UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 89-0697	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): A Numerical Study of an Augmented Lagrangian Method for the Estimation of Parameters in Elliptic Systems.	5. TYPE OF REPORT & PERIOD COVERED: Preprint
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s): K. Ito, M. Kroller and K. Kunisch	8. CONTRACT OR GRANT NUMBER(s): F49620-86-C-0111
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Lefschetz Center for Dynamical Systems, Division of Applied Mathematics, Brown University, Providence, RI 02912	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F, 2304/A1
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research, Bolling Air Force Base, Washington, DC 20332	12. REPORT DATE: January 1989 13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Same as 11	15. SECURITY CLASS. (of this report): Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A method based on an augmented Lagrangian formulation is developed which allows one to estimate coefficients in an elliptic differential equation from measurements of the state. This is a hybrid method combining the output-least-squares and the equation-error technique. Seminorm regularization is employed, and convergence and stability properties are discussed. Several aspects of an efficient implementation are decribed. Finally the effectiveness of the method is demonstrated by means of one and two dimensional examples.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

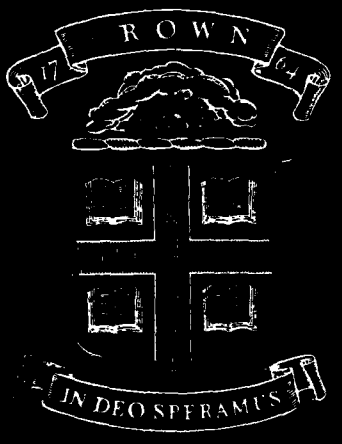
UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A Numerical Study of an Augmented Lagrangian Method for the Estimation of Parameters in Elliptic Systems.

by

K. Ito, M. Kroller and K. Kunisch

January 1989 LCDS/CCS # 89-7



Lefschetz Center for Dynamical Systems
and
Center for Control Sciences

Division of Applied Mathematics Brown University Providence RI 02912

A Numerical Study of an Augmented Lagrangian Method for the Estimation of Parameters in Elliptic Systems.

by

K. Ito, M. Kroller and K. Kunisch

January 1989 LCDS/CCS # 89-7

A Numerical Study of an Augmented Lagrangian Method for the Estimation of Parameters in Elliptic Systems

by
K.Ito[1], M.Kroller[2] and K.Kunisch[2]

January 1989

[1]This research was supported in part by the Air Force Office of Scientific Research under Contract F49620-86-C-0111, by NASA under Grant NAG-1-517 and by NSF under Grant UINT-8521208.

[2]This research was supported in part by the Fonds zur Foerderung der wissenschaftlichen Forschung under S3206 and P6005.

Addresses of Authors

K.ITO
Center of Control Sciences
Division of Applied Mathematics
Brown University
Providence, RI 02912, USA

M.KROLLER
Institut für Mathematik
Universitat Graz
Elisabethstr 12
8010 Graz, AUSTRIA

K.KUNISCH
Institut für Mathematik
Technische Universitat Graz
Kopernikusgasse 24
8010 Graz, AUSTRIA

Abstract

A method based on an augmented Lagrangian formulation is developed which allows one to estimate coefficients in an elliptic differential equation from measurements of the state. This is a hybrid method combining the output-least-squares and the equation-error technique. Seminorm regularization is employed, and convergence and stability properties are discussed. Several aspects of an efficient implementation are decribed. Finally the effectiveness of the method is demonstrated by means of one and two dimensional examples.

1 Introduction

In this paper we discuss our numerical experience with the hybrid method [IK2] based on the augmented Lagrangian formulation for estimating the coefficient q in the elliptic equation,

$$-\text{div}\,(q\ \text{grad}u) = f \text{ in } \Omega \tag{1.1}$$

$$u|_\Gamma = 0$$

from a measurement z of the solution $u\,(q^*)$ which corresponds to the "true" coefficient q^*. Here Ω is a bounded open set in $\Re^n, n = 1,2,3$ with piecewise smooth boundary Γ, and q is a scalar valued function of the spatial variable x. Frequently $u(q^*)$ is not known exactly; this may be due to model and/or measurement error and also because measurements may be known only at discrete points x_i in the domain Ω so that z is constructed by interpolation of pointwise data. A motivation for developing the hybrid method is to combine the output least squares method and the equation error method [C] into one algorithm while retaining the favorable properties of both. The augmented Lagrangian formulation, just as the equation error method, reduces to a quadratic programing problem and, like the output least squares method, it is versatile with regards to availability of observations as will be demonstrated in Section 4. The method is based upon formulating the problem as a constrained and regularized minimization problem (P^β) for two

independent variables $(q, u) \in H^2(\Omega) \times H_0^1(\Omega)$, and, as will be shown shortly, the minimization (P^β) is solved interatively by the augmented Lagrangian method. Let us consider:

$$(P^\beta) \qquad \text{minimize } \tfrac{1}{2} \mid u - z \mid^2_{H_0^1} + \tfrac{\beta}{2} N(q)$$
$$\epsilon(q, u) = (-\Delta)^{-1}(div(q\, grad\, u) + f) = 0$$
$$\text{subject to} \qquad l(q) = \alpha - q(x) \leq 0$$
$$g(q) = \tfrac{1}{2}(\mid q \mid^2_{H^2} - \gamma^2) \leq 0,$$

where we assume that $f \in H^{-1}(\Omega)$ and $z \in H_0^1(\Omega)$, that α and γ, are positive constants satisfying α meas$(\Omega)^{1/2} < \gamma$ and where Δ denotes the Laplacian with Dirichlet boundary condition as an operator from $H_0^1(\Omega)$ onto $H^{-1}(\Omega)$. Note that $\epsilon : H^2(\Omega) \times H_0^1(\Omega) \to H_0^1(\Omega)$ is a continuous bilinear map, and that $g : H^2(\Omega) \to \Re^1$ and the affine map $\ell : H^2(\Omega) \to H^2(\Omega)$ are continuous. Finally $\frac{\beta}{2} N(q)$ is a regularization term.

It is easy to show [IK2] that there exists a Lagrange multiplier $\lambda^* \in H_0^1(\Omega)$ associated with the equality constraint $\epsilon(q, u) = 0$ and it is given by

$$\lambda^* = (\nabla(q^\beta \nabla))^{-1} \Delta (u^\beta - z). \tag{1.2}$$

where (q^β, u^β) denotes a solution of (P^β).

A number of remarks are in order (see also Section 3 for a further discussion).

(1) *The pointwise constraint ℓ guarantees the strong ellipticity* and the

norm constraint g allows us to argue the existence of solutions of (P^β), $\beta \geq 0$. Without the norm constraint, (P^β) may not have a solution in general.

(2) The second term in the objective functional represents a regularization term in which $\beta \geq 0$ and $N(\beta)$ is a seminorm on $H^2(\Omega)$; e.g.

$$N(q) = |\nabla q|^2_{L^2} + \sum_{1 \leq i,j \leq n} |q_{x_i x_j}|^2_{L^2}. \tag{1.3}$$

For an appropriate choice of $\beta \geq 0$ the use of a regularization term in (P^β) induces the continuity of the mapping from the observation $z \in H^1_0$ to the solution $(q^\beta(z), u^\beta(z))$ of (P^β), see [CK1, CK2]. In general β cannot be taken equal to zero if continuous dependence of the solutions of (P^β) on z is desired.

(3) Since u is uniquely determined from q through the equality $e(q, u) = 0$, the objective functional can be considered to be dependent on q only. In this way (P^β) is reduced to the well known regularized least squares method given by

$$\text{minimize} \qquad |u(q) - z|^2_{H^1_0},$$

$$\text{subject to} \qquad -\operatorname{div}(q \operatorname{grad} u(q)) = f, \qquad l(q) \leq 0, \quad \text{and} \quad g(q) \leq 0.$$

From an optimization point of view (P^β) and the least squares method are equivalent. But in (P^β) q and u are independent variables and hence the implementations based on (P^β) are very different from those of the regularized least squares formulation.

(4) If $n = 1$ the parameter space $H^2(\Omega)$ can be replaced by $H^1(\Omega)$. In this case $g(q) = \frac{1}{2}(|q|^2_{H^1} - \gamma^2)$ and $N(q) = |q_x|^2$.

The augmented Lagrangian method applied to (P^β) involves a sequence of minimizations of the functionals

$$L_{c_k}(q, u, \lambda^k) = \tfrac{1}{2} \mid u - z \mid^2_{H_0^1} + \tfrac{\beta}{2} N(q) + \langle \lambda^k, e(q,u) \rangle_{H_0^1} + \tfrac{c_k}{2} \mid e(q,u) \mid^2_{H_0^1} \tag{1.4}$$

subject to $\ell(q) \leq 0$ and $g(q) \leq 0$

and the multiplier sequence $\{\lambda^k\}$ in $H_0^1(\Omega)$ is generated by the updating rule

$$\lambda^{k+1} = \lambda^k + c_k e(q_k, u_k) \tag{1.5}$$

where the pair (q_k, u_k) is a (local) minimizer of $L_{c_k}(\cdot, \cdot, \lambda_k)$. To carry out this iterative scheme a sequence of monotonically nondecreasing, positive, real numbers $\{c_k\}$ and the start-up value $\lambda^1 \epsilon H_0^1(\Omega)$ for the Lagrange multiplier for the equality constraint $e(q,u) = 0$ need to be chosen. In view of (1.2) we suggest $\lambda^1 = 0$ but convergence will be guaranteed for any other choice of λ^1 as well. The inequality constraint $g(q) \leq 0$ can be augmented in a similar manner as the equality constraint $e(q,u) = 0$ (see [IK1,IK2,KK1] for details). The hybrid property of our algorithm is now evident since the term $\frac{c_k}{2} \mid e(q,u) \mid^2_{H_0^1}$ involves the equation error

$$\mid div(q \, grad \, u) + f \mid^2_{H^{-1}}$$

and the term $\frac{1}{2} \mid u - z \mid^2_{H_0^1}$ represents an output least squares criterion.

2 Algorithm and Implementation

We carried out our computations for the choices $\Omega = [0,1]$ and $\Omega = [0,1] \times [0,1]$. For the computations, we discretized the problem using the finite element method [AB]; i.e. we represented q and u by

$$q^m(x) = \sum_{i=1}^{m} q_i^m \psi_i^m(x) \in H^m \tag{2.1}$$

$$u^n(x) = \sum_{j=1}^{n} u_j^n \phi_j^n(x) \in V^n \tag{2.2}$$

where $\phi_i^n \in H_0^1(\Omega)$ and $\psi_i^m \in L^\infty(\Omega)$. We used the following types of basis functions for ψ_i^m and ϕ_i^n in the two dimensional case.

Type I The functions ψ_i^m and ϕ_i^n are piecewise linear tensor splines [Sch]: i.e. $m = (M+1)^2$

$$\psi^m_{j(M+1)+i+1} = B_i^m(x_1) B_j^m(x_2) \quad \text{for } i,j = 0,\cdots,M$$

and $n = (N-1)^2$

$$\phi^n_{(j-1)(N-1)+i} = B_i^N(x_1) B_j^N(x_2) \quad \text{for } i,j = 1,\cdots,N-1$$

where

$$B_i^N(x) = \begin{cases} N(x - \frac{i-1}{N}) & \text{on } [\frac{i-1}{N}, \frac{i}{N}] \\ N(\frac{i-1}{N} - x) & \text{on } [\frac{i}{N}, \frac{i+1}{N}] \\ 0 & \text{, elsewhere.} \end{cases}$$

Type II The functions ψ_i^m and ϕ_i^n are piecewise linear basis functions on triangular elements with nodes at $\left(\frac{i}{M}, \frac{j}{M}\right)$, $i,j = 0, \cdots, M$ and nodes at $\left(\frac{i}{N}, \frac{j}{N}\right)$, $i,j = 1, \cdots, N-1$, respectively, where $m = (M+1)^2$ and $n = (N-1)^2$.

Type III The functions ϕ_i^n are the same as in Type II and the functions ψ_i^m are piecewise constant: i.e.,

$$\psi^m_{(j-1)M+i} = \chi_{(\frac{i-1}{M}, \frac{i}{M})}(x_1)\chi_{(\frac{j-1}{M}, \frac{j}{M})}(x_2)$$

$$\text{for } i,j = 1, \cdots, M.$$

where $\chi_{(a,b)}$ is the indicator function of the interval (a,b) and $m = M^2$. In the one dimensional case we used linear splines [Sch] for the approximation of both q^m and u^n.

As a regularization term we took

$$N(q) = \int_0^1 \int_0^1 \left(|q_{x_1}|^2 + |q_{x_2}|^2\right) dx_1 dx_2, \tag{2.3}$$

with the obvious modification in the one-dimensional case. Here we did not follow the theory developed in [IK2] which requires that for n =2 or 3 the second order terms be included and that N be chosen as in (1.3).

With these specifications made, the description of $(P^\beta), \beta \geq 0$, is given by

$$(P_\beta^{m,n}) \quad \text{minimize} \quad J^{m,n}(\underline{q}^m, \underline{u}^n) = \tfrac{1}{2}\underline{u}^{n^T} H \underline{u}^n + \underline{u}^{n^T}\underline{b} + \tfrac{1}{2}|z|^2_{H_0^1} + \tfrac{\beta}{2}\underline{q}^{m^T} W \underline{q}^m$$

subject to

$$\tilde{\epsilon}_i^{m,n}(\underline{q}^m, \underline{u}^n) := \langle q^m \bigtriangledown u^n, \bigtriangledown \phi_i^n \rangle - \langle f, \phi_i^n \rangle = 0 \quad 1 \leq i \leq n$$

$$q_j^m \geq \alpha \quad \text{for } 1 \leq j \leq m$$

$$\underline{q}^{m^T} \tilde{W} \underline{q}^m \leq \gamma^2$$

where $\underline{q}^m = \text{col}(q_1^m, \cdots, q_m^m) \in \mathcal{R}^m$ is the coordinate vector of $q^m(x) \in L^\infty(\Omega)$ in (2.1) and $\underline{u}^n = \text{col}(u_1^n, \cdots, u_n^n) \in \Re^n$ is the coordinate vector of $u^n(x) \in H_0^1(\Omega)$ in (2.2). The coordinate vector $\underline{\epsilon}^{m,n}$ in $\Re^n$ of $\epsilon^{m,n}$ is given by $\underline{\epsilon}^{m,n} = H^{-1}\tilde{\underline{\epsilon}}^{m,n}$. In what follows we will use the symbol q^m for both the element in $L^\infty(\Omega)$ and its vector coordinate in $\mathcal{R}^m$ and similarly for u^n. Moreover the bar is deleted from the notation of $\tilde{\underline{\epsilon}}^{m,n}$. The matrices $H \in \Re^{n\times n}$ and $W \in \Re^{m\times m}$ and the vector $b \in \Re^n$ are given by

$$H_{i,j} = \langle \bigtriangledown \phi_i^n, \bigtriangledown \phi_j^n \rangle, \qquad i,j = 1, \cdots, n,$$

$$W_{i,j} = \langle \bigtriangledown \psi_i^m, \bigtriangledown \psi_j^m \rangle, \qquad i,j = 1, \cdots, m,$$

$$b_i^n = \langle \bigtriangledown \phi_i^n, \bigtriangledown z \rangle \qquad i = 1, \cdots, n.$$

We did not implement the norm constraint $|q|^2_{H^2} \leq \gamma$ in the two dimensional case, see also section 3 in this respect. If one were to implement this con-

straint, one efficient way is to choose $\tilde{W}$ as the symmetric positive definite matrix $\tilde{W}$ on $\Re^m$ given by

$$\tilde{W} = W^2 + W + \tilde{Q}$$

with $\tilde{Q}_j = \langle \psi_i^m, \psi_j^m \rangle_{L^2}, \quad , j = 1, \cdots, m.$

The augmented Lagrangian function associated with $(P_\beta^{m,n})$ with only the equality constraint augmented has the form

$$\begin{aligned} L_{c_k}^{m,n}(q^m, u^n, \lambda_k) = \; & \tfrac{1}{2} {u^n}^T H u^n + {u^n}^T b + \tfrac{1}{2}|z|_{H_0^1}^2 \\ & + \tfrac{\beta}{2} {q^m}^T W q^m + \tfrac{c_k}{2} (\tilde{\epsilon}^{m,n})^T H^{-1} \tilde{\epsilon}^{m,n} + {\lambda_k^n}^T \tilde{\epsilon}^{m,n}. \end{aligned} \tag{2.4}$$

The Lagrange multiplier $\lambda_k^n \in \Re^n$ is updated by

$$\lambda_{k+1}^n = \lambda_k^n + c_k H^{-1} \tilde{\epsilon}^{m,n}(q_k^m, u_k^n) \tag{2.5}$$

where $\lambda_k^n(x) = \sum_{i=1}^n (\lambda_k^n)_i \phi_i^n(x) \in H_0^1$ approximates the Lagrange multiplier $\lambda^* \in H_0^1$ for the equality constraint $\epsilon(q, u) = 0$ in (P^3). We recall that (2.5) can be considered as a steepest ascend algorithm for the dual problem associated with $(P_\beta^{m,n})$.

Next we describe two algorithms to solve $(P_\beta^{m,n})$.

Algorithm 1

(1) Choose $\lambda_1 = 0$ and $\{c_k\}$ monotonically increasing with c_1 sufficiently large.

(2) Set $k = 1$.

(3) Determine (q_k^m, u_k^n) from

$$\text{minimize} \quad L_{c_k}^{m,n}(q^m, u^n, \lambda_k^n) \quad \text{subject to} \quad q^m \geq \alpha.$$

(4) If convergence is achieved, then stop. Otherwise, put

$$\lambda_{k+1}^n = \lambda_k^n + c_k H^{-1} \tilde{\epsilon}^{m,n}(q_k^m, u_k^n).$$

Return to (3) with $k = k + 1$.

The second algorithm takes advantage of the quadratic property of $L_{c_k}^{m,n}$: i.e., $L_{c_k}^{m,n}$ is quadratic in q^m(resp. u^n) if u^n is fixed (resp. if q^m is fixed).

Algorithm 2

(1) Same as in Algorithm 1.

(2) Set $k = 1$ and $u_0^n = H^{-1}b$.

(3) Determine q_k^m from

$$(P_{equ}) \text{ minimize } \frac{c_k}{2} \tilde{\epsilon}^{m,n}(q^m, u_{k-1}^n)^T H^{-1} \tilde{\epsilon}^{m,n}(q^m, u_{k-1}^n) + \lambda_k^{n^T} \tilde{\epsilon}^{m,n}(q^m, u_{k-1}^n) \quad \text{over } q^m$$

subject to $q^m \geq \alpha$.

(4) Determine u_k^n from

$$(P_{out}) \text{ minimize } \frac{1}{2} u^{n^T} H u^n + u^{n^T} b \; + \lambda_k^{n^T} \tilde{\epsilon}^{m,n}(q_k^m, u^n) + \frac{c_k}{2} \tilde{\epsilon}^{m,n}(q_k^m, u^n)^T H^{-1} \tilde{\epsilon}^{m,n}(q_k^m, u^n) \quad \text{over } u^n$$

(5) is identical to (4) in Algorithm 1.

We point out that the choice $u_0^n = H^{-1}b$ guarantees that the solution q_1^m of (P_{equ}) coincides with the solution of the equation error formulation in H^{-1}.

Based on our theoretical work [IK1, IK2], the choice of c_1 must be made such that the Lagrangian function of (P^β) is uniformly convex at the solution (q^β, u^β). Thus c_1 depends on $z, f, \gamma, \beta, \alpha$ and embedding constants. In our calculations, we chose c_1 heuristically. The monotonically increasing sequence $\{c_k\}_{k=2}^{\infty}$ could be determined according to known heuristics for augmented Lagrangian method, see [PT], for example. We took c_k to be a fixed positive constant in most of our calculations. The gradient of $L_{c_k}^{m,n}$ with respect to q^m is given by

$$\beta W q^m + \Phi^T(u^n)\left[\lambda_k^n + c_k H^{-1}(\Phi(u^n)q^m - f^n)\right]$$

and the gradient of $L_{c_k}^{m,n}$ with respect to u^n is given by

$$Hu^n + b^n + H(q^m)\left[\lambda_k^n + c_k H^{-1}(H(q^m)u^n - f^n)\right]$$

where $\tilde{\epsilon}^{m,n}(q^m, u^n) = \Phi(u^n)q^m - f^n = H(q^m)u^n - f^n$ and the matrices $\Phi(u^n) \in \mathcal{R}^{n\times m}$ and $H(q^m) \in \mathcal{R}^{n\times n}$ are defined by

$$\Phi(u^n)_{k,l} = \langle \psi_l^m \nabla u^n, \nabla \phi_k \rangle, \quad k = 1, \cdots, n; \quad l = 1, \cdots, m;$$

$$H(q^m)_{i,j} = \langle q^m \nabla \phi_i, \nabla \phi_j \rangle, \quad i, j = 1, \cdots, n.$$

With our choice of basis functions for ψ_i^m and ϕ_i^n, the matrices $\Phi(u^n)$ and $H(q^m)$ are sparse. In the two dimensional case, the gradient calculations can

be performed by the order of $N^2 + M^2$ operations. The calculation of $H^{-1}\tilde{e}$ for given $\tilde{e} \in \mathcal{R}^m$ requires the order of N^3 operations.

Remarks (1) In Algorithm 2, Steps (3) and (4) are a quadratic minimization in q^m and u^n, respectively.

(2) One can successively use the Steps (3) and (4) in Algorithm 2 to obtain the solution (q_k^m, u_k^m) to the minimization of $L_{c_k}^{m,n}(q^m, u^n, \lambda_k^n)$ in Step (3) of Algorithm 1. It is recommended that in the initial stages (i.e., $k = 1, 2$) of Algorithm 2, Steps (3) and (4) are repeated a few times before updating the Lagrange multiplier in Step (5).

(3) We used the conjugate gradient method [AB] to perform Step (3) in Algorithm 1 and Steps (3) and (4) in Algorithm 2 (we used the $ZXCGR$-routine in the IMSL library in our calculations).

(4) The matrix operation H^{-1} (which approximates $(-\Delta)^{-1}$) plays the role of (pre-) conditioning. In fact, (P_{out}) can be written as

minimize

$$\begin{aligned} &\tfrac{1}{2}{u^n}^T(H \quad + c_k H(q_k^m)H^{-1}H(q_k^m))u^n \\ &\qquad + {u^n}^T(b^n + H(q_k^m)(\lambda_k^n - c_k H^{-1} f^n)) + \text{ constant} . \end{aligned} \tag{2.6}$$

Since $H(q_k^m)$ is the approximation to $-\nabla(q_k^m(\nabla u))$ on V^n, H and $H(q_k^m)H^{-1}H(q_k^m)$ possess the same order of spectral condition number. Without the operation H^{-1} the equation error term in (P_{out}) would exhibit the square of the condition number of H. For the case $\Omega = [0,1] \times [0,1]$ (in general,in the

multi-dimensional case) the operation $H^{-1}\epsilon$ for given $\epsilon \in \Re^n$ is the most costly calculation in our algorithms. Hence one may use a smoothing matrix V based on the SSOR (symmetric successive overrelaxation) method [AB] or on a multi-grid method [HT] in place of H^{-1} so that less calculation is required. In this case the condition number of $H(q_k^m)VH(q_k^m)$ is improved (see [AB]).

(5) It is recommended to use a pre-conditioned conjugate gradient algorithm for the minimization of (2.6) since the condition number of the matrix $H(q_k^m)H^{-1}H(q_k^m)$ is still of the order of N^2. The preconditioned conjugate gradient method can be formulated as follows: setting

$$Q = H + c_k H(q_k^m)H^{-1}H(q_k^m).$$

the j-th step consists of

$$\begin{cases} \alpha_j = g_j^T h_j / d_j^T Q d_j \\ u_{j+1} = u_j + \alpha_j d_j \\ g_{j+1} = g_j + \alpha_j Q d_j \\ h_{j+1} = H^{-1} g_{i+1} \\ \beta_j = g_{j+1}^T h_{j+1} / g_j^T h_j \\ d_{j+1} = -h_{j+1} + \beta_j d_j \end{cases}$$

where u_j is the j-th iterate for the minimizer, g_j is the corresponding gradient vector, and d_j is the conjugate direction. Initially one chooses u_0 and puts

$g_o = Hu_0 + b^n + H(q_k^m)(\lambda_k^n - c_k H^{-1} f^n)$, $h_0 = H^{-1} g_0$ and $d_0 = -h_0$. Again one could replace the smoothing operation H^{-1} by one of the alternates described above.

3 Theoretical Considerations

In this section we will discuss some of the theoretical issues with regards to estimation problems in elliptic PDEs and their solution by means of the augmented Lagrangian technique.

Convergence of algorithm. The convergence of the augmented Lagrangian method that we described in the previous sections has been studied in [IK2]. Under appropriate conditions on the problem data $(z, f, \alpha, \beta, \gamma) \in H_0^1(\Omega) \times H^{-1}(\Omega) \times (\Re^+)^3$ appearing in (P^β), convergence of the sequence (q_k, u_k, λ_k) to $(q^\beta, u^\beta, \lambda^*)$ in $H^2(\Omega) \times H_0^1(\Omega) \times H_0^1(\Omega)$ was shown. Here (q_k, u_k) are minimizers (in a neighborhood of (q^β, u^β)) of the functionals $L_{c_k}(q, u, \lambda^k)$ subject to $l(q) \leq 0$ and $g(q) \leq 0$ as specified in (1.4), λ_k is determined by (1.5) and (q^β, u^β) is a solution of (P^β) with associated Lagrange multiplier λ^*. A critical step in the proof of the convergence results involves showing the positivity of the Hessian of the Lagrangian associated with (P^β), evaluated at $(q^\beta, u^\beta, \lambda^*)$. It is our future interest to study convergence properties of the solutions (q^m, u^n) of the discretized problem $(P_\beta^{m,n})$ as $(m, n) \to \infty$.

Semi-norm regularization. The use of the regularization term $\beta N(q)$ in (P^β) is common in solving ill-posed inverse problems. In [CK1,CK2,KS] for example, the stability of the solutions to the regularized output least squares method with respect to perturbations in the problem data in (P^β) is studied.

We point out the fact that the regularization term that we propose here is only a semi norm on H^2. More specifically, it does not contain the $|q|_{L_2}$-term so that only the variation of q is regularized. If the L_2-norm is included in the regularization term, we normally obtain under-estimated solutions.

Identifiability and Singular set. It is easy to argue the existence of a solution (q^β, u^β) of (P^β) for $\beta \geq 0$. Convergence of (q^β, u^β) to a minimum norm solution of (P^0) (that is, a solution of (P^0) that minimizes $N(q)$) as $\beta \to 0^+$ was studied in [IK2,EKN]. Uniqueness of the solutions of (P^0) and (P^β), $\beta > 0$ cannot be guaranteed in general. We mention here a related result concerning the injectivity of the mapping $q \to u(q)$ and refer to [K] for further discussion of these matters. Injectivity of $q \to u(q)$ is a necessary but not a sufficient condition to obtain the continuity of the inverse of the map $q \to u(q)$. Let q^* and q satisfy the constraints in (P^β) where q^* is the "true" reference coefficient and q is a perturbed one. Let us denote by $u(q^*)$ and $u(q)$, the corresponding solutions. Then we obtain the estimate [K].

$$\int_\Omega |q - q^*| \; |\nabla u(q^*)|^2 \, dx \leq K \, |u(q^*) - u(q)|_{W^{2,1}(\Omega)} . \tag{3.1}$$

where the constant $K > 0$ depends on

$$|q^*|_{W^{1,p}}, |q|_{W^{1,p}}, \quad p > n \text{ and } |u(q^*)|_{L^\infty}.$$

Let us define the singular set

$$S = \left\{ x \in \bar{\Omega}, \nabla u(q^*)(x) = 0 \right\}. \tag{3.2}$$

Clearly, if meas$(S) = 0$ and $u(q^*) = u(q)$, then $q^* = q$ a.e., so that the injectivity of the map $q \to u(q)$ holds at q^*.

Choice of the Output Least Squares criterion. In our calculations, we use the H_0^1-norm as an output least squares criterion. Such a choice is based on the facts that (i) the H_0^1-topology for the output least squares criterion and the H^{-1}-topology for the equation error term are natural from the point of view of the second order sufficient optimality condition for (P^3) (see [IK2] for details), (ii) the choice of these topologies also leads to a method that requires the same amounts of numerical differentiations in both the equation error and the output least squares term, and (iii) it enhances the sensitivity of the map $q \to u(q)$. In terms of computational efforts there are no differences between the use of the L_2-norm criterion and that of the H_0^1-norm criterion. With a view on balancing the output least squares and the equation error terms, one may take the H^{-2}-topology for the equation error term if the L^2-norm is used for the output least squares criterion, so that the augmented Lagrangian function (1.4) is replaced by

$$
\begin{aligned}
L_{c_k}(q, u, \lambda^k) \;&= \tfrac{1}{2}|u - z|_{L^2} + \beta \mathcal{N}(q) \\
&\quad + \langle \lambda, e(q,u) \rangle_{L^2} + \tfrac{c_k}{2} |e(q,u)|^2_{L^2}
\end{aligned}
\tag{3.3}
$$

where $e(q,u) = (-\Delta)^{-1}(\mathrm{div}(q \ \mathrm{grad}\ u) + f)$.

A general class of inverse problems in elliptic partial differential equations. In this paper we restricted our attention to the Dirichlet boundary value problem (1.1). Without difficulties one can modify our formulation to treat other types of boundary conditions as well. For example, consider the problem of determining $q(x)$ in

$$-\text{div}(q \text{ grad } u) + p(x)u = f$$

with the boundary condition $q\frac{\partial}{\partial n}u = g$ on Γ where $\frac{\partial}{\partial n}u$ denotes the outward normal derivative, $p(x) \in L^\infty$ with $p(x) \geq w > 0$ and $g \in H^{-1/2}(\Gamma)$. In this case, we define the equation error term by

$$e(q,u) = (-\Delta + 1)^{-1}(-\text{div}(q \text{ grad } u) + pu - f)$$

where Δ is the Laplacian with $\text{dom}(\Delta) = \{u \in H^2(\Omega) \text{ and } \frac{\partial}{\partial n}|_\Gamma = 0\}$ and H_0^1 is replaced by H^1 in (P^β). While we are concerned with estimating the diffusion coefficient in this paper, our formulation enables us to consider the estimation problem for other coefficients in elliptic partial differential equations (e.g., $p(x) \in L^\infty(\Omega)$ in the above example).

Inequality constraints. As pointed out in the Introduction, the inequality constraints (i.e., the pointwise constraint l and the norm constraint g) are required for existence of solutions to (P^β). In our calculations, we ignored the pointwise lower bound without harm. Besides, in our formulation the positivity of $q(x)$ is not a hard constraint since we do not need to solve the

equation (1.1) for u in each iteration. The relevance of the norm constraint was investigated for the one dimensional case in [KK1]. It was observed that its use improved the results if the interior of the singular set $S = \{x \in [0,1], u_x(q^*) = 0\}$ is not empty and it did not change them significantly if meas $S = 0$, provided, of course, that γ is chosen such that $|q^*|_{H^1(0,1)} \leq \gamma$. We did not use the norm constraint in our implementation for the case $\Omega = [0,1] \times [0,1]$ since we do not expect that any new phenomena arise. Moreover the regularization term also provides a semi-norm bound for the solutions.

Miscellaneous comments.

An implicit regularization is often achieved by choosing coarser basis elements for representing $q(x)$ than for $u(x)$. For example, in some of our calculations, we took $N = 2M$ (see [KK1, KK2, KK3]).

4 Numerical Results

Numerous numerical experiments were carried out for the two algorithms that we explained in Section 2. The tests were conducted in the following way. For all examples except for Example 7 we chose the "true" parameter q^* as well as the solution $u^* = u(q^*)$ and calculated f from (1.1). For Example 7 we chose q^* and f and calculated $u^* = u(q^*)$ numerically. The data z were determined from $u(q^*)$ either by putting $z = u(q^*)$ ("distributed data") or by evaluating $u(q^*)$ at points $x_i \in \Omega$ and determining z as a cubic Hermite spline interpolation of $z_i = u(q^*, x_i)$. Noisy data were obtained by specifying

$$z_i = u(q^*, x_i) + \delta\xi_i,$$

with ξ_i uniformly distributed random number in [-1,1], and choosing z as the cubic Hermite spline interpolation of these z_i. The numerical examples to be discussed below are divided into two groups an are based on two different packages. The first six examples demonstrate the performance of the algorithm while a specific aspect, as for instance the behavior near the singular set, is under investigation and the number of unknowns is no larger than 81. In Examples 7 and 8 we show that the algorithm that we propose also works very well for an extremely fine resolution (64×64) of the unknown coefficient. For the calculations of Examples 1-6 we took Type I basis functions with the grid for the state approximation twice as fine as the grid for the

coefficient approximation ($N = 2M$). This choice introduces some inherent regularization. The $(-\Delta)^{-1}$ operation is resolved by Cholesky factorization and the optimization problems are solved with the IMSL version $ZXCGR$ of the conjugate gradient algorithm. The discretization for Examples 7 and 8 used Type III basis functions with $N = M$. The $(-\Delta)^{-1}$ operation is realized by a multigrid step involving one V-cycle with finite element injection, while the optimization problems are solved by a preconditioned conjugate gradient algorithm as explained in (5) of Section 2. Moreover, Algorithm 2 was slightly modified by iterating several times between steps (3) and (4) before going to step (5). The results from a coarser mesh are then used as start-up values for the next finer mesh. The calculations for Examples 1-6 were carried out on an IBM-AT and those for Examples 7-8 on an IBM 3081, both in double precision. For further specification of the numerical aspects and for additional test examples we refer to [KK 1-3].

We summarize some general observations.

(1) In all examples the augmented Lagrangian algorithm in the form of Algorithm 1 as well as Algorithm 2 performed well.

(2) In all examples where complete data were available (i.e. $z = u(q^*)$), where q^* was a smooth function and where the singular set S was small, the first or second iteration gave a good approximation to both q^* and $u(q^*)$. If these requirements were not satisfied, and m was sufficiently large, then

higher iterations gave an improvement over the first one. Recall that for Algorithm 2, the determination of the first iteration q_1^m coincides with the solution of the equation error algorithm.

(3) Algorithm 2 is significantly faster than Algorithm 1.

(4) For the estimation of q^* the form of the singular set S is of great importance. If the dimension of S is smaller than the dimension of Ω, (and $\beta = 0$ or β is very small), then the maximum error of the approximation to q^* generally occurs within a small neighborhood of S. If S contains an open set, then q^* cannot be estimated there.

(5) We could observe conversion and rates of convergence of the solutions to $(P_\beta^{m,n})$ as $m \to \infty$, $n \to \infty$, see [KKI, KK2].

(6) The use of a regularization term, especially if the same grid is chosen for both the coefficient and the state space approximation, or to avoid oscillations in the neighborhood of a singular set, is useful. We are currently working on an adaptive algorithm to specify the level of regularization that is required for a specific problem.

(7) The augmented Lagrangian algorithm that we propose also works very well when applied to problems with discontinuous unknown coefficients. Comparing the results where a gridpoint of the discretization for the coefficient coincides with a discontinuity of q^*, to those where discontinuities and grid points do not coincide, the latter were clearly better.

Most of the plots that we present below are endowed with the following information:

Example X, $\beta = X$, $M = X$, $k = X$, $L^2 = x$, additional comments.

Here L^2 stands for the L^2-error $|q_k^M - q^*|_{L^2}$. Unless otherwise specified, the data were taken to be error free distributed observations, $z = u(q^*)$. Examples 3-6 were calculated with Algorithm 1, the remaining examples with Algorithm 2. Since the numerical results always produced a very good approximation to $u(q^*)$ we only show the numerical results for the approximation of the coefficient q^* (except in Example 6). In the one dimensional examples $\Omega = (0,1)$ and in the two dimensional examples $\Omega = (0,1) \times (0,1)$. The start-up value for q^m for the finite dimensional approximating problems $(P_\beta^{m,n})$ was chosen to be identically 1 in Examples 1-6 and 1.5 in Examples 7 and 8. For all calculations c_k was taken to be equal to 1.

Example 1. Here we take

$$\begin{aligned} z(x,y) &= u(q^*)(x,y) = w(x)w(y), \\ q^* &= 2 + \sin(x^2 y) \end{aligned}$$

where

$$w(x) = \begin{cases} -9x^2 + 6x & , x \in [0, \frac{1}{3}] \\ 1 & , x \in (\frac{1}{3}, \frac{2}{3}) \\ -9x^2 + 12x + 3 & , x \in [\frac{2}{3}, 1], \end{cases}$$

with f calculated from (1.1). In Plot 1 we give the graph for z and q^* as well as the numerical result for q after 8 iterations without and with the use of a

regularization term. Here the singular set has non-empty interior. Observe that regardless of the use of a regularization term, q^* is identified well over the complement of the singular set.

Example 2. Here we took

$$\begin{aligned} z(x,y) &= u(q^*)(x,y) = (y - 2y^2)\sin 2\pi x \sin 2\pi y \\ q^* &= 2 + \sin(x^2 y). \end{aligned}$$

and f is calculated from (1.1).

The singular set for this example consists of isolated points and the line characterized by $y = 0$ and $y = \frac{1}{2}$. Plot 2 gives a graph of $|\nabla x|$ and Plot 3 shows the results after the first and eighth iteration. We point out the improvement in the neighborhood of the singular set characterized by $y = \frac{1}{2}$.

Example 3. Here we took

$$\begin{aligned} u(q^*) &= \sin 2\pi x \sin 2\pi y \\ q^* &= 1 + 6x^2 y(1 - y), \end{aligned}$$

and calculated f from (1.1). Then tests were carried out assuming that successively more data points become available. The results after one iteration of Algorithm 1 are shown in Plot 4. Further iterations so not change the results significantly for this example. This is probably due to the fact that the discretization of q is rather coarse.

Example 4. This is an example with a discontinuous coefficient:

$$z = u(q^*) = \sin \pi x$$

$$q^* = \begin{cases} 1 & for\ x \in [0, \frac{1}{3}] \\ 2 - x & for\ x \in (\frac{1}{3}, \frac{2}{3}) \\ -9x^2 + \frac{21}{2}x - 3 & for\ x \in [\frac{2}{3}, 1]. \end{cases}$$

The results for $M = 21$, $M = 22$ and $M = 42$ are given in Plot 5. For $M = 21$ and $M = 42$, the discontinuities are also grid points of the discretization for q. We point out that these estimates are obtained without a-priori assumptions on the location of the discontinuity and the size of the jump. The results can be improved by using regularization, see Plot 6.

Example 5. Here q^* is chosen as in Example 4. We compare the numerical results for two different observations $z_1 = \sin \pi x$ and $z_2 = \sin 2\pi x$. Observe that the singular set for z_1 is $S(z_1) = \frac{1}{2}$ and similarly $S(z_2) = \{\frac{1}{4}, \frac{3}{4}\}$. The last graph in Plot 7 shows the numerical result, when Algorithm 1 was modified in the obvious way to account for two observations simultaneously.

Example 6. For this example we chose

$$\begin{aligned} u(q^*) &= \sin \pi x, \\ q^* &= 1 + x. \end{aligned}$$

We put $z(x_i) = u(q^*)(x_i)$ with $x_i = \frac{i}{42}$, for $i = 0, \cdots, 7, 9, \cdots, 42$, and $z(x_8) = u(q^*)(x_8) + \frac{1}{4}$, thus producing an outlier at $\frac{4}{21}$. Then we passed a cubic Hermite spline interpolate through z_i, see the thicker lines in Plot 8. To estimate the unknown parameter q^* we used Algorithm 1 as well

as Algorithm 1 with the $|u - z|_{H_0^1}$ output criterion replaced by an $|u - z|_{W_0^{1,1}}$ criterion. The second curve in the graphs of Plot 8 give the numerical result for u^n. Differently from the H^{-1}-criterion, the outlier is essentially ignored by the numerical solution u^n, when the $W^{1,1}$-criterion is used. In this case the approximation q^m to q^* is still qualitatively correct while the H^1 criterion without regularization produces a useless estimate for q^*. For further discussion of this example see [KK3].

Example 7. Here we took

$$f = \sin 2\pi x \sin 2\pi y,$$

$$q^* = \begin{cases} 2 & for (x,y) \in (.3,.3) \times (.3,.6) \\ 1 & elsewhere \end{cases}$$

and data were assumed to be available on a 20×20 grid for the first result and on a 40×40 grid for the second result. Observe that adding more data points results in a sharper resolution of the discontinuous coefficient q^*.

Example 8. For this example, with

$$\begin{aligned} u(q^*) &= \sin 2\pi x \sin 2\pi y, \\ q^* &= 1 + 6x^2y(1-y) \end{aligned}$$

consecutively more noise was added to $u(q^*)(x_i)$ at the observation points x_i which were assumed at a uniform 10×10 grid. Observe that we are referring to absolute, rather than relative noise here.

Example 1
function z

Example 1
function q^*.

Plot 1

Example 1
$\beta = 0,\ M = 6,\ k = 8,\ L^2 = 0.12$

Example 1
$\beta = 0,\ M = 6,\ k = 8,\ L^2 = 0.0036$

Plot 2

Example 2
function $|\nabla z|$

Plot 3

Example 2
$\beta = 0,\ M = 9,\ k = 1,\ L^2 = 0.034$

Example 2
$\beta = 0,\ M = 9,\ k = 8,\ L^2 = 0.0065$

Plot 4

Example 3
$\beta = 0,\ M = 5,\ k = 1$
interpolated data at $\{\frac{i}{3}, \frac{j}{3}\}$

Example 3
$\beta = 0,\ M = 5,\ k = 1$
interpolated data at $\{\frac{i}{4}, \frac{j}{4}\}$

Example 3
$\beta = 0,\ M = 5,\ k = 1$
interpolated data at $\{\frac{i}{5}, \frac{j}{5}\}$

Example 4
$\beta = 0,\ M = 21,\ k = 2,\ L^2 = 0.101$

Example 4
$\beta = 0,\ M = 22,\ k = ,\ L^2 = 0.064$

Plot 5

Example 4
$\beta = 0,\ M = 41,\ k = 2,\ L^2 = 0.055$

Example 4
$\beta = 0,\ M = 42,\ k = ,\ L^2 = 0.060$

Plot 6

Example 4
$\beta = 10^{-5},\ M = 41,\ k = 3,\ L^2 = 0.0445$

Example 5
$\beta = 0,\ M = 41,\ k = 3,\ L^2 = 0.0525$

Example 5
$\beta = 0,\ M = 41,\ k = 3,\ L^2 = 0.047$

Plot 7

Example 5
$\beta = 0,\ M = 41,\ k = 3,\ L^2 = 0.0447,$
two distributed observations

Plot 8

Example 6
$\beta = 0,\ M = 6,\ k = 3,\ W^{1,1}$-criterion

Example 6
$\beta = 0.1,\ M = 6,\ k = 3,\ H^1$-criterion

Example 7
$\beta = 10^{-4},\ M = 64,\ k =$ 20 × 20 point measurements

Plot 9

Example 7
$\beta = 10^{-4},\ M = 64,\ k =$ 40 × 40 point measurements

Example 8

Plot 10 $\beta = 10^{-3}$, $M = 64$, $k =$ 10×10 point data with 0% noise
$\beta = 5 \times 10^{-3}$, $M = 64$, $k =$ 10×10 point data with 1% noise
$\beta = 5 \times 10^{-3}$, $M = 64$, $k =$ 10×10 point data with 3% noise

References

[AB] O.Axelsson and V.A.Barker: Finite Element Solution of Boundary Value Problems, Theory and Computation, Academic Press, New York, 1984.

[C] G. Chavent: Identification of distributed parameter systems: about the least square method, its implementation and identifiability. Proceedings of the 5^{th} IFAC Symposium on Identifiability and System Parameter Estimation, Darmstadt, 1979, Pergamon Press, 85-97.

[CK1] F. Colonius and K.Kunisch: Output least squares stability in elliptic systems, Appl. Math. Optimization, 19(1988), 33-63.

[CK2] F. Colonius and K. Kunisch: Stability of perturbed optimization problems with applications to parameter estimation, submitted to SIAM J. Appl. Math.

[EKN] H.W. Engl, K. Kunisch and A. Neubauer: Tikhonov regularization for the solution of nonlinear ill-posed problems I, submitted.

[HT] W.Hackbusch and U.Trottenberg (eds.): Multigrid Methods. Lecture Notes in Mathematics 960, Springer, Berlin, 1982.

[IK1] K.Ito and K.Kunisch: The augmented Lagrangian method for equality and inequaltiy constraints in Hilbert spaces, to appear in Math. Programming.

[IK2] K.Ito and K.Kunisch: The augmented Lagrangian method for parameter estimation in elliptic systems, to appear in SIAM J. Control and Optimization.

[K] K.Kunisch: Inherent identifiability of parameters in elliptic equations, J. Math. Anal. Appl. 132(1988), 453-472.

[KS] C.Kravaris and J.H.Seinfeld: Identification of parameters in distributed systems by regularization, SIAM J. Control and Optimization, 23(1985), 217-241.

[KK1] M.Kroller and K.Kunisch: A numerical study of an augmented Lagrangian method for the estimation of parameters in a two point boundary value problem, Institute for Mathematics, Technical University of Graz, Technical Report, No.87, March 1987.

[KK2] M.Kroller and K.Kunisch: A numerical study of an augmented Lagrangian method for the estimation of parameters in elliptic systems, Institute for Mathematics, Technical University of Graz, Technical Report, No.101, September 1987.

[KK3] M.Kroller and K.Kunisch: A numerical study of an augmented Lagrangian method for the estimation of parameters in elliptic systems: noisy data, outliers and discontinuous coefficients, Institute for Mathematics, Technical University of Graz, Technical Report, No.123, September 1988.

[PT] V.T.Polyak and N.Y.Tret'yakov: The Method of Penalty Estimates for Conditional Extremum Problems, Z.Vychisl. Mat. i Mat. Fiz. 13 (1973) 34-36.

[Sch] M.H.Schulz: Spline Analysis, Prentice-Hall, Englewood Cliffs, N.J., 1973.