To appear in "Computational Science for the 21st Century",
J. Periaux et al (eds.), Volume to be published by J. Wiley &
Sons in 1997 in honor of Roland Glowinski's 60th birthday.

UCLA

COMPUTATIONAL AND APPLIED MATHEMATICS

Augmented Lagrangian and Total Variation Methods for Recovering Discontinuous Coefficients from Elliptic Equations

Tony F. Chan
Xue-Cheng Tai

January 1997

CAM Report 97-2

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90024-1555

Augmented Lagrangian and Total Variation Methods for Recovering Discontinuous Coefficients from Elliptic Equations

Tony F. Chan [1] and Xue-Cheng Tai[2]

Abstract: Estimation of coefficients of partial differential equations is ill-posed. Output-least-squares method is often used in practice. Convergence of the commonly used minimization algorithms for the inverse problem is often very slow. By using the augmented Lagrangian method, the inverse problem is reduced to a coupled linear algebraic system, which can be solved efficiently. Total variation techniques have been successfully used in image processing. Here, we use it with the augmented Lagrangian approach to recover discontinuous coefficients. The numerical results show that our approach can recover discontinuous coefficients with large jumps from noisy observations.

1. Introduction

Consider the partial differential equation:

$$\begin{cases} -\nabla \cdot (q\nabla u) = f, \quad \text{on} \quad \Omega \,, \\ u = 0 \,, \quad \text{on} \quad \partial\Omega \,. \end{cases} \tag{0.1}$$

Our concern is that we know an approximate observation u_d for u, i.e. $u_d \approx u$, and need to recover the coefficient q. A common approach is the output-least-squares method, i.e. find the minimizer for

$$\min_{e(q,u)=0,\ q\in K} \frac{1}{2}\|u(q) - u_d\|^2 + \beta R(q) \,, \tag{0.2}$$

[1] Department of Mathematics, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90095-1555. Email: chan@math.ucla.edu, URL: http://www.math.ucla.edu/~chan. The work of this author is partially supported by the Office of Naval Research under contract ONR N00014-96-1-0277 and by the National Science Foundation under grant DMS-9626755.

[2] Department of Mathematics, University of Bergen, Alleg. 55, N-5007 Bergen, Norway. Email: Tai@mi.uib.no. URL: http://www.mi.uib.no/~tai. The work of this author is supported by the Research Council of Norway, under contract SEP-115837/431.

where $u(q)$ solves $e(q,u)=0$ and $e(q,u)=0$ represents the equation constraint (0.1) in a suitable space. The set K is the set of admissible coefficients. Generally, $\|u-u_d\|$ represents the H^1 or L^2-norm of $u-u_d$, or the summation of errors between u and u_d at some points. $R(q)$ is a regularization term which is needed to guarantee stability and existence.

It is well known that inverse problems are ill-posed. Consequently the norm used for $R(q)$ and $u-u_d$ must be properly chosen. To guarantee existence of q, we need to identify the parameter in a compact set in a suitable space. Moreover, when $q_n \to q$ weakly, the corresponding solutions $u(q_n)$ and $u(q)$ must satisfy

$$u(q_n) \to u(q)$$

in the norm of the space chosen for $u-u_d$. To guarantee uniqueness, one needs

$$\|u(q_n)-u(q)\| \le C\|q_n - q\|. \tag{0.3}$$

When the coefficient q is smooth, suitable techniques are available to identify the parameter, see Ito and Kunisch [IK90], Kunisch and Tai [KT97]. By using $R(q)=\|q\|^2_{H^2}$ for the coefficient q and H^1 or L^2-norm for the state $u-u_d$, the existence, uniqueness and convergence of problem (0.2) are shown for the identification of q in Kunisch and Tai [KT97]. However, in many applications we need to identify possibly discontinuous coefficient q with large jumps, for example, reservoir mechanics, [Ewi83]; image reconstruction [CGM96, ALM92], electrical impedance tomography [DS94]. For these applications, the use of H^2-norm for q will smear out the discontinuities of the coefficient and prevent us from getting accurate estimations.

Recent research from image processing reveals that by using the TV-norm (total variation norm), we can identify objects with discontinuities and sharp edges, see [CGM96, ALM92]. In this work, we combine the total variation regularization with an augmented Lagrangian method, see Glowinski [Glo84], to recover the coefficient q. More specifically, we take

$$R(q) = TV(q) = \int_\Omega |\nabla q| dx \ , \tag{0.4}$$

where, $TV(q)$ denotes the total variation of q, see Ziemer [Zie89] and Giusti [Giu84] for definitions. When q is not differentiable, $|\nabla q|$ is understood as a measure, see p. 111 of [Zie89]. In applications, u_d carries random noisy errors and it is better to use $L^2(\Omega)$ norm for $u-u_d$. Therefore, we shall solve:

$$\min_{e(q,u)=0,\ q\in K} \frac{1}{2}\|u-u_d\|^2_{L^2(\Omega)} + \beta TV(q) \tag{0.5}$$

to find the coefficient q from u_d.

2. The Augmented Lagrangian Approach

To solve problem (0.5), we use an augmented Lagrangian approach, see [Glo84, KT97]. Let Δ_0 be the Laplace operator with zero Dirichlet boundary condition. We let

$e = e(q,u)$ be the solution of

$$\begin{cases} -\Delta_0 e = -\nabla(q\nabla u) - f \text{ in } \Omega \ , \\ e = 0 \text{ on } \partial\Omega \ , \end{cases}$$

i.e.

$$e = \Delta_0^{-1}(\nabla(q\nabla u) + f).$$

For any $q \in L^\infty(\Omega)$, $u \in H_0^1(\Omega)$ and $\lambda \in H_0^1(\Omega)$ and a constant $c > 0$, define

$$\begin{aligned} L_c(q,u,\lambda) &= \frac{1}{2}\|u - u_d\|^2_{L^2(\Omega)} + \beta TV(q) \\ &+ (\lambda, e(q,u))_{L^2(\Omega)} + \frac{c}{2}\|e(q,u)\|^2_{L^2(\Omega)}. \end{aligned}$$

If $(q^*, u^*; \lambda^*)$ is a saddle point of L_c, i.e.

$$L_c(q^*,u^*,\lambda) \le L_c(q^*,u^*,\lambda^*) \le L_c(q,u,\lambda^*) \ , \qquad \forall q,u,\lambda \ ,$$

then (q^*, u^*) is a minimizer of (0.5), see p. 168 of [Glo84].

The following algorithm is often used to find a saddle point for $L_c(q,u,\lambda)$:

Algorithm 1 .

Given $\lambda^0, c > 0$.
Step 1. Find q_n, u_n *such that*

$$(q_n, u_n) = \arg\min_{q,u} L_c(q,u,\lambda^{n-1}). \tag{0.6}$$

Step 2. Set $\lambda^n = \lambda^{n-1} + ce(q^n, u^n)$.

From the definition of L_c, we can calculate that:

$$\begin{aligned} \frac{\partial L_c}{\partial q} &= -\beta\nabla \cdot \left(\frac{\nabla q}{|\nabla q|}\right) - \nabla u \cdot \nabla(\Delta_0^{-1}\lambda) \\ &- c\nabla u \cdot \nabla \cdot (\Delta_0^{-2}(\nabla(q\nabla u) + f)) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial L_c}{\partial u} &= u - u_d + \nabla \cdot (q\nabla(\Delta_0^{-1}\lambda)) \\ &+ c\nabla \cdot (q\nabla[\Delta_0^{-2}(\nabla(q\nabla u) + f)]). \end{aligned}$$

To find a minimizer of (0.6), we shall search for the points that satisfy:

$$\frac{\partial L_c}{\partial q} = 0, \quad \frac{\partial L_c}{\partial u} = 0. \tag{0.7}$$

The boundary condition for q is the natural boundary condition $\frac{\partial q}{\partial n} = 0$ on $\partial\Omega$ and the boundary condition for u is $u = 0$ on $\partial\Omega$. These two equations are coupled. An inner iteration alternating between them will be used to find a q and a u such that (0.7) is fulfilled.

3. The Discrete Formulation

The TV-norm functional $TV(q)$ is not differentiable with respect to q. For numerical purpose, we introduce

$$F(q) = \int_\Omega \frac{|\nabla q|^2}{\sqrt{|\nabla q|^2 + \varepsilon}}.$$

We shall use this functional as the regularization functional for our identification problem. If q is a function of $H^1(\Omega)$, then ∇q is just the generalised derivative of q, if not, we shall give an explicit formula for the discrete case.

Let $\Omega \subset R^n$, $n = 1, 2, 3$ be a bounded domain. We first divide Ω into finite elements $\mathcal{T}_h = \{e_i\}$. Let S_h denote the piecewise linear finite element space over $\mathcal{T}_h$ with zero Dirichlet boundary value on $\partial\Omega$. Let P_h denote the piecewise constant finite element space over $\mathcal{T}_h$. The space S_h will be used to approximate u and the space P_h will be used to approximate q. For a given $q \in P_h$, it is easy to calculate that

$$TV(q) = \int_\Omega |\nabla q| dx = \sum_{i<j} |q_i - q_j| |\bar{e}_i \cap \bar{e}_j|.$$

In the above, $q_i = q|_{e_i}$ and $|\bar{e}_i \cap \bar{e}_j|$ denotes the $(n-1)$-dimensional measure of the interface between $\bar{e}_i$ and $\bar{e}_j$. Here, e_i and e_j are the elements of the finite element division $\mathcal{T}_h$. Correspondingly, we define the discrete functional $F(q)$ as

$$F(q) = \sum_{i<j} \frac{|q_i - q_j|^2}{\sqrt{|q_i - q_j|^2 + \varepsilon}} |\bar{e}_i \cap \bar{e}_j| \ .$$

In identifying the coefficient q, equation (0.1) is regarded as a constraint. In order to get a suitable representation of the equation in our discretized space, let us define $e : P_h \times S_h \to S_h$ as the solution to:

$$(\nabla e, \nabla v)_\Omega = (q \nabla u, \nabla v)_\Omega - (f, v)_\Omega \ , \quad \forall v \in S_h. \tag{0.8}$$

Let $\{\theta_i\}$ be the basis functions of P_h, $\{\varphi_i\}$ be the basis functions of S_h. In the following, we shall use u to denote both the function u and the vector of its nodal values. We do the same thing for λ and e. In a similar way, q denotes both the function and the vector of its values on the finite elements. We denote

$$\begin{aligned} A_q &= [(q\nabla\varphi_i, \nabla\varphi_j)], \\ B_u &= [(\nabla u \cdot \nabla\varphi_i, \theta_j)], \\ C_0 &= [(\nabla\varphi_i, \nabla\varphi_j)], \\ g &= (f, \varphi_i). \end{aligned}$$

Equation (0.8) can now be written as:

$$C_0 e = A_q u - g = B_u q - g.$$

The matrix A_q is symmetric positive definite for any $q > 0$, but the matrix B_u is not a square matrix and is singular. For any parameter $q \in P_h$ and state $u \in S_h$,

$$\begin{aligned} e = e(q, u) &= C_0^{-1}(A_q u - g) \\ &= C_0^{-1}(B_u q - g) \ , \end{aligned}$$

and $e(q,u)=0$ indicates that (q,u) satisfies equation (0.1) in the discrete sense. Define

$$\begin{aligned} K \quad &= \{q \in P_h \,|\, 0 < k_1 \le q \le k_2\} \ , \\ G(u) \quad &= \frac{1}{2}\|u - u_d\|^2_{L^2(\Omega)}. \end{aligned}$$

We now consider the discrete version of (0.5):

$$\min_{e(q,u)=0,\ q\in K} \{\beta F(q) + G(u)\}. \tag{0.9}$$

The solution (q^*,u^*) of (0.9) will be taken as the identified coefficient and the corresponding state. In this discrete setting, every bounded set is compact, therefore it is easy to show the existence of solution (q^*,u^*) to (0.9).

We shall use the augmented Lagrangian method to solve (0.9). From now on, we use the notation $x=(q,u)$ and

$$L(x,\lambda) = \beta F(q) + G(u) + (\lambda, e(q,u))_{L^2(\Omega)}.$$

In the above, $\lambda \in S_h$. Corresponding to $x^* = (q^*,u^*)$, there exists a Lagrange multiplier $\lambda^* \in S_h$, i.e.

$$L(x^*,\lambda^*)(x-x^*) \ge 0, \quad \forall x \in K \times S_h.$$

By checking on the optimality conditions, we see that x^*, λ^* satisfy:

$$\beta F'(q^*) + B^*_{u^*}(C_0^{-1})\lambda^* = 0, \ q^* \in K \ , \tag{0.10}$$
$$u^* - u_d + A_{q^*}(C_0^{-1})\lambda^* = 0 \ , \tag{0.11}$$
$$e(q^*,u^*) = 0 \ . \tag{0.12}$$

In the above, $B^*_{u^*}$ denotes the transpose of the matrix B_{u^*}, and we note that A_{q^*} and C_0 are symmetric and positive define matrices.

We can use an iterative method to solve equations (0.10), (0.11) and (0.12) sequentially. However, due to the illposedness of the problem, such an iteration does not converge when observation errors are present. We therefore propose now an augmented Lagrange method to solve (0.10)–(0.12). For $c>0$, we define

$$L_c(x,\lambda) = L(x,\lambda) + \frac{c}{2}\|e(q,u)\|^2_{L^2(\Omega)} \ .$$

(x^*,λ^*) is a saddle point of $L_c(x,\lambda)$ if and only if it is a saddle point of $L(x,\lambda)$. We use the following algorithm to compute a saddle point of L_c. This algorithm is studied in detail in [KT97] when H^2 norm is used for the regularization term.

Algorithm 2 .

Step 1: *Choose* $u_0, \lambda_0, c > 0$ *and set* $n=1$.

Step 2: *Solve* $q_n \in K$ *from* $q_n = \arg\min_q L_c(q, u_n, \lambda_{n-1})$, *which gives:*

$$\begin{aligned} &\beta F'(q_n) + B^*_{u_n} C_0^{-1}\lambda_{n-1} \\ + \quad &cB^*_{u_n} C_0^{-2}(B_{u_n} q_n - g) = 0 \ . \end{aligned} \tag{0.13}$$

Step 3: *Solve $u^n \in S_h$ from $u^n = \arg\min_u L_c(q_n, u, \lambda_{n-1})$, which gives:*

$$\begin{aligned} & u^n - u_d + A_{q_n} C_0^{-1} \lambda_{n-1} \\ + \quad & c A_{q_n} C_0^{-2} (A_{q_n} u_n - g) = 0 \ . \end{aligned} \tag{0.14}$$

Step 4: *update λ_n as*

$$\lambda_n = \lambda_{n-1} + c e(q_n, u_n) \ . \tag{0.15}$$

and go to Step 2 for $n = n + 1$.

It is easy to see that (0.13) and (0.14) have unique solutions q_n and u_n. If the observation error is zero or very small, it can be proved that $\{q_n, u_n, \lambda_n\}$ converges to the solution of (0.10)–(0.12) with first order accuracy. In more general situations, we can prove:

Theorem 3..1 *Suppose q_n or a subsequence of q_n is in the interior of K, then there exists a subsequence $\{q_{n_k}, u_{n_k}, \lambda_{n_k}\}$ of $\{q_n, u_n, \lambda_n\}$ such that*

$$\{q_{n_k}, u_{n_k}, \lambda_{n_k}\} \to \{\tilde{q}, \tilde{u}, \tilde{\lambda}\} \ ,$$

and $\{\tilde{q}, \tilde{u}, \tilde{\lambda}\}$ satisfies system (0.10)–(0.12).

Proof. $q_n \in K$ is a solution of (0.13) actually means

$$\begin{aligned} \beta(F'(q_n), q - q_n) + (C_0^{-1}\lambda_n, B_{u_n}(q - q_n)) \\ + c(B_{u_n}^* C_0^{-2}(B_{u_n} q_n - g), q - q_n) \\ \geq 0, \quad \forall q \in K. \end{aligned} \tag{0.16}$$

However, if $q_n \in \text{int } K$, then it satisfies (0.13). Combining (0.13) with (0.15), we see that

$$\begin{aligned} \beta F'(q_n) \quad + \quad & B_{u_n}^* C_0^{-1} \lambda_n \\ + \quad & B_{u_n}^* C_0^{-1} (\lambda_{n+1} - \lambda_n) = 0 \ , \end{aligned} \tag{0.17}$$

and so:

$$\beta F'(q_n) \quad + \quad B_{u_n}^* C_0^{-1} \lambda_{n+1} = 0 \ . \tag{0.18}$$

Similarly, substituting (0.15) into (0.14) gives:

$$u^n - u_d + A_{q_n} C_0^{-1} \lambda_{n+1} = 0 \ . \tag{0.19}$$

Checking on (0.18) and (0.11), one finds that

$$\beta F'(q_n) + B_{u_n}^* C_0^{-1} A_{q_n}^{-1} (u^n - u_d) = 0, \tag{0.20}$$

which implies:

$$\beta(F'(q_n), q_n) + (C_0^{-1} A_{q_n}^{-1} (u^n - u_d), B_{u_n} q_n) = 0. \tag{0.21}$$

However, $B_{u_n} q_n = A_{q_n} u_n$, and so

$$\beta(F'(q_n), q_n) + (C_0^{-1} A_{q_n}^{-1} (u^n - u_d), A_{q_n} u_n) = 0. \tag{0.22}$$

Since q_n is bounded and positive, relation (0.22) indicates that $\{u_n\}$ must be bounded, and (0.19) again implies that $\{\lambda_n\}$ must be bounded. Thus, there exists a subsequence $\{q_{n_k}, u_{n_k}, \lambda_{n_k}\}$ such that

$$\{q_{n_k}, u_{n_k}, \lambda_{n_k}\} \rightarrow \{\tilde{\lambda}, \tilde{u}, \tilde{\lambda}\}.$$

In our finite dimensional situation, by passing to their limit for equations (0.10)–(0.12) one finds that $\{\tilde{q}, \tilde{u}, \tilde{\lambda}\}$ satisfies (0.10)–(0.12).

Remark 3..2 . *We note that for a given u_n, we can obtain a unique q_n from (0.13) and for a given q_n, we get a unique u_n from (0.14). Equations (0.13) and (0.14) are coupled. We use an inner iteration between (0.13) and (0.14) to solve these coupled equations.*

Remark 3..3 . *Due to the term $F'(q_n)$, equation (0.13) is nonlinear. In our computations, we use a fixed point iteration similar to a method suggested in [VO96].*

Remark 3..4 . *For a given parameter q, we can choose k_1 sufficient small and k_2 sufficient large in the constraint set K. Thus, if the algorithm is convergent, then $q_n \in int\, K$. This is observed in our computations. However, this condition is difficult to verify theoretically.*

4. Numerical Experiments

We shall test our algorithm on some one dimensional problems. Let $f(x)$ be a known function. For any given $q \in P_h$, let $u \in S_h$ be the solution of

$$(q\nabla u, \nabla v)_\Omega = (f, v)_\Omega, \ \forall v \in S_h,$$

on $\Omega = [0, 1]$. The observation u_d is obtained from

$$u_d = u + \delta * R_d * \max_\Omega |u|,$$

where δ is a constant and R_d is a vector of uniformly distributed random number in $[-1, 1]$ and $\dim(R_d) = \dim(u_d)$.

In using Algorithm 1, it is necessary to choose a good initial guess for u_0 and λ_0. A natural choice is $u_0 = u_d$ and $\lambda_0 = 0$. However, u_d contains random errors. When the mesh size h is very small, then $||u - u_d||_{H^1(\Omega)} = O(h^{-1})$. Thus the matrix B_{u_0} has very large entries. In order to obtain a smoother initial value u_0, we shall try to find a w such that it is the solution of

$$-\varepsilon\nabla \cdot \left(\frac{\nabla w}{\sqrt{|\nabla w| + \varepsilon}} \right) + (w - u_d) = 0 \ . \qquad (0.23)$$

This has the effect of smoothing u_d while preserving the sharp discontinuities, see [CGM96]. After computing the solution for (0.23), we take it as the initial value u_0. We find that a smooth initial value which is also close in L^2 norm to u_d is critical for getting an accurate identified parameter. In some cases, use the function w as the observation for u can produce a better identified parameter.

In the examples that follow, we shall present the computational results by using our algorithm. In the figures, $q(x)$ is the true parameter, $q_n(x)$ is the identified parameter, $u_n(x)$ is the identified state, $u_d(x)$ is the observation and u_h is the true state u.

Example 1. In this example, we try to identify the parameter $q(x) = c_1(x)e^x$ from an observation without noise. The function $c_1(x)$ is piecewise constant, i.e. $c_1(x) = 1$ in intervals $[\frac{i}{7}, \frac{i+1}{7}]$, $i = 0, 2, 4, 6$ and $c_1(x) = 10$ in intervals $[\frac{i}{7}, \frac{i+1}{7}]$, $i = 1, 3, 5, 7$. As no noise is present, the output error and the equation constraint can be minimized to zero. Accordingly, we take a large constant for c and a very small value for ε. The computational result presented in Figure 1 is obtained with $h = 1/200$, $\varepsilon = h^5$, $c = 10^6$, $\beta = 10^{-8}$. In this example and the following ones, we carry out two inner loops between step 2 and step 3 of Algorithm 2 before we go to step 4 to update the multiplier. From the figure we see that the identified parameter $q_n(x)$ is indistinguishable from the true parameter $q(x)$. In this example, and also the following ones, the identified parameter is oscillatory in the first few iterations, but the oscillation disappears after a few iterations.

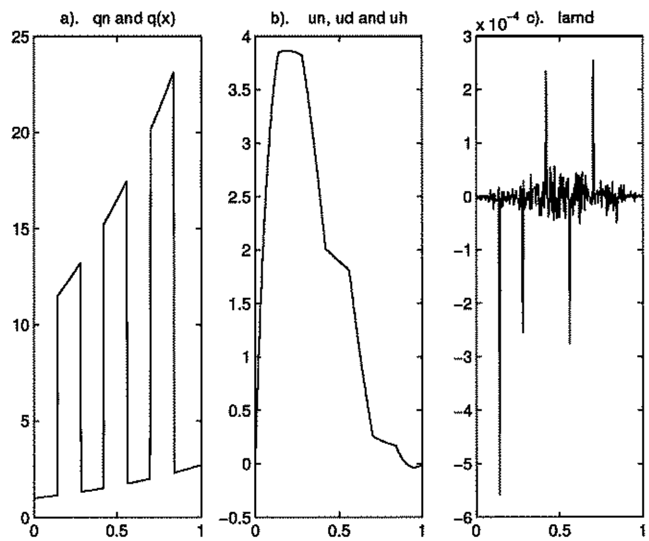

Figure 1 The computed result for Example 1. Noiseless case. Iteration number $n = 10$.

Example 2. In this example, we identify a piecewise constant coefficient from an observation with noise. We take $\delta = 10^{-3} = 0.1\%$. The other constants are taken as: $h = 1/200$, $\varepsilon = 10^{-4}$, $\beta = 10^{-4}$, $c = 10^2$. The identified parameter after 20 iterations are given in Figure 2. We see that, even in the presence of noise, the location of the discontinuities are located extremely accurately but the fine details are not recovered as well as in the noiseless case.

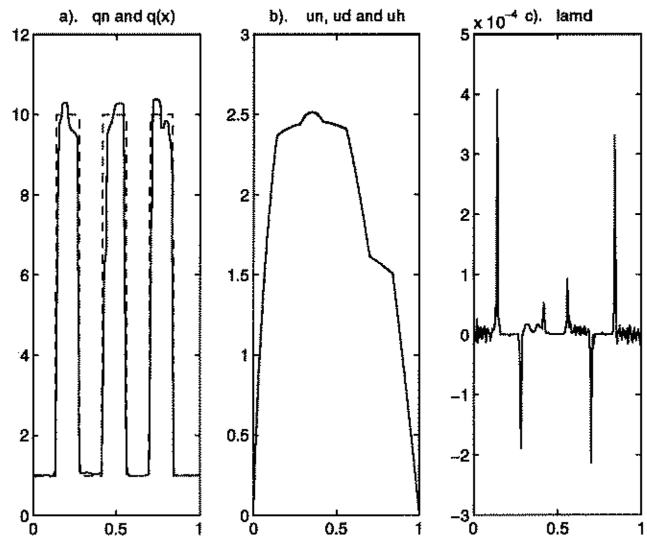

Figure 2 The computed result for Example 2. Noisy case. Iteration number $n = 20$.

Example 3. We try to identify the same parameter of Example 2, but we increase the noise level to $\delta = 10^{-2} = 1\%$. From Figure 3, we see that the location of the discontinuity of the coefficient is successfully identified, but the fine details are not identified as well as the less noisy cases. In the computation, we have taken $h = 1/300$, $\beta = 10^{-3}$, $c = 10^2$, $\varepsilon = 10^{-2}$.

Example 4. When the proposed algorithm is used to identify continuous parameters, it can allow much larger observation errors. We try to identify parameter $q(x) = e^{(2x-1)^2}$ from an observation with 10% of noise, i.e. we take $\delta = 0.1 = 10\%$. The identified parameter is shown in Figure 4. The constants are taken as $h = 1/200$, $\varepsilon = 0.1$, $\beta = 10^{-3}$, $c = 10$. We observe that the recovered coefficient q_n is a bit oscillatory but the general shape of q is recovered relatively well given the large observation errors.

5. Conclusions

By using the augmented Lagrangian method, we avoided the using of minimization algorithms in solving the inverse problem. Instead, only linear algebraic systems need to be solved in the iteration procedure. This enables us to use a larger number of unknowns and use much less CPU time in comparison with minimization algorithms. By combining the augmented Lagrangian method with total variation techniques, we can recover discontinuous coefficients with sharp jumps.

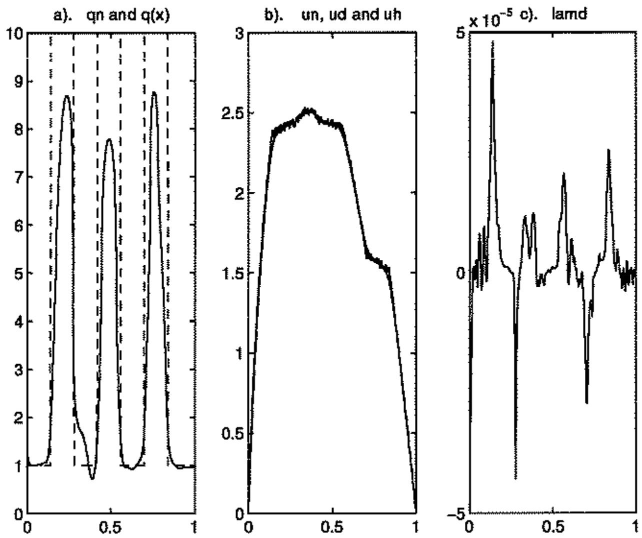

Figure 3 The computed result for Example 3. Noisy case. Iteration number $n = 20$.

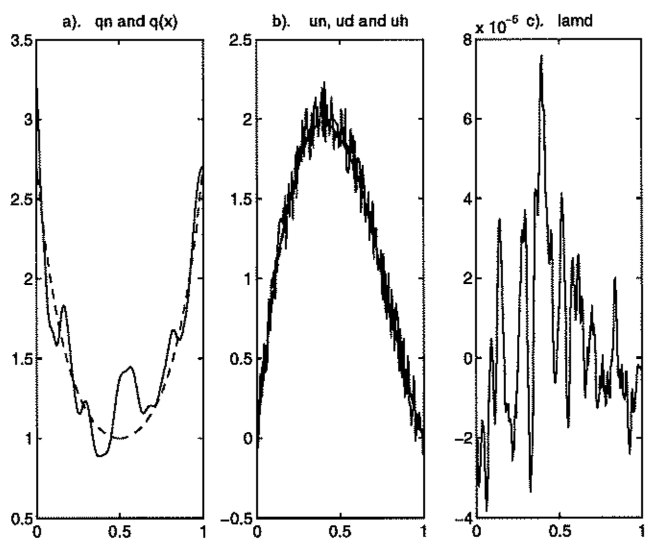

Figure 4 The computed result for Example 4. Iteration number $n = 20$.

REFERENCES

[ALM92] Alvarez L., Lions P. L., and Morel J. M. (1992) Image slective smoothing and edge detection by nonlinear diffusion. II. *SIAM J. Numer. Anal.* 29: 845–866.

[CGM96] Chan T. F., Golub G. H., and Mulet P. (1996) A nonlinear primal-dual method for total variation-based image restoration. In *Proc. of ICAOS'96, 12 th Int'l Conf. on Analysis and Optimization of Systems: Images, Wavelets and PDE's, Paris, June 26-28, 1996, M. Berger et al (eds)*, number 219 in Lecture Notes in Control and Information Sciences, pages 241–252.

[DS94] Dobson D. C. and Santosa F. (1994) An image enhancement technique for eletrical impedance tomography. *Inverse problems* 10: 317–334

[Ewi83] Ewing R. (ed) (1983) *The mathematics of reservoir simulation.* SIAM, Philadelphia.

[Giu84] Giusti E. (1984) *Minimal surfaces and functions of bounded variations.* Birkhäuser.

[Glo84] Glowinski R. (1984) *Numerical methods for nonlinear variational problems.* Springer-Verlag.

[IK90] Ito K. and Kunisch K. (1990) The augmented lagrangian method for parameter estimation in elliptic systems. *SIAM J. Control Optim.* 28: 113–136.

[KT97] Kunisch K. and Tai X.-C. (1997) Sequential and parallel splitting methods for bilinear control problems in Hilbert spaces. *SIAM J. Numer. Anal.* 34: 91–118.

[VO96] Vogel C. and Oman M. (1996) Iterative methods for total variation denoising. *SIAM J. Sci. Comp.* 17: 227–238.

[Zie89] Ziemer W. P. (1989) *Weakly differentiable functions.* Springer-Verlag.