for parallelization.

Algorithm 5.1. Right transformations for solving the system (1)

% Initialization:
Let the right hand side $b = (b_y, b_\lambda, b_u)^T$ and the matrix K with blocks as in (1) be given.
Specify a tolerance $\epsilon \ll 1$ and set $x^{(0)} := 0$.

% Pre-processing: Compute the matrices G_y, $G_\lambda \in \mathbb{R}^{n \times m}$ columnwise:
`for` $j = 1 : m$
  Solve $S\, G_y(:, j) = B_2(j, :)^T$ iteratively.
  Solve $S^T G_\lambda(:, j) = B_1(j, :)^T - Q\, G_y(:, j)$ iteratively.
`end`

Compute the matrix $D := C - B_1 G_y - B_2\, G_\lambda$.

% Iteration:
`for` $i = 0, \ldots$ until convergence
  Compute the residual $\xi = (\xi_y, \xi_\lambda, \xi_u)^T = b - K x^{(i)}$.
  % Check for convergence:
  `if` $\|\xi\| < \epsilon \, \|b\|$
    Solution is $x := x^{(i)}$, stop.
  `end`
  % Computation of $z := M_1(k)^{-1} \xi$:
  Solve $S z_y = \xi_\lambda$ iteratively.
  Solve $S^T z_\lambda = \xi_y - Q z_y$ iteratively.
  Solve $D\, z_u = \xi_u - B_1 z_y - B_2\, z_\lambda$ directly.
  % Update of x:
  Set $x^{(i+1)} := x^{(i)} + (z_y - G_y z_u, z_\lambda - G_\lambda z_u, z_u)^T$.
`end`

We point out that the right transforming iterative scheme requires the approximate solution of m subproblems $Sz = f$ and $S^T z = f$ (with generic right hand sides f) plus 1 solution in each iteration. From a computational point of view it is therefore applicable only if fast PDE solvers are available or if m is "not too large", for example in case of shape optimization problems where only a small number of design parameters u_i, $1 \leq i \leq m$, is involved.

6. RIGHT TRANSFORMATIONS ALGORITHM (II)

To be able to treat a larger class of problems, we propose a different version of the RT algorithm where matrix $D(k)$ is not computed explicitly. Instead, an iterative solver for the system $Dz_u = \xi_u - B_1 z_y - B_2\, z_\lambda$ is used which requires only evaluations of the form Dv, $v \in \mathbb{R}^m$, that can be computed implicitly.

Algorithm 6.1. Right transformations (iterative version)

% Initialization:
Let the right hand side $b = (b_y, b_\lambda, b_u)^T$ and the matrix K with blocks as in (1) be given. Specify a tolerance $\epsilon \ll 1$ and set $x^{(0)} := 0$.

% Iteration:
`for` $i = 0, \ldots$ until convergence

Compute the residual $\xi = (\xi_y, \xi_\lambda, \xi_u)^T = b - Kx^{(i)}$.

% Check for convergence:
`if` $\|\xi\| < \epsilon\, \|b\|$
Solution is $x := x^{(i)}$, stop.
`end`

% Computation of $z := M_1(k)^{-1}\xi$*:*
Solve $Sz_y = \xi_\lambda$ iteratively.
Solve $S^T z_\lambda = \xi_y - Q\, z_y$ iteratively.
Solve $Dz_u = \xi_u - B_1 z_y - B_2\, z_\lambda$ **iteratively**, e.g., by GMRes. $(*)$
Thereby, each matrix-vector multiplication Dv is done implicitly by:
Compute $Sw_1 = B_2^T v$ iteratively.
Compute $S^T\, w_2 = B_1^T v - Q\, w_1$ iteratively.
Set $Dv := Cv - B_1 w_1 - B_2\, w_2$.

% Update of x*:*
Solve $S\tilde{z}_y = B_2^T z_u$ iteratively.
Solve $S^T \tilde{z}_\lambda = B_1^T z_u - Q\, \tilde{z}_y$ iteratively.
Set $x^{(i+1)} := x^{(i)} + (z_y - \tilde{z}_y, z_\lambda - \tilde{z}_\lambda, z_u)^T$.

`end`

Comments & Discussion:

- If $S = S^T$, $Q = Q^T$, and $C = C^T$, then also the matrix D is symmetric and we can

apply, for example, Lanczos' method in $(*)$, resulting in less memory requirements.
- Based on experience, we recommend to solve $(*)$ at least to an accuracy (w.r.t. the relative residual) of roughly $10^{-1}r^k$, i.e., one magnitude lower than the predicted contraction rate. Otherwise, the obtained contraction rates may be significantly worse than those obtained from Algorithm 5.1.
- In the version above, each RT iteration requires $(2 + 2\ell)$ approximate solutions of subproblems (3), where ℓ is the number of iterations spent for the solution of $(*)$. Thus, if i_{RT} denotes the overall number of RT iterations, it is necessary to solve $2\,i_{RT}(1 + \ell)$ times the subproblems (3), in contrast to $m + i_{RT}$ for Algorithm 5.1. Hence, only for $\ell < \frac{m}{2\,i_{RT}}$ Algorithm 6.1 turns out to be more economical than Algorithm 5.1 (assuming that the overall computational cost is dominated by the subproblems). Therefore, one is well advised to use a good preconditioner for D.
- The iterative version forfeits the possibility of parallelizing the code due to the successive character of its main loop.

The latter two issues lead to a nearby resort: Use a combination of Algorithm 5.1 and Algorithm 6.1, where the matrices G_y, G_λ and D are computed in a pre-processing step explicitly and, preferably, in a parallelized way. Then, in each iteration step, the linear system $Dz_u = \xi_u - B_1 z_y - B_2\, z_\lambda$ is solved iteratively as in Algorithm 6.1, but without the necessity to compute matrix-vector products Dv implicitly.

ACKNOWLEDGEMENTS

This work has been partially supported by the German Science Foundation DFG under SPP 1253 *Optimization with Partial Differential Equations* and by the NSF under the project *Multilevel Methods in PDE Constrained Optimization*, Grant-No. DMS-0511611.

REFERENCES

1. Antil H, Hoppe RHW, Linsenmann C. Adaptive multilevel interior-point methods in PDE constrained optimization. In *Domain Decomposition Methods in Science and Engineering XVIII*, Bercovier M, Gander MJ, Kornhuber R, Widlund O (eds). Lecture Notes in Computational Science and Engineering 2009, vol. 70. Springer: Berlin–Heidelberg–New York; 15–26.
2. —. Optimal design of stationary flow problems by path-following interior point methods. *Control and Cybernetics* 2008; **37**(4):771–796.
3. —. Adaptive path-following primal-dual interior point methods for shape optimization of linear and nonlinear Stokes flow problems. *Lecture Notes in Computer Science* 2008, vol. 4818. Springer: Berlin–Heidelberg–New York; 259–266.
4. —. Path-following primal-dual interior point methods for shape optimization of stationary flow Problems. *Journal of Numerical Mathematics* 2007; **15**(2):81–100.
5. Forsgren A, Gill PE, Griffin JD. Iterative solution of augmented systems arising in interior methods. *SIAM Journal on Optimization* 2007; **18**(2):666–690.
6. Fortin M, Glowinski R. *Augmented Lagrangian Method: Applications to the Numerical Solution of Boundary-Value Problems*. North-Holland: Amsterdam–New York–Oxford, 1983.
7. Gay DM, Overton ML, Wright MH. A primal-dual interior method for nonconvex nonlinear programming. In *Advances in Nonlinear Programming*, Yuan Y (ed). Kluwer: Dordrecht, 1998; 31–56.
8. Golub G. *Matrix Computations*. Johns Hopkins University Press: Baltimore, 1996.
9. Golub G, Greif C. On solving block-structured indefinite linear systems. *SIAM Journal on Scientific Computing* 2003; **24**(6):2076–2092.
10. —. *Techniques for Solving General KKT Systems*. TR SCCM-00-05, Stanford University, 2000.
11. Gill PE, Murray W, Wright MH. *Practical Optimization*. Academic Press: London–New York, 1981.

12. Hagemann LA, Young DM. *Applied Iterative Methods*. Academic Press: New York, 1981.
13. Hoppe RHW, Linsenmann C, Petrova SI. Primal-dual Newton methods in structural optimization. *Computing and Visualization in Science* 2006; **9**(2):71–87.
14. Hoppe RHW, Petrova SI. Primal-dual Newton interior point methods in shape and topology optimization. *Numerical Linear Algebra with Applications* 2004; **11**(5-6):413–429.
15. —. Applications of primal-dual interior methods in structural optimization. *Computational Methods in Applied Mathematics* 2003; **3**(1):159–176.
16. Hoppe RHW, Petrova SI, Schulz V. Primal-dual Newton-type interior-point method for topology optimization. *Journal of Optimization Theory and Applications* 2002; **114**(3):545–571.
17. Horn RA, and Johnson CR. *Matrix Analysis*. Cambridge University Press: Cambridge–New York–Melbourne, 1991.
18. Maar B, Schulz V. Interior point multigrid methods for topology optimization. *Structural and Multidisciplinary Optimization* 2000; **19**(3):214–224.
19. Schulz V, Wittum G. Transforming smoothers for PDE constrained optimization problems. *Computing and Visualization in Science* 2008; **11**(4-6):207–219.
20. —. Multigrid optimization methods for stationary parameter identification problems in groundwater flow. In *Multigrid Methods V*, Hackbusch W, Wittum G (eds). Lecture Notes in Computational Science and Engineering 1998, vol. 3. Springer: Berlin–Heidelberg–New York; 276–288.
21. Varga RS. *Matrix Iterative Methods*. Prentice-Hall: Englewood Cliffs, New Jersey, 1962.
22. Wittum G. Multi-grid methods for Stokes and Navier-Stokes equations. Transforming smoothers: Algorithms and numerical results. *Numerische Mathematik* 1989; **54**:543–563.
23. —. On the convergence of multi-grid methods with transforming smoothers: Theory with applications to the Navier-Stokes equations. *Numerische Mathematik* 1990; **57**:15–38.