

On level set regularization for highly ill-posed distributed parameter estimation problems

K. van den Doel* U. M. Ascher†

December 22, 2005

Abstract

The recovery of a distributed parameter function with discontinuities from inverse problems with elliptic forward PDEs is fraught with theoretical and practical difficulties. Better results are obtained for problems where the solution may take on at each point only one of two values, thus yielding a shape recovery problem.

This article considers level set regularization for such problems. However, rather than explicitly integrating a time embedded PDE to steady state, which typically requires thousands of iterations, methods based on Gauss-Newton are applied more directly. One of these can be viewed as damped Gauss-Newton utilized to approximate the steady state equations which in turn are viewed as the necessary conditions of a Tikhonov-type regularization with a sharpening sub-step at each iteration. In practice this method is eclipsed, however, by a special “finite time” or Levenberg-Marquardt-type method which we call *dynamic regularization* applied to the output least squares formulation. Our stopping criterion for the iteration does not involve knowledge of the true solution.

The regularization functional is applied to the (smooth) level set function rather than to the discontinuous function to be recovered, and the second focus of this article is on selecting this functional. Typical choices may lead to flat level sets that in turn cause ill-conditioning. We propose a new, quartic, non-local regularization term that penalizes flatness and produces a smooth level set evolution, and compare its performance to more usual choices.

Two numerical test cases are considered: a potential problem and the classical EIT/DC resistivity problem.

*Department of Computer Science, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada. (kvdool@cs.ubc.ca). Supported in part under IRIS grant on Reality Based Modeling and in part under NSERC Discovery Grant 84306.

†Department of Computer Science, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada. (ascher@cs.ubc.ca). Supported in part under NSERC Discovery Grant 84306.

Keywords: Inverse problem, Dynamic regularization, Level set, Gauss-Newton, Levenberg-Marquardt, Electrical impedance tomography, Potential, DC resistivity

1 Introduction

Consider the following data inversion problem. A *forward operator*, $F(m)$, is given, and a *model* $m(\mathbf{x})$ is sought over a discretized domain Ω in 2D or 3D, such that $F(m)$ matches given data b up to the noise level in the data measurements. The forward model is further given by

$$F(m) = Qu, \quad (1a)$$

$$u = G(m), \quad (1b)$$

where Q is a matrix which projects to data locations (e.g. along the boundary $\partial\Omega$), and $G(m)$ is the inverse of an elliptic PDE system discretized on a grid at least as fine as that of m using a finite volume or finite element method.

Several applications give rise to such a problem formulation. These include DC resistivity [35], linear potential problems [23, 7], magnetotelluric inversion [30], diffraction tomography [12], oil reservoir simulation [14] aquifer calibration [17], electrical impedance tomography (EIT) [5, 10, 6], and Maxwell's equations in low frequencies [26, 27, 20, 21].

It is well-known that while the forward problem (1) is well-posed the inverse problem is not. Indeed, the applications mentioned above and those considered in this article are *highly* ill-posed. In practice for the available noisy data typically there is no unique solution, i.e., there are many models m which yield a field u such that Qu is close to b to within the noise level, and moreover, such models m may vary wildly and depend discontinuously on the data. A direct application of the *output least squares method*, which is to solve the optimization problem

$$\min_m \phi_0 = \frac{1}{2} \|F(m) - b\|^2 \quad (2)$$

using the least squares norm of the data fitting term, typically runs into trouble. In a Tikhonov-type regularization [36], therefore, one approximately solves the optimization problem

$$\min_m \phi_\beta = \frac{1}{2} \|F(m) - b\|^2 + \beta R(m) \quad (3)$$

where $R(m)$ is a regularization term, and $\beta > 0$ is the regularization parameter whose choice has been the subject of many papers (see, e.g., [37]).

For the regularization term, we have considered in previous articles a same-grid discretization of

$$R(m) = \int_{\Omega} (\rho(|\nabla m|) + \hat{\alpha}(m - m_{\text{ref}})^2) d\mathbf{x} \quad (4)$$

where $\hat{\alpha}$ is a (typically very small, positive) parameter and m_{ref} is a given reference function (typically, the half-space solution). A least squares regularization is achieved by choosing $\rho(\tau) = \frac{1}{2}\tau^2$, although we have used also a weighted least squares penalty function, see [19, 1, 2, 21, 4].

However, the least squares functional is well-known to be unsuitable if a priori information that the model m contains *discontinuities* is to be respected. Total variation regularization has been proposed and successfully applied to denoising and mildly ill-posed problems such as deblurring [31]. In [4, 3] we have discussed and developed further the use of modified total variation (TV), or (occasionally slightly better) Huber switching between TV and least squares. But we also demonstrated that these methods may fail when applied to highly ill-posed problems such as those considered in this article. Specifically, examples showed that simply trusting the reconstruction because the data misfit is small cannot be advocated, even if the discrepancy principle is obeyed and even under the unrealistic assumption that data is available everywhere. It can then be argued that displaying a smooth blob, such as when using weighted least squares regularization, is less committing and more truthful to the quality of the actual information at hand than displaying a discontinuous solution. For recovering sharp discontinuous solutions more a priori information is required.

Such additional information is available if we know that the model function $m(\mathbf{x})$ is piecewise constant. In fact, let us assume in this article that m may only take on one of two known values, m_1 and m_2 (say, a homogeneous body and a homogeneous background). The problem becomes that of shape optimization, and a level set approach [34, 29, 8] for incorporating this additional information is natural.

Following [32, 24, 18] and others we consider $m(\mathbf{x})$ as a function of a smoother one $\psi(\mathbf{x})$ and apply regularization to ψ ,

$$\min_{\psi} \phi = \frac{1}{2} \|F(m) - b\|^2 + \beta R(\psi), \quad (5a)$$

$$m = \chi(\psi). \quad (5b)$$

The function χ is (a grid-smoothing of) the characteristic function based on the values that m may take. We choose, e.g.,

$$\chi(s) = \frac{m_1 - m_2}{2} \tanh(s/h) + \frac{m_1 + m_2}{2}. \quad (5c)$$

Note that χ depends on the grid resolution h , and

$$\lim_{h \rightarrow 0} \chi(s) = \begin{cases} m_2 & s < 0 \\ m_1 & s > 0 \end{cases}.$$

Thus, sharpening happens across the 0-level set of $\psi(\mathbf{x})$.

Selecting the regularization functional R is one focus of the present paper. The regularization should capture the idea that ψ is smooth, but it should not be too

flat near its 0-level, so that the interface will not change significantly upon a minor perturbation in ψ . We consider the form

$$R(\psi) = \hat{R}(|\nabla\psi|) + \alpha R_{TV}(|\nabla m|) \quad (6)$$

with R_{TV} a suitably modified version (e.g., [31, 4]) of

$$R_{TV} = \int_{\Omega} |\nabla m(\psi)| d\mathbf{x} \quad (7)$$

penalizing the length of the level set interface [16, 9, 11], and $\alpha \geq 0$ is yet another parameter to be selected. The choice of \hat{R} is considered in Section 2. In addition to the usual quadratic term we end up recommending a non-local quartic term which mildly and smoothly penalizes the distance of $|\nabla\psi|$ from the value 1 and appears to produce a smoother level set evolution.

More usual level set formulations appear in the literature as a function of (artificial) time. In fact, already the formulation (3) can be generalized as follows: Denoting the sensitivity matrix $J = \frac{\partial F}{\partial m}$, the necessary conditions for the optimization problem (3) can be written as the steady state equations for the time-dependent problem

$$\begin{aligned} M(m) \frac{\partial m}{\partial t} &= -[J^T(F(m) - b) + \beta R'(m)], \\ m(0) &= m_0, \end{aligned} \quad (8)$$

where $t \geq 0$ is the artificial time variable and the preconditioner M is positive definite. A forward Euler discretization of (8) with a special choice of time step in fact coincides with a preconditioned steepest descent method for (3). However, these methods as such are not known for their efficiency. It is more important to realize that special cases such as $M = I$, $\beta = 0$, and $M = \beta I$, $\beta \rightarrow \infty$, have regularizing effects for a *finite time*, i.e., without reaching the steady state solution (see, e.g., [37]). These choices (as well as $M = 0$ and others where M involves an anisotropic term) have all been considered in practice and yield various instances of Tikhonov-type regularization and scale space methods [33, 24, 16]. We prefer the interpretation (8),(9) over the connection made in these references between the different resulting methods.

The method described in (5) now generalizes into

$$\begin{aligned} M(\psi) \frac{\partial \psi}{\partial t} &= -[\hat{J}^T(\hat{F}(\psi) - b) + \beta R'(\psi)], \\ \psi(0) &= \psi_0, \end{aligned} \quad (9)$$

where we set

$$\hat{F}(\psi) = F(m(\psi)), \quad \hat{J}(\psi) = \frac{\partial \hat{F}}{\partial \psi} = J\chi'. \quad (10)$$

Variants of this have been considered in [13, 16, 9] and elsewhere. We caution, however, that in some such approaches involving explicit discretization of the time-dependent PDE thousands of iterations (or time steps) are required before convergence is deemed to have been achieved, often by unclear criteria.

In the present article we consider two techniques that can be viewed as modifications of the *Gauss-Newton* method (see, e.g., [28]). In the first, we apply a damped Gauss-Newton method for the solution of (5). In other words, we solve directly for the steady state of (9) with a positive β , rather than following the dynamics of the time-dependent system. Convergence is typically obtained in 20 iterations and less. Implementation issues and results are discussed in Section 4.

Unfortunately, the relatively large magnitude of the updates implied by the small number of iterations often causes the method to “overshoot”: In experiments where we know the true solution we typically observe a rapid approach towards that solution which is not followed by clearly indicated convergence. The question of how to stop such an iterative procedure (for problems where we do not know the true solution!) becomes tricky to resolve.

Better results in the latter respect are obtained by a different, *dynamic regularization* approach: Starting with the output least squares formulation, i.e., (5) with $\beta = 0$, we apply a damped Levenberg-Marquardt version of the Gauss-Newton method, with the iteration regularized by the Hessian of the same operators \hat{R} as above. This iteration is carried out only to “finite time” in the language of (9); indeed, the method can be viewed as an instance of (9) with $\beta = 0$ and M given by (27) below. Its advantage in practice is that typically ϕ decreases consistently until some point where the iteration starts stalling, whereby the solution process can be reasonably stopped. In general, theory for level set methods for our inverse problems is currently in an unsatisfactory state, and the present article does nothing to improve upon this, but our computational effort consistently yields pleasing results using this dynamic regularization. We elaborate on it further in Section 3.

In Section 4.1 we describe our implementation of the methods. Following the development of suitable algorithms we present the results of two experiments supporting our findings. The first, described in Section 4.2, involves a simple looking potential problem [23, 16] which still yields tough inverse problems. The typical total running time for our method is well below 10 seconds, and that is significantly faster than solving the forward problem for thousands of steps. The second, described in Section 4.3, involves a multi-source experiment for a nonlinear DC resistivity problem. Conclusions detailing our contributions are offered in Section 5.

2 Tikhonov Regularization

At first, let us note that the problem

$$\hat{F}(\psi) = b \tag{11}$$

in the continuum limit ($h \rightarrow 0$) is invariant under the transformation $\psi(\mathbf{x}) \rightarrow \lambda(\mathbf{x})\psi(\mathbf{x})$ for any continuous, bounded, positive (or negative) function $\lambda(\mathbf{x})$. Therefore, even if the original problem $F(m) = b$ were well-posed, the level set formulation (11) is not.¹

Let us rewrite (5) as

$$\min_{\psi} \phi = \phi_e(\psi) + \beta R(\psi), \quad (12a)$$

$$\phi_e = \frac{1}{2} \|\hat{F}(\psi) - b\|^2. \quad (12b)$$

For a suitable regularization functional R the parameter β must be appropriately tuned for the problem at hand, for example using the discrepancy principle or by error-free methods [15]. The choice of R is usually aided by a priori knowledge about the system. Since the level set formulation inherently has a large null-space, reflecting the fact that only the 0-level set plays a role in the problem, an additional role of R now is to remove this null-space, thereby shaping the level set function ψ in some desirable form. A common goal is to approximate a signed distance function.

A well-used choice for \hat{R} is the discretized form of

$$\hat{R}_2 = \frac{1}{2} \int_{\Omega} |\nabla \psi|^2 d\mathbf{x}. \quad (13)$$

However, while this term penalizes excessive roughness it also encourages flat level set functions, which may cause m to change drastically upon a small change in ψ .

It has been suggested by various authors (e.g., [16, 9]) to add to this regularization operator \hat{R}_2 a term whose gradient depends on curvature, as in (6). In particular, the (modified) total variation term (7) penalizes the length of the level set interface. Note that we cannot have just (7) by itself as regularizer, because (7) depends explicitly only on the interface and therefore does not remove the level set null-space. Unfortunately, this means that we now have another parameter α besides β to choose.

The motivation for adding a term (7) to \hat{R}_2 is to penalize fragmentation of the recovered shape. Another reason is to regularize the inverse problem itself, rather than just the level set formulation with its large null-space. While it is not always clear that this is really necessary for the quadratic term (13), the quartic regularization term introduced below has a minimum which is independent of the 0-level set of ψ and will therefore need additional regularization, possibly by a term of the form (7).

Both \hat{R}_2 and R_{TV} have a drawback in that they admit very flat functions ψ which may hover around the 0-level (that determines the interface). The usual technique for preventing flatness of the level set is to periodically restart the iterative procedure

¹Note that sometimes in the literature the term “level set method” implies the use of a signed distance function. Other methods such as those based on a double well potential in ψ are sometimes called phase-field methods; see for instance [8]. In this paper we call any method that uses the 0-level set of a function to describe a boundary a “level set method”.

to evolve the level set function re-initializing ψ to be approximately a signed distance function [29]. The iterative methods we describe here usually require just a few iterations, however, and therefore we seek a regularization term that achieves the dual goals of regularizing the ill-posed problem and of encouraging the level set function to vary smoothly and not be flat near the interface.

An “ideal” level set function is a signed distance function, and will have $|\nabla\psi(\mathbf{x})| = 1$ almost everywhere. A natural choice for \hat{R} may therefore seem to be the quartic penalty form

$$\hat{R}_4 = \frac{1}{2} \int_{\Omega} (1 - |\nabla\psi|^2)^2 d\mathbf{x}. \quad (14)$$

In practice we have found that (14) sometimes works, provided one is extremely careful with the discretization procedure, but there are several reasons why it is not entirely satisfactory. First, in order to converge to a particular solution starting from some initial configuration, (14) may prevent ψ from moving from a configuration with, say, one connected region, to a configuration with two regions, because artificial potential barriers are created. Indeed, one may argue that we should add (14) to the objective function with a *large* coefficient β , effectively trying to solve the problem with the added constraint $|\nabla\psi| = 1$. However, this constraint makes it very difficult, and occasionally impossible in our experiments, to find the minimum of the objective function since we are quite limited in updates we can make to ψ . Even if convergence is obtained, a relatively large number of iterations is required to pass the artificial potential barrier.

Second, the discretization picks up spurious terms at the locations where ψ changes direction, which happens at the points that are equidistant from two or more 0-level sets. At such points $\nabla\psi$ has a discontinuity that, unless miraculously aligned with the grid, introduces spurious contributions into the objective function ϕ . During the optimization procedure these contributions may push the field ψ in the wrong direction, thereby denying smooth convergence to the solution.

Finally, (14) is more difficult to handle from an implementation perspective, since it is non-convex. The Hessian of (14) is given by

$$\hat{H}_4 = 2 \sum_{\mathbf{x} \in \Omega_h} (\psi^T L_{\mathbf{x}} \psi - 1) L_{\mathbf{x}} + 4 \sum_{\mathbf{x} \in \Omega_h} (L_{\mathbf{x}} \psi)(L_{\mathbf{x}} \psi)^T, \quad (15)$$

with ψ a function on the grid Ω_h reshaped as a vector. The positive definite, local grid operator $L_{\mathbf{x}}$ is constructed such that $\psi^T L_{\mathbf{x}} \psi$ is the discretized form of $|\nabla\psi(\mathbf{x})|^2$ for each grid point \mathbf{x} . A sufficient, though not necessary, condition for the positive definiteness of the Hessian (15) is

$$\psi^T L_{\mathbf{x}} \psi > 1, \quad \forall \mathbf{x} \in \Omega_h. \quad (16)$$

Clearly this Hessian is more complicated than the Hessian of \hat{R}_2 , which is just the discrete Laplacian.

Since the insistence on $|\nabla\psi(\mathbf{x})| = 1$ at every point \mathbf{x} is just to prevent difficulties due to flat level set functions hovering around zero, we can certainly afford to relax this requirement somewhat and try to encourage ψ to behave well only in an average sense. Thus, we really only want

$$\int_{\Omega} K(\mathbf{x} - \mathbf{y}) |\nabla\psi(\mathbf{y})|^2 d\mathbf{y} = 1, \quad (17)$$

where $K(\mathbf{x})$ is a smearing function (a Gaussian for example) with some support to be determined empirically. Our regularization term now becomes

$$\frac{1}{2} \int_{\Omega} \int_{\Omega} (1 - |\nabla\psi(\mathbf{x})|^2) \hat{K}(\mathbf{x} - \mathbf{y}) (1 - |\nabla\psi(\mathbf{y})|^2) d\mathbf{x} d\mathbf{y}, \quad (18)$$

with \hat{K} the convolution of K with itself.

If we choose a Dirac delta function for $K(\mathbf{x})$ then (14) is recovered. At the other extreme where we smear over the entire domain, i.e., $K(\mathbf{x}) = 1/|\Omega|$ with $|\Omega| = \text{volume}(\Omega)$, we obtain the regularization term

$$\hat{R}_{4n} = \frac{1}{2} \left[\int_{\Omega} (1 - |\nabla\psi|^2) d\mathbf{x} \right]^2. \quad (19)$$

This penalizes flatness in ψ only in the average sense, and yet it proves to be quite effective in our experiments as described in Section 4.4. In fact, the best results in our experiments have been obtained with this non-local term (19). The Hessian of (19) is

$$\hat{H}_{4n} = 2(\psi^T L\psi - |\Omega|)L + 4(L\psi)(L\psi)^T \quad (20)$$

with $L = \hat{H}_2$ a positive definite matrix based on the standard 5-point discrete Laplacian. A sufficient condition for \hat{R}_{4n} to be positive definite is that

$$\psi^T L\psi > |\Omega|. \quad (21)$$

This is a much milder condition than the pointwise condition (16). In our experiments negative eigenvalues of the Hessian were never encountered during the iteration, so long as the initial ψ satisfies (21).

Note that both quartic terms have minima which are independent of the 0-level set, in the continuum, because any interface can be described as the 0-level set of a signed distance function which has $\hat{R}_{4n} = \hat{R}_4 = 0$. This implies that additional regularization is required for the interface problem, which can be achieved through a term R_{TV} or through dynamic regularization as discussed following (8) and in Section 3 below.

As detailed in Section 4 we have observed experimentally that quite frequently a very good solution (as judged by the eye norm) is obtained after a few iterations,

especially using (6) with \hat{R}_{4n} and also with \hat{R}_2 . But detecting where to stop the iteration without the aid of the true solution is hard; upon continuing to iterate the solution overshoots, and the actual solution to the minimization problem (12) is not as good. If β and/or α are too small the solution overshoots by fitting the model to the noise. If they are too large, the solution overshoots by being pushed in the wrong direction by the regularization term. At the experimentally “optimal” β and α , with which the solution converges with a misfit as dictated by the discrepancy principle, we have observed that usually a much better solution is obtained before convergence, but with a misfit larger than dictated by the discrepancy principle. In addition, considerable fine-tuning of the additional parameter α is occasionally required to obtain the best fit, which is an uncertain procedure since all we have to go on is the misfit.

We observe in summary that a Tikhonov regularization of the error functional leaves something to be desired in the present context.

3 Dynamic regularization

Let us consider the minimization problem (12) with $\beta = 0$. Instead of the regularization term βR , we now rely on the “finite time” corresponding to the limited number of iterations defined below to provide the regularization effect. The approach is similar to what is called iterative regularization in [15, 33, 25]. The most common stopping criterion is again the discrepancy principle. The usual criticism against this approach, namely, that it does not take a priori information into account in the same essential way that a Tikhonov-type regularization can, is muted in our level set setting for a shape optimization problem as described in the closing paragraphs of the previous section.

The Levenberg-Marquardt approach linearizes $\hat{F}(\psi)$ in ϕ_e and adds a penalty $\beta_0 \delta\psi^T X \delta\psi = \beta_0 \|\delta\psi\|_X^2$ to the equations for the update $\delta\psi$, with X some appropriate positive definite matrix. Let us write

$$g_e(\psi) = \frac{\partial \phi_e}{\partial \psi} = \hat{J}^T(\psi)(\hat{F}(\psi) - b)$$

and

$$H_e = \hat{J}^T \hat{J}, \tag{22}$$

which is the Gauss-Newton approximation to the Hessian of ϕ_e . We assume here that the problem has been discretized on an $N \times N$ grid so that ψ is represented by a grid function which can be reshaped into an N^2 -vector. With this notation the Levenberg-Marquardt method defines the updates for ψ by the equation

$$(H_e(\psi) + \beta_0 X(\psi)) \delta\psi = -\gamma g_e(\psi), \tag{23}$$

where X and β_0 are to be chosen, and $0 < \gamma \leq 1$ is a damping parameter determined by line search. Certain choices for X were considered in [7] in a somewhat different algorithmic framework. Specifically, a Hamilton-Jacobi sub-step is required at each iteration of [7] and our best choice for X , which turns out to be (20), is not considered there.

The parameter β_0 has a purely algebraic role here in ensuring that the eigenvalues of the resulting matrix are bounded away from 0. The result therefore should not be very sensitive to the choice of β_0 . Note that β_0 is considered to be constant, whereas the usual trust-region based Levenberg-Marquardt technique adaptively chooses β_0 at each iteration (with $\gamma = 1$) depending on the assessed region where the local linearized model can be trusted. Here, X has the function of removing the large near null-space of H_e . Perhaps a better interpretation of (23) is simply as a regularized version of the damped Gauss-Newton method.

Comparing this to the update rule for a damped Gauss-Newton scheme for the Tikhonov regularized problem (12) which reads

$$(H_e(\psi) + \beta R''(\psi))\delta\psi = -\gamma(g_e(\psi) + \beta R'(\psi)), \quad (24)$$

it is tempting to choose $X = R''$ in (23) and interpret $R(\psi)$ in a similar fashion as in the Tikhonov regularization. This interpretation is justified because we can interpret the iterative scheme (23) as the steps in a perturbation series expansion around the minimum of the Tikhonov functional (12) at β_0 towards $\beta = 0$, as described next.

Suppose for a moment that we know the solution, denoted by ψ_0 , to (12) for a particular value of $\beta = \beta_0$. For example, β_0 may be a value which is large enough such that the problem is sufficiently regularized to be smooth and well-behaved. We now write the solution for generic β (thought of as being “close” to β_0) as a power series in $\beta - \beta_0$. Using standard methods we can write

$$\psi = \psi_0 + \sum_{i=1}^{\infty} (\beta - \beta_0)^i y_i,$$

with y_i to be determined. Define also

$$\psi_k = \psi_0 + \sum_{i=1}^k (\beta - \beta_0)^i y_i.$$

We can interpret ψ_k as being the solution to (12) accurate to order k in $(\beta - \beta_0)$. Next, impose the optimality condition

$$\frac{\partial \phi_e}{\partial \psi} = 0$$

on ψ_k , keeping only terms to order k :

$$g_e(\psi_k) + (\beta - \beta_0)R'(\psi_k) + \beta_0 R'(\psi_k) = 0. \quad (25)$$

Using Taylor’s expansion up to linear terms on (25) and keeping again only terms to order k ,

$$g_e(\psi_{k-1}) + H_e(\psi_{k-1})(\psi_k - \psi_{k-1}) + (\beta - \beta_0)R'(\psi_{k-1}) + \beta_0 R'(\psi_{k-1}) + \beta_0 R''(\psi_{k-1})(\psi_k - \psi_{k-1}) = 0,$$

which can be rewritten (bumping k up by 1 for notational ease) as an iterative scheme:

$$(H_e(\psi_k) + \beta_0 R''(\psi_k))(\psi_{k+1} - \psi_k) = -(g_e(\psi_k) + \beta R'(\psi_k)). \quad (26)$$

For $\beta = 0$ we obtain the Levenberg-Marquardt scheme (23) with $X = R''$, $\gamma = 1$.

Another connection between X in (23) and R'' in (24) can be made by considering again a solution ψ to (12) for a given β and deriving an ODE for ψ as a function of $t = \beta^{-1}$. Multiplying ϕ by t , the necessary conditions for optimality in (12a) are

$$t g_e + R' = 0.$$

Differentiating with respect to t and then dividing by it gives

$$(H_e + \beta R'') \frac{\partial \psi}{\partial t} + \beta g_e = 0.$$

A forward Euler discretization with step size τ gives the update rule

$$(H_e + \beta R'') \delta \psi = -\tau \beta g_e,$$

and this yields (23) upon setting $X = R''$, $\beta = \beta_0$ and $\tau = \gamma/\beta_0$. Note again that β_0 need not be small and that t , hence τ , can be stretched arbitrarily. Note also that the differential system $(H_e + \beta_0 R'') \frac{\partial \psi}{\partial t} = -g_e$ is a special case of (9) with $\beta = 0$ and

$$M = H_e + \beta_0 R''. \quad (27)$$

On the other hand, our choice of the step size τ has little to do with ODE methods.

In our experiments, as detailed in Section 4, the most remarkable difference from the Tikhonov regularization of Section 2 is that with dynamic regularization no overshooting has occurred. Instead, the present method appears to stall at just the right time, when the solution looks best according to the eye norm! Based on this we propose a practical stopping criterion by monitoring the misfit

$$\text{misfit} = \|F(m) - b\|/\|b\| \quad (28)$$

and the reconstructed shapes, and stopping the iteration when the algorithm seems to have stalled, or when the misfit drops below some small multiple $\mu > 1$ of the theoretically optimal value based on a known noise level of the data, i.e., the discrepancy principle. (See also [22].) Stalling occurs occasionally before the data misfit has reached the critical value. In the Tikhonov case, too small a choice for μ is a disaster as the solution will immediately overshoot, but here it just means that the algorithm stalls and the misfit only decreases very slowly. If we then continue for thousands of iterations the method will of course eventually overshoot, but this is prevented by the “stalling” based stopping criterion.

An important additional advantage of the present algorithm is that there is no need for an interface penalty term (7), as the “finite time” (i.e., the limited number of iterations) provides the necessary shape regularization, so we set $\alpha = 0$ in (6).

4 Numerical experiments

4.1 Implementation details

For simplicity the inverse problem is considered on a uniform, $N \times N$ grid. Our reported experiments use $N = 32$, although we have experimented with $N = 64$ as well. The evaluation of the forward operator F of (1) uses the same $N \times N$ grid (except when generating artificial data, see below), although a finer grid can be easily utilized instead.

For each modified Gauss-Newton iteration we have to solve a linear system of the form

$$(\hat{J}^T \hat{J} + \beta R'') \delta\psi = -\gamma p, \quad (29)$$

where $p = \nabla\phi$. For the line search procedure determining γ we use a step-halving scheme and utilize the Wolfe conditions to decide if sufficient decrease in ϕ has been obtained [28].

We have

$$\hat{J} = \frac{\partial \hat{F}}{\partial \psi} = \frac{\partial F(m)}{\partial m} \frac{\partial \chi(\psi)}{\partial \psi},$$

where $\frac{\partial \chi(\psi)}{\partial \psi}$ is a grid-smoothed version of a Dirac delta function. Further, on the interface we apply thresholding to give it a support of $4h$ with h the grid spacing. This makes the matrix $\hat{J}^T \hat{J}$ in (29) sparse. For $R'' = L$ the entire matrix in (29) is subsequently sparse. For $R'' = \hat{H}_{4n}$ of (20) there is an additional rank-1 term that is dealt with as such using the Sherman-Morrison formula. With this, for small grids of up to 64×64 a direct solution method using MATLAB's backslash proves rather efficient: Some specific performance data are given towards the end of Section 4.4. Note that the construction of $\hat{J}^T \hat{J}$ requires the solution of $O(N)$ PDEs differing only in their right hand sides (see, e.g., (34) below). This is done quite efficiently by a single sparse LU decomposition followed by $O(N)$ sparse forward and backward substitutions. In our implementation the assembly of $\hat{J}^T \hat{J} + \beta R''$ takes typically 10 times longer than the solution of (29). The cost of a single iteration of this scheme is not much higher (about 3-4 times more in our implementation) than the cost of a single step using a first order method such as gradient descent, a method that typically requires thousands of iterations to yield satisfactory results.

To generate data for the test problems described below we discretize the given PDE ((30) or (34)) on a finer $(2N - 1) \times (2N - 1)$ grid. Solving for a “true model” m and downsampling back to an $N \times N$ grid gives u of (1b). The purpose of the finer grid is to avoid so-called “inverse crimes”. Then Q of (1a) is applied and the result is sprinkled with Gaussian noise.

In the following we present two test problems. This is followed in Section 4.4 by numerical results.

4.2 A potential problem

Here we follow [16] in using a model problem from [23] to test and demonstrate our methods. Consider the Poisson equation

$$\begin{aligned}\Delta u &= m, & \mathbf{x} \in \Omega, \\ u|_{\partial\Omega} &= 0,\end{aligned}\tag{30a}$$

where m is a characteristic function

$$m(\mathbf{x}) = \begin{cases} m_2 & \mathbf{x} \in D \\ m_1 & \mathbf{x} \notin D \end{cases},\tag{30b}$$

with $m_1 = 0$, $m_2 = 1$, and $D \subset \Omega = [0, 1]^2$.

The inverse problem is to recover the shape D from the normal derivative of u on the boundary, $[\frac{\partial u}{\partial \nu}]_{\partial\Omega}$. A physical interpretation is for example the recovery of a mass distribution of a single material from measurements of the gravitational field.

Note that for any given m of the form (30b) there is a unique $u \in H^2(\Omega)$ which satisfies (30a). Since u is two derivatives smoother the inverse problem is highly ill-posed. For this problem, one set of Neumann data corresponding to the homogeneous Dirichlet case in (30a) furnishes as much information as the entire Dirichlet to Neumann map [23], so no multiple data sets are used.

We discretize (30a) on a uniform grid with spacing $h = 1/N$, employing the most straightforward discretization

$$h^{-2}[u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j}] = m_{i,j}, \quad 0 < i, j < N\tag{31}$$

and closing the system with the homogeneous Dirichlet BC. Denoting the resulting algebraic system

$$\Delta_h u = m$$

we have $u = \Delta_h^{-1}m$.

For the normal derivative on the boundary we use forward differencing, hence the predicted data are the values of $-u/h$ at the (interior) grid points adjacent to the boundary. Denote these by Qu , with the resulting Q a rather sparse, $4N \times (N-1)^2$ matrix. Thus, the forward problem is defined by

$$F(m) = Q\Delta_h^{-1}m, \quad \text{hence } J = Q\Delta_h^{-1}.\tag{32}$$

Note that F is linear in m . The additional fact that m may take on only two values must be dealt with (or rather, taken into account) separately.

We now formulate the inverse problem in terms of a level set function ψ which is related to the field m by $m = \chi(\psi)$, cf. (5).

4.3 EIT and DC resistivity

The prototype forward PDE

$$\nabla \cdot (\sigma \nabla u) = q, \quad \mathbf{x} \in \Omega, \quad (33)$$

arises in many applications. We consider it in resistivity form on the unit square, $\Omega = [0, 1]^2$, subject to natural (homogeneous Neumann) boundary conditions, for multiple data sets $q^i, i = 1, \dots, M$. This reads

$$\begin{aligned} \nabla \cdot (m^{-1} \nabla u^i) &= q^i, \quad i = 1, \dots, M, \\ \frac{\partial u^i}{\partial \nu} \Big|_{\partial \Omega} &= 0. \end{aligned} \quad (34)$$

The forward model is discretized on a staggered grid as described in [2], and the constant null-space is removed in a standard way. The inverse problem is to recover m , which at each grid point takes on the values $m_1 = 1$ or $m_2 = 10$, say, from measurements of u^i on the boundary.

It is well-known that a single configuration for q^i , corresponding physically to current sources injected into the system, usually does not provide enough information to reconstruct m . Instead, in practical applications, a number of different configurations for q^i are set up and boundary data are measured for each configuration. The more independent data we have, the better our chances for a faithful reconstruction are. In our example we configure q^i to consist of a positive point source on the left boundary and an opposite source on the right boundary:

$$q^i(\mathbf{x}) = \delta_{\mathbf{x}, \mathbf{p}_L^i} - \delta_{\mathbf{x}, \mathbf{p}_R^i},$$

where \mathbf{p}_L^i and \mathbf{p}_R^i are located on the left and right boundaries. Different data sets are obtained by varying the positions \mathbf{p}_L^i and \mathbf{p}_R^i of the two opposing sources. We place the left source at s equidistant points including the corners, and similarly for the right source, in all possible combinations. This gives a total of $M = s^2$ data sets.

Ignoring the piecewise constant property of m the resulting forward problem can again be clearly written as $F(m)$, where now F is a (smooth) nonlinear function of m .

4.4 Numerical results

Below is a representative sample of our experiments. First we show the best reconstructions we were able to obtain with the various methods. We have experimented with two data sets consisting of two and three squares depicted in Figure 1.

We have tested the dynamic regularization and Tikhonov-type regularization with the quadratic and the two quartic regularization terms. The local quartic term was generally found to be an inferior choice, because the resulting Hessian is often plagued

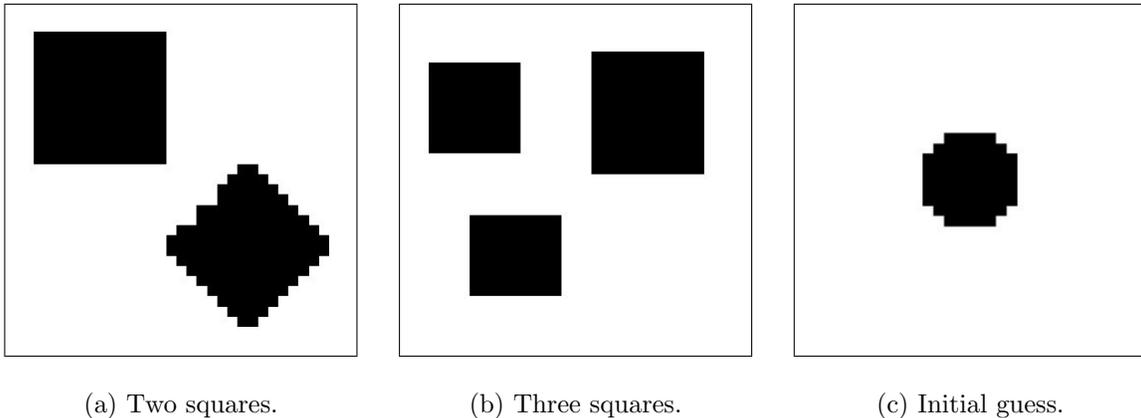


Figure 1: The two data sets, or “true models”, that were used in our experiments. The boundary data was polluted with 5% noise. We also display the initial guess used by the iterative methods.

by negative eigenvalues: When these are removed by some ad-hoc modification of the Hessian we are effectively no longer using \hat{R}_4 . Results are depicted in Figure 2.

These results were obtained after carefully tuning parameters for all combinations of regularization method, regularization functional, forward problem, and data set. The notation *3P2D* refers to reconstructing three squares for the potential problem, with the quadratic regularization term \hat{R}_2 and dynamic regularization; likewise, *2E4nT* refers to reconstructing two squares for the EIT problem, with the non-local quartic regularization \hat{R}_{4n} using the Tikhonov-type approach, etc. The regularization parameter(s) are indicated below these codes. The first refers to β (for Tikhonov) or β_0 (for dynamic regularization), the second, only used for Tikhonov, is α (see (6)). We also indicate the misfit, defined by (28), followed in parenthesis by its theoretical expected value, which is controlled by an input noise level. For the EIT problem we used one data set for the two squares problem and four data sets for the three squares problem, i.e., $s = 1$ or 2 respectively. For the two squares EIT adding a nonzero α prevented any splitting of the initial configuration into separate parts, so we set $\alpha = 0$. No reasonable reconstruction could be obtained with any of the parameter choices we have tried for the EIT problem with three squares using our Tikhonov-type program. Note that neither theoretical nor empirical reasons would lead one to expect to be able to reconstruct the exact interface, even with no noise.

To illustrate the noise sensitivity, we repeat experiments (d) and (m), i.e., the EIT problem with two squares and the \hat{R}_{4n} term employing dynamic and Tikhonov regularization, but now with 20% noise. Results are displayed in Figure 3. The Tikhonov reconstruction has overshoot its target, so β and possibly α must be increased to prevent this. However, since the misfit is close to ideal there is no practical way to detect this disaster unless the solution is known in advance. The dynamic reconstruction on

the other hand has stalled after 7 iterations and does not change much anymore.

As explained in Section 2, one of the motivations for using the quartic regularization term is to discourage flat level sets. We illustrate that this is a practical advantage in Figure 4, where the level set functions resulting from using \hat{R}_2 and \hat{R}_{4n} are depicted. Observe that the level set function of \hat{R}_2 is flatter near the interface than the one obtained using \hat{R}_{4n} , which could be the reason for the better performance of the non-local quartic term. However, we have also found many examples where neither function is particularly flat, and yet the quartic term almost always leads to a smoother level set function.

To illustrate that the same value of β_0 can be used at a wide range of noise levels, we show in Figure 5 the reconstruction of a non-convex shape with the potential problem inverted in the forward operator and dynamic regularization employed at various noise levels from 1% up to 50%. Reducing the noise level below 1% does not result in any improvement in the reconstructed shape. Also, the misfit remains at about 0.17 when the iterations stall in this case. Note that there is no reason to expect a precise reconstruction, even in theory: For the potential problem, the model $m(\mathbf{x})$ is not uniquely determined by boundary data. For example, the gravitational far-field of a hollow sphere is the same as that of a smaller solid sphere of the same mass.

In Figure 6 we show the shape evolution utilizing dynamic regularization towards the solution depicted in Figure 2(f). It is smooth and orderly.

In Figure 7 we illustrate the overshooting problem of the Tikhonov-type regularization. Visually the reconstruction looks best after 4 iterations. However, the misfit at this point gives no indication of success, so there is no way we could stop the iteration at this point if we did not know the solution already. In contrast, the solution with dynamic regularization depicted in Figure 2(f) has stalled and does not overshoot in further iterations.

Based on these results and on further numerical experimentation we conclude that in our setting:

1. The dynamic regularization produces better results than the Tikhonov-type regularization, as can be seen by comparing the corresponding reconstructions in Figure 2 (e.g., (d) vs (m)). The iteration counts are generally smaller than those reported in [7], although they are of the same order of magnitude.
2. The non-local quartic regularization term is often clearly superior to the usual quadratic term: Compare for example Figures 2(g) and 2(h). The quartic term produces better behaved level set functions as illustrated in Figure 4.
3. The local quartic term does not perform well in practice; see Figure 2(n) for a typical example.

In addition, there are clear practical advantages for the dynamic regularization over the Tikhonov approach. Specifically:

1. The R_{TV} term is not needed: set $\alpha = 0$.
2. The results are not very sensitive to β_0 (cf. [22]), although more iterations are required if β_0 is increased significantly.
3. The iteration stalls at the appropriate time without overshooting. Since the explicit goal is shape reconstruction the use of the “eye-norm” to determine if the method has stalled or not is both practical and appropriate to the problems we are considering. A more rigorous and objective stopping criterion that is somehow equivalent to the eye-norm would of course very useful but is not straightforward and falls outside the scope of this paper.
4. The same value of β_0 can be used independent of the noise level.

For a 32×32 grid and one data set, one iteration typically requires $0.5s$ in MATLAB on a 3GHz Pentium IV based desktop computer. Thus, many of the runs reported above require well below 10 seconds of total CPU time. For a 64×64 grid the CPU time per iteration is about $2s$. The iteration time varies somewhat as it depends on the length of the interface, which influences the sparsity of the Hessian. The EIT problem on a 64×64 grid with 4 data sets requires about $5s$ per iteration. Of this, 78% is spent in assembling the Hessian and the gradient, and 5% is spent in solving (29).

5 Conclusions and further thoughts

Shape optimization of the type considered in this article, where the forward operator involves the inversion of an elliptic PDE, is notoriously difficult. These problems are highly ill-posed. Consequently, there are several papers in the literature where numerical experiments are synthesized with data assumed available throughout the spatial domain or at least a substantial part of it, and/or where the “true solution” is used in the algorithm itself to decide when the reconstruction is satisfactory. We have avoided here both of these temptations, because reality often does not provide for such luxury.

The data in our experiments are only provided at the domain’s boundary (occasionally in multiple sets, though). The proposed dynamic regularization method (23) with $X = R''$ (or, (9) with $\beta = 0$, (27) and forward Euler) performs very well under these conditions without consulting the “true solution”, using a “stalling termination criterion” instead. We have displayed obtained reconstructions for several examples involving both a potential problem and the classical EIT problem, the latter sometimes requiring multiple data sets. The noise levels in our synthetic experiments are not very small.

The reconstructions are particularly satisfactory when a novel regularization term, \hat{R}_{4n} , defined in (19), is employed. This provides for a smooth evolution of non-flat level set functions.

Our focus has been on numerical methods that terminate reasonably fast, requiring only relatively few iterations. On a typical 32×32 grid our MATLAB code requires a few seconds to converge. Indeed, for our better results fewer than 20 iterations were usually sufficient. Such performance cannot be expected from methods that require thousands of forward solves.

On the grids we have experimented with, standard direct sparse solvers are sufficiently fast. For the inverse problem system in 2D probably this is all that is required [5, 11]. An important advantage of direct methods is that they are relatively hassle-free and easy to implement (especially in MATLAB); for instance, the dependence of the condition number of the matrix in (29) on β is less crucial than it could be if iterative methods were utilized.

Forward problems may well require finer discretization grids, but fast iterative techniques for elliptic PDEs are well known. The efficient iterative solution of the inverse linear system (29) is much more challenging, and it is needed for larger problems, especially in 3D (see, e.g., [21, 19, 7]). We are currently considering this latter problem and expect to report our results in the near future.

In addition, we have demonstrated that the dynamic regularization algorithm requires no TV regularization term and is not very sensitive to β_0 , which in turn is insensitive to the noise level.

We cannot reconstruct shapes which are too complex, for example shapes that have holes, using our algorithm with similar ease. (This can be a very difficult task, though, as discussed towards the end of the previous section.) Moreover, the minimal number of data sets required for a practical EIT reconstruction is unclear and depends on the relevant true solution.

Finally, there is no full theoretical backing to our results. We hope, though, that our encouraging numerical experiments can shed some light on and be founded in a more solid theoretical framework in the future.

References

- [1] U. Ascher and E. Haber. Grid refinement and scaling for distributed parameter estimation problems. *Inverse Problems*, 17:571–590, 2001.
- [2] U. Ascher and E. Haber. A multigrid method for distributed parameter estimation problems. *J. ETNA*, 18:1–18, 2003.
- [3] U. Ascher and E. Haber. Computational methods for large distributed parameter estimation problems with possible discontinuities. *Proc. Symp. Inverse Problems, Design & Optimization*, pages 201–208, 2004. M. Colaco, H. Orlande and G. Dulikravich (eds.).
- [4] U. Ascher, E. Haber, and H. Huang. On effective methods for implicit piecewise smooth surface recovery. *SIAM J. Scient. Comput.*, 2006. To appear.

- [5] L. Borcea, J. G. Berryman, and G. C. Papanicolaou. High-contrast impedance tomography. *Inverse Problems*, 12:835–858, 1996.
- [6] L. Borcea, G. Gray, and Y. Zhang. Variationally constrained numerical solution of electrical impedance tomography. *Inverse Problems*, 19:1159–1184, 2003.
- [7] M. Burger. Levenberg-marquardt level set methods for inverse obstacle problems. *Inverse problems*, 20:259–282, 2004.
- [8] M. Burger and S. J. Osher. A survey on level set methods for inverse problems and optimal design. *European J. Appl. Math.: To Appear*, 2005.
- [9] T. Chan and X. Tai. Level set and total variation regularization for elliptic inverse problems with discontinuous coefficients. *J. Comp. Phys.*, 193:40–66, 2003.
- [10] M. Cheney, D. Isaacson, and J.C. Newell. Electrical impedance tomography. *SIAM Review*, 41:85–101, 1999.
- [11] E. Chung, T. Chan, and X. Tai. Electrical impedance tomography using level set representations and total variation regularization. *J. Comp. Phys.*, 205:357–372, 2005.
- [12] A. J. Devaney. The limited-view problem in diffraction tomography. *Inverse Problems*, 5:510–523, 1989.
- [13] O. Dorn, E.L. Miller, and C.M. Rappaport. A shape reconstruction method for electromagnetic tomography using adjoint fields and level sets. *Inverse Problems*, 16, 2000. 1119-1156.
- [14] R. Ewing (Ed.). *The mathematics of reservoir simulation*. SIAM, Philadelphia, 1983.
- [15] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer, 1996.
- [16] F Fruhauf, O. Scherzer, and A. Leitao. Analysis of regularization methods for the solution of ill-posed problems involving unbounded operators and a relation to constraint optimization. *SIAM J. Numer. Anal.*, 43:767–786, 2005.
- [17] S. Gomez, A. Perez, and R. Alvarez. Multiscale optimization for aquifer parameter identification with noisy data. In *Computational Methods in Water Resources XII, Vol. 2*, 1998.
- [18] E. Haber. A multilevel, level-set method for optimizing eigenvalues in shape design problems. *J. Comp. Phys.*, 198:518–534, 2004.

- [19] E. Haber and U. Ascher. Preconditioned all-at-one methods for large, sparse parameter estimation problems. *Inverse Problems*, 17:1847–1864, 2001.
- [20] E. Haber, U. Ascher, D. Aruliah, and D. Oldenburg. Fast simulation of 3D electromagnetic using potentials. *J. Comput. Phys.*, 163:150–171, 2000.
- [21] E. Haber, U. Ascher, and D. Oldenburg. Inversion of 3D electromagnetic data in frequency and time domain using an inexact all-at-once approach. *Geophysics*, 69:1216–1228, 2004.
- [22] Martin Hanke. A regularizing levenberg-marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse problems*, 13:79–95, 1997.
- [23] F. Hettlich and W. Rundell. Iterative methods for the reconstruction of an inverse potential problem. *Inverse Problems*, 12:251–266, 1996.
- [24] A. Leitao and O. Scherzer. On the relation between constraint regularization, level sets, and shape optimization. *Inverse Problems*, 19:L1–L11, 2003.
- [25] J. Nagy and K. Palmer. Steepest descent, cg and iterative regularization of ill-posed problems. *BIT*, 43:1003–1017, 2003.
- [26] G. Newman and D. Alumbaugh. Three-dimensional massively parallel electromagnetic inversion—I. theory. *Geophysical J. Int.*, 128:345–354, 1997.
- [27] G. Newman and D. Alumbaugh. Three-dimensional massively parallel electromagnetic inversion—II, analysis of a crosswell electromagnetic experiment. *Geophysical J. Int.*, 128:355–367, 1997.
- [28] J. Nocedal and S. Wright. *Numerical Optimization*. New York: Springer, 1999.
- [29] S. Osher and R. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2003.
- [30] R. L. Parker. *Geophysical Inverse Theory*. Princeton University Press, Princeton NJ, 1994.
- [31] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [32] F. Santosa. A level-set approach for inverse problems involving obstacles. *ESAIM Controle Optim. Calc. Var.*, 1:17–33, 1996.
- [33] O. Scherzer. Scale-space methods and regularization for denoising and inverse problems. *Advances in Image and Electron Physics*, 128:445–530, 2003.

- [34] J.A. Sethian. *Level Set Methods and Fast Marching Methods : Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision, and Material Sciences*. Cambridge, 1996.
- [35] N.C. Smith and K. Vozoff. Two dimensional DC resistivity inversion for dipole dipole data. *IEEE Trans. on geoscience and remote sensing*, GE 22:21–28, 1984.
- [36] A.N. Tikhonov and V.Ya. Arsenin. *Methods for Solving Ill-posed Problems*. John Wiley and Sons, Inc., 1977.
- [37] C. Vogel. *Computational methods for inverse problem*. SIAM, Philadelphia, 2002.

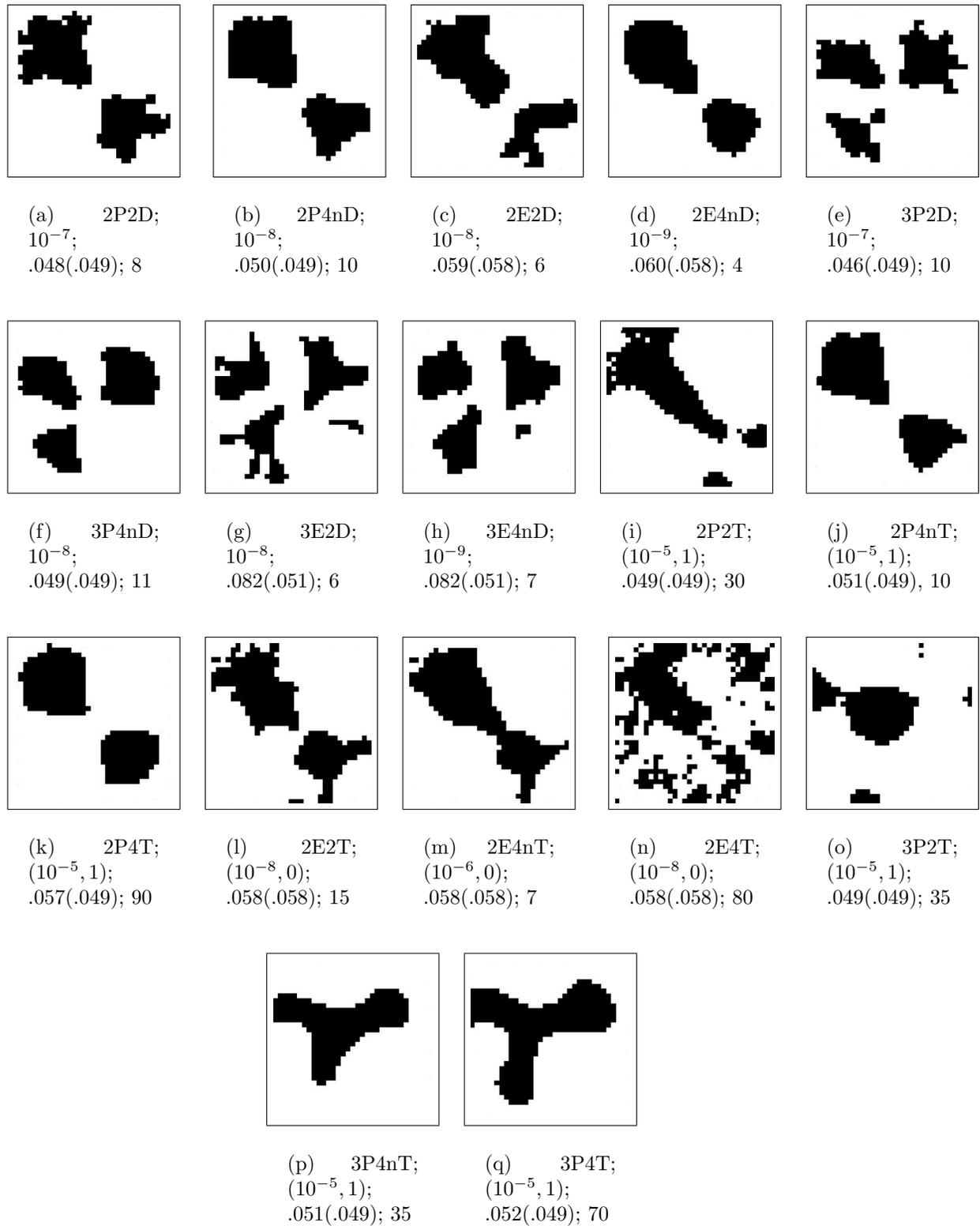


Figure 2: Best reconstructions. In each subfigure caption the code dLeM describes the selection of true model, forward problem, regularization term and regularization method; β and for Tikhonov also α follow; then misfit computed (theoretical); and finally the number of iterations. No reasonable shape was obtained for 3E2T, 3E4nT, and 3E4T.

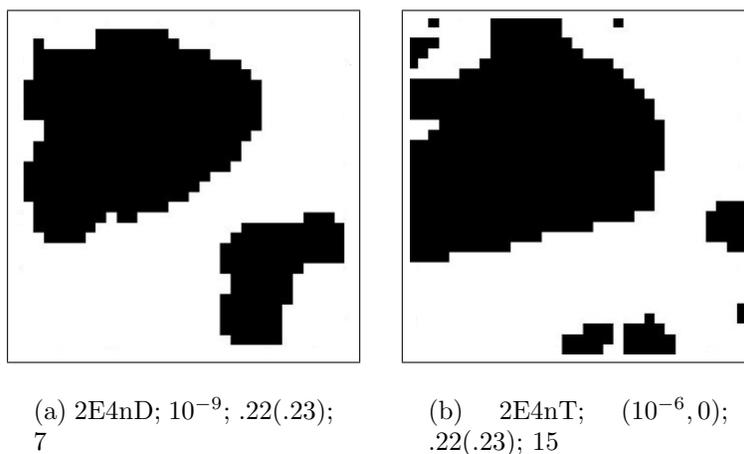


Figure 3: EIT experiments with 20% noise and the same parameters as with 5% noise in Figure 2. The dynamic reconstruction converges after about 7 iterations and does not overshoot.

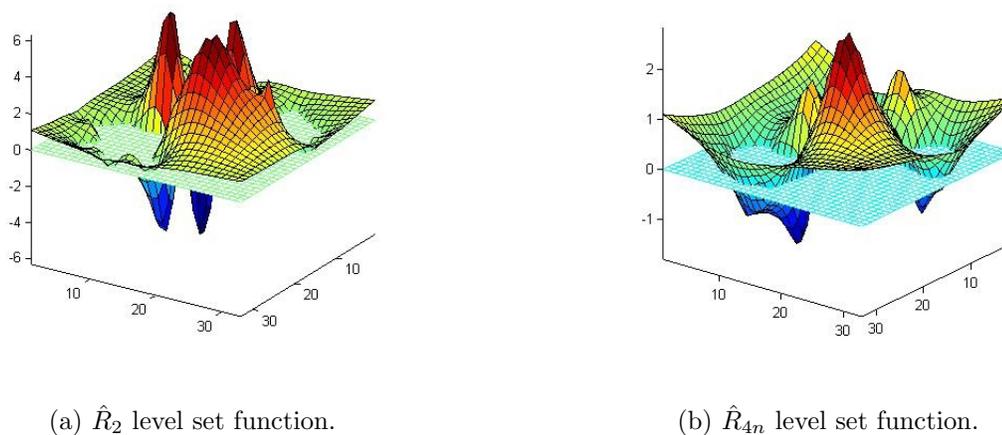


Figure 4: Level set functions for the dynamic regularization reconstruction of the three squares potential problem with the quadratic (left) and non-local quartic terms (right). These correspond to the solutions (e) and (f) in Figure 2. Observe the flatness generated by use of \hat{R}_2 as compared to the level set using \hat{R}_{4n} .

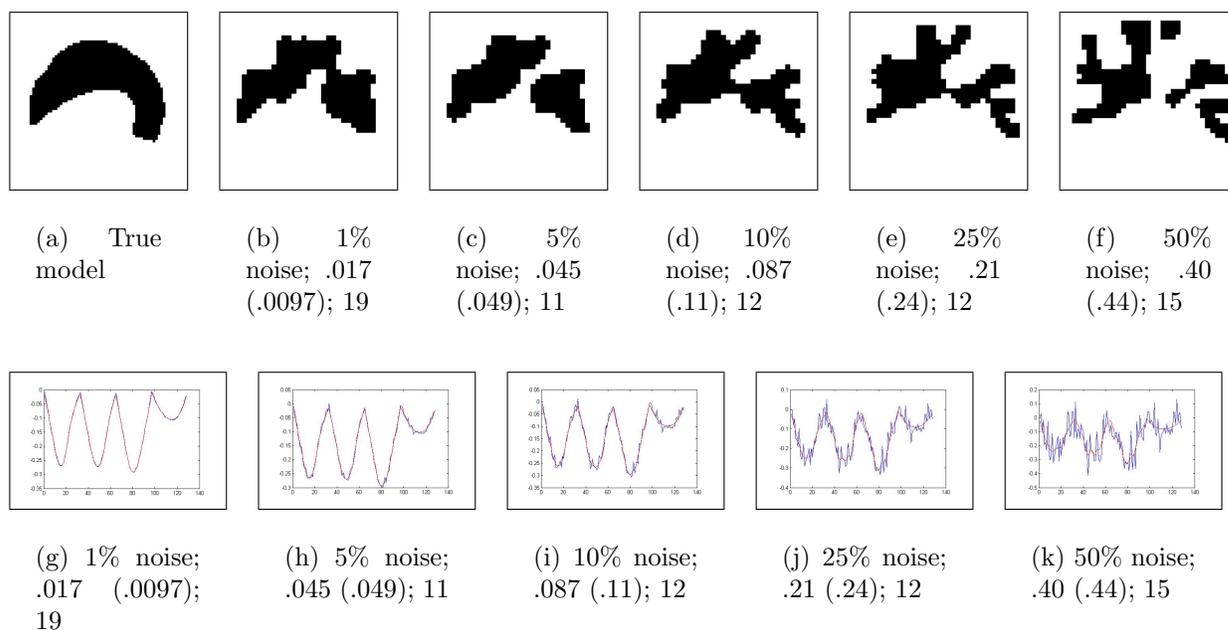


Figure 5: A non-convex shape reconstruction with dynamic regularization using the potential model at various noise levels for fixed $\beta_0 = 1.e - 8$. The non-local quartic form \hat{R}_{4n} was used. The misfit (and theoretical value) and number of iterations are indicated. We also show the actual (noisy curve) and the reconstructed boundary data. The forward fit of the predicted data is very good.

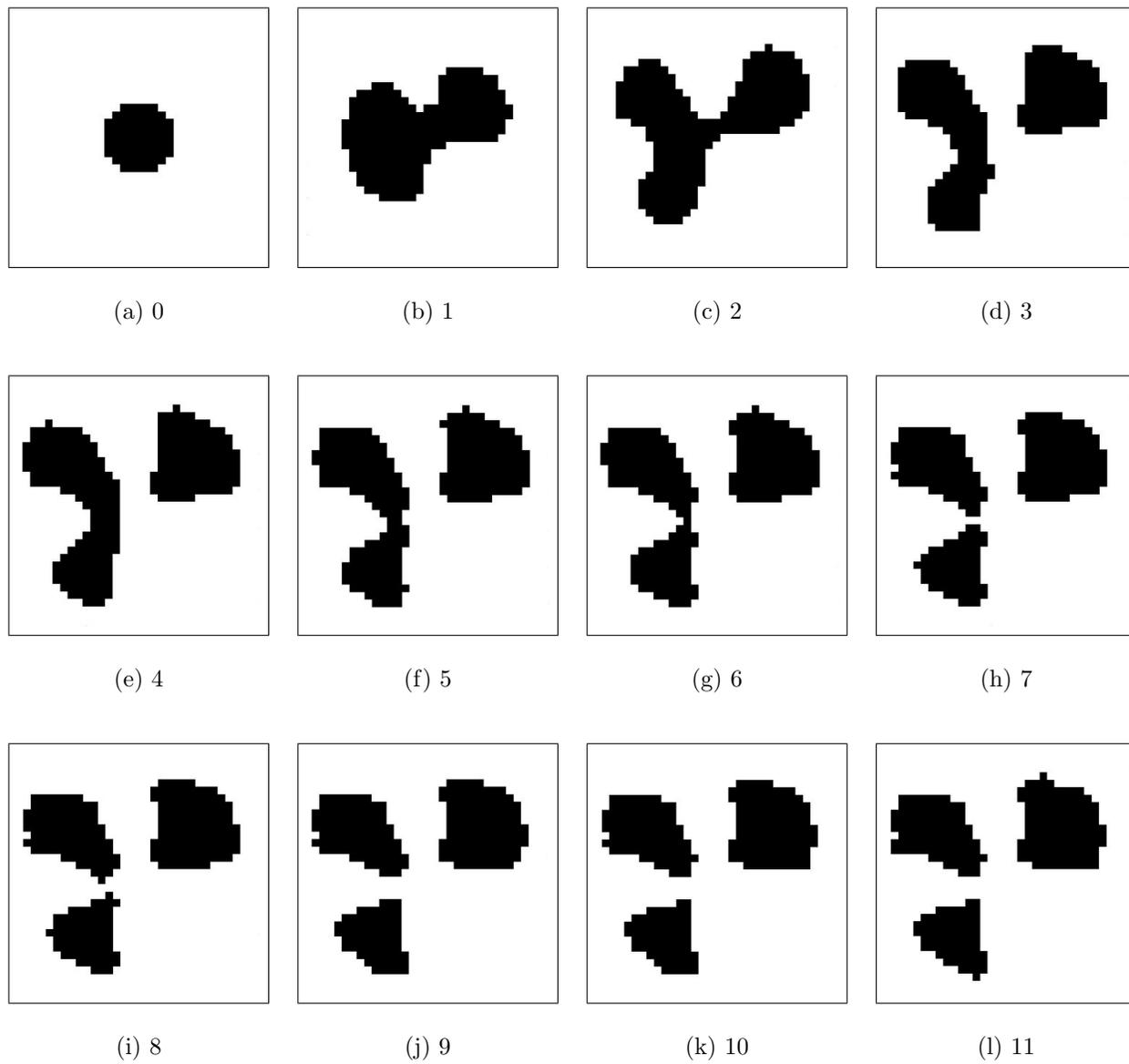


Figure 6: Shape evolution of the problem (f) of Figure 2.

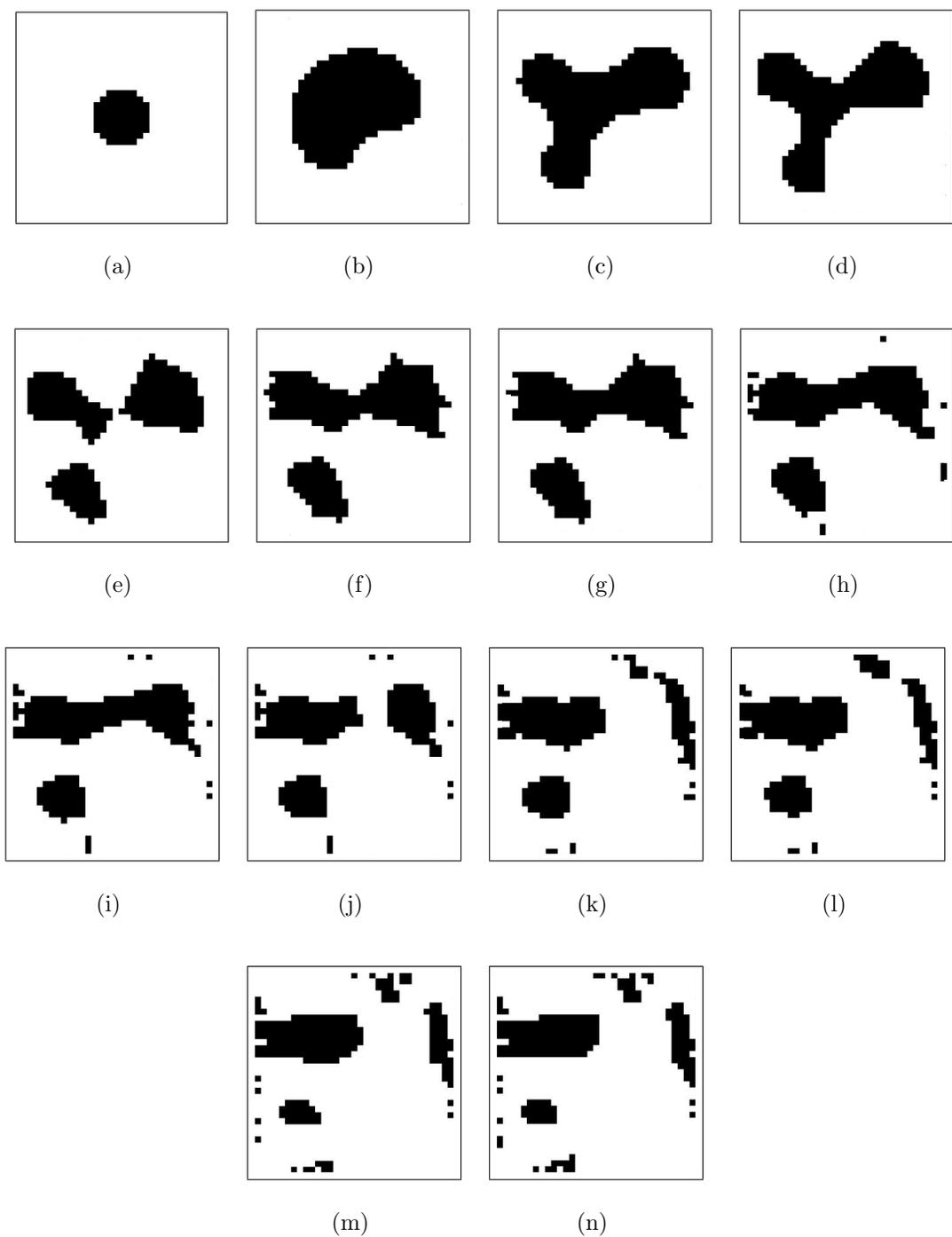


Figure 7: Tikhonov frequently overshoots as illustrated here. We show the iterations for the potential problem with three squares using Tikhonov with $\beta = 1.e - 7$, $\alpha = 1$. The final misfit is 0.034, theoretical value 0.049. After 4 iterations, when the reconstruction looks best, the misfit is 0.07. Compare with (f) and (p) in Figure 2.