An Abstract Analysis of Differential Semblance Optimization

Mark S. Gockenbach

June 29, 1994

Abstract Differential Semblance Optimization (DSO) is a novel way of approaching a class of inverse problems arising in exploration seismology. The promising feature of the DSO method is that it replaces a nonsmooth, highly nonconvex cost function (the Output Least–Squares (OLS) objective function) with a smooth cost function that is amenable to standard (local) optimization algorithms.

The OLS problem can be written abstractly as a partially linear least–squares problem with linear constraints. The DSO objective function is derived from the associated quadratic penalty function. It is shown that one way to view the DSO objective function is as a regularization of a function that is dual (in a certain sense) to the OLS objective function. By viewing the DSO problem as a perturbation of this dual problem, this method can be shown to be effective. In particular, it is demonstrated that, under suitable assumptions, the DSO method defines a parameterized path of minimizers converging to the desired solution, and that for certain values of the parameter, standard optimization techniques can be used to find a point on the path.

The predictions of the theory are motivated and illustrated on two simple model problems for seismic velocity inversion, the plane wave detection problem and the "layer–over–half–space" problem. It is shown that the theory presented in this thesis extends the existing theory for the plane wave detection problem.

Dedicated to my wife Joan, in appreciation for her love and support.

Acknowledgements: This work was partially supported by the Center for Research on Parallel Computation, DOE contract DE-FG03-93ER25178, the Texas Geophysical Parallel Computation Project, and The Rice Inversion Project. TRIP Sponsors for 1994 are Advance Geophysical, Amoco Production Co., Conoco Inc., Cray Research Inc., Exxon Production Research Co., Interactive Network Technologies, Mobil Research and Development Corp., and Texaco Inc.

1 Introduction

The goal of reflection seismology is to use measurements made at the earth's surface to determine the structure of the subsurface. The data typically consist of recordings of waves that originate at a controlled source at the surface and propagate through the subsurface. From these data, various parameters describing the rock properties are sought. The velocity at which the waves propagate is particularly important.

Since the data (the seismogram) can be predicted from the velocity and other parameters by means of accepted models, the determination of the parameters can be formulated as an inverse problem. The model parameters are to be chosen so the model predicts a seismogram as close to the observed seismogram as possible. A popular formulation of this problem is Output Least–Squares (OLS), in which the data are to be fitted in the L^2 sense. Differential Semblance Optimization (DSO), as proposed by Symes (see, for instance, [10], [11]), is an alternative to the least–squares approach. DSO is based on the quadratic penalty function method, and produces a parameterized family of optimization problems related to the least–squares problem.

In our abstract notation, these formulations are written as follows. The linear operator $A(x)$ depends nonlinearly on the variable x and acts as the "forward map", that is, for given model parameters x and y, $A(x)y$ represents the predicted data. The linear operator W defines a constraint; the only physically meaningful model parameters are those satisfying $Wy = 0$. The observed data are denoted by d. The "OLS" problem is to find x to solve

$$\min_{x} \quad J_\infty(x),$$

where

$$J_\infty(x) = \min\left\{\|A(x)y - d\|^2 : Wy = 0\right\}.$$

The "DSO" problem is to find x to solve

$$\min_{x} \quad J_r(x),$$

where

$$J_r(x) \quad = \quad \min\left\{\|A(x)y - d\|^2 + r^2\|Wy\|^2 : y \in Y\right\}.$$

The parameter r is the penalty parameter; the first problem is indexed by ∞ because it is a standard result that the quadratic penalty function approaches the constrained problem (in some suitable sense) as $r \uparrow \infty$. In the above definitions, the partially linear nature of the forward map is exploited to reduce the problem by minimizing first over the variable y. This approach requires the analysis of both of the minimizations, the inner minimization (over the variable y) and the outer minimization (over the variable x).

In the following section, we introduce the methods of Output Least–Squares and Differential Semblance Optimization, illustrating them on two simple problems. These problems can be easily stated and understood, yet it will be shown that they present nontrivial difficulties encountered in more realistic velocity inversion problems.

In Section 3, we establish the above abstract notation and discuss realistic hypotheses governing the inner minimization. Our aim in this section, as well as in Section 5, is to describe assumptions on our abstract optimization problem which can be verified for the motivating examples from velocity inversion, and which allow us to prove the results which validate the DSO method.

Section 4 is devoted to an analysis of the dependence of the inner minimization problem on r. Our analysis shows that there exists a constrained optimization problem related to the OLS inner problem. We will call this related problem the dual problem, although we have in mind a different notion of duality than is standard in linear and nonlinear programming. The DSO inner problem can be viewed as a regularization of this dual problem. With this point of view, we can describe the limiting behavior of the method as $r \downarrow 0$. We will show that, for "exact data", the penalty function converges to the dual problem as $r \downarrow 0$. We also study the behavior of the inner problem as $r \uparrow \infty$, which is important in understanding the relationship of the DSO function to the original OLS function.

In Section 5, we discuss the dependence of the forward map on the parameter x, which has a nonlinear influence. This aspect of the problem is unfortunately much harder to understand in the abstract. We will make rather strong assumptions, which can be verified in the motivating examples, and which allow us to prove significant results concerning the behavior of the solution to the outer minimization.

We then proceed to analyze the nature of the solutions to the outer problem in Section 6. We show that for "exact data", the original OLS problem, the dual problem, and the family of DSO problems all have the same solution. For data that has been perturbed by noise, we show that this single solution is perturbed to a parameterized path of solutions. The basis for the effectiveness of the DSO method lies in the fact that for small values of the penalty parameter r, the penalty function J_r is easy to minimize, and thus a point on the path can be located. As $r \uparrow \infty$, the path of solutions converges to the desired solution of the original problem. These results are proved in Section 6.

In Section 7, we return to the simple examples from Section 2, and show how our analysis applies. In particular, we show that the assumptions we made are satisfied. We also present the results of numerical experiments which illustrate the predictions of the theory.

2 Simple examples

2.1 Determining velocity and depth

The canonical problem of reflection seismology, which we call the layer–over–half–space problem, is to determine the depth of a reflector and the velocity of the (homogeneous) layer above it. Figure 2.1 shows the appropriate experiment: a source of waves is located at the surface, and a line of receivers record the vibrations produced by waves bouncing off the reflector. These recordings comprise the seismogram.

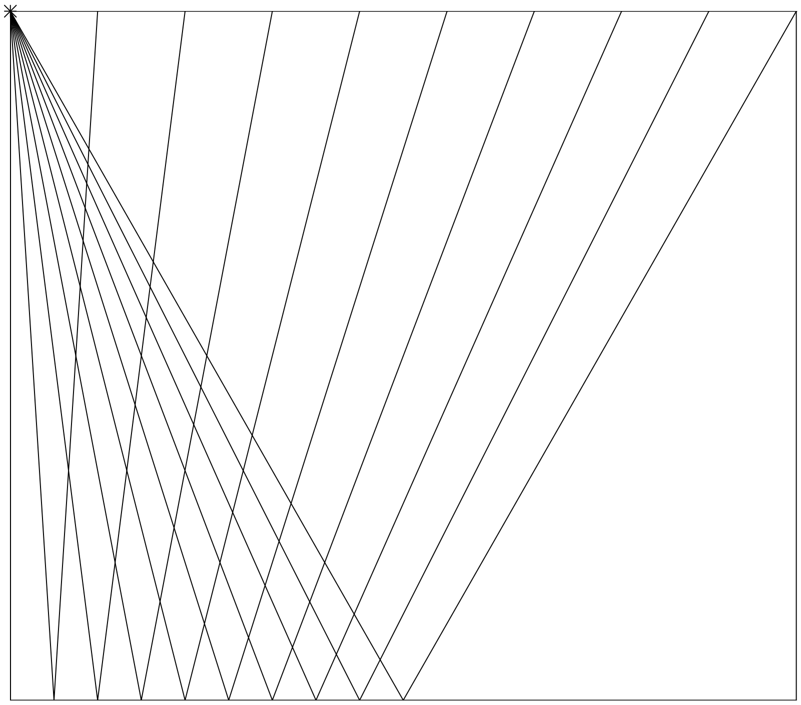

Figure 2.1: Seismic experiment

We use the following notation:

$$\begin{array}{rcl} c & = & \text{velocity of layer,} \\ d & = & \text{depth of reflector,} \\ f(t) & = & \text{source waveform,} \\ h & = & \text{half of the distance from the source to a given receiver.} \end{array}$$

We shall assume that the seismogram is recorded over the time–interval $[0, T]$ and that the receivers are located continuously on the interval $[0, 2H]$ (of course, in reality, the receivers would be at discrete locations in this interval). We define $\Omega = [0, H] \times [0, T]$. The seismogram is assumed to be a function $D \in L^2(\Omega)$.

For the purpose of this example, the following simplified model is assumed: given the velocity c, the depth d, and the source f, the predicted seismogram is

$$f\left(t - \frac{2\sqrt{d^2+h^2}}{c}\right).$$

This model captures correctly the kinematic aspects of the experiment, that is, the waves arrive at the receivers at the correct times, but it ignores the decrease in the amplitudes due to various effects.

2.1.1 Output least–squares

We define the *forward map* $\tilde{S} : U \to \mathcal{L}(L^2[0, T], L^2(\Omega))$ by

$$\tilde{S}(c, d)f(h, t) = f\left(t - \frac{2\sqrt{d^2+h^2}}{c}\right),$$

where $U \subset \mathcal{R}^2$ is the positive first quadrant: $U = \left\{(c, d) \in \mathcal{R}^2 : c > 0, D > 0\right\}$. The output least–squares problem is then

$$\min_{c,d,f} \quad \|\tilde{S}(c, d)f - D\|^2_{L^2(\Omega)}. \tag{2.1}$$

An obvious difficulty with this formulation is that the objective function is not in general differentiable with respect to c and d for an arbitrary L^2 function f. Moreover, suppose we require the source waveform f to lie in a more restricted space, such as a Sobolev space $H^k([0, T])$ (where k is chosen large enough to induce the desired number of derivatives in the objective function). Then the problem immediately becomes ill–posed when the data D is not correspondingly smooth (that is, if $D(h, t) = \bar{f}\left(t - \frac{2\sqrt{d^2+h^2}}{c}\right)$ for some $\bar{f}$ which belongs to L^2 but not to H^k, then (2.1) clearly has no solution in H^k) .

In order to get around the objections raised in the previous paragraph, we can proceed as follows. We insist that the data $D(h, t)$ be smooth, say twice-continuously differentiable, and write (2.1) as

$$\min_{c,d} \quad \tilde{J}(c, d), \tag{2.2}$$

where

$$\tilde{J}(c, d) = \min_{f} \quad \|\tilde{S}(c, d)f - D\|^2_{L^2(\Omega)}. \tag{2.3}$$

It is not difficult to find the f that solves this inner minimization problem for a given c and d:

$$f_\infty[c,d](t) = \frac{1}{H}\int_0^H D\left(h, t + \frac{2\sqrt{d^2+h^2}}{c}\right)\,\mathrm{dh}.$$

Moreover, under the assumption that D is smooth, $f_\infty[c,d]$ is also smooth, which gives us a smooth objective function J. Hence (2.2) can be solved using smooth optimization techniques.

Unfortunately, the fact that we have produced a smooth optimization problem does not guarantee that the problem can be solved easily. In fact, the objective function $\tilde{J}$ tends to have certain unfavorable characteristics. As an example, the values of $\tilde{J}(c,d)$ were computed for exact data generated by a source waveform equal to a Ricker wavelet (with peak frequency 50 Hz.), with velocity $c = 5m/ms$ and $d = 500m$. (Ricker wavelets are commonly used to to generate synthetic seismic data. A Ricker wavelet of frequency f Hz. is the function $2\lambda(2\lambda(t-t_0)^2-1)\exp(-\lambda(t-t_0)^2)$, where $\lambda = (f\pi/1000)^2$ and time t is measured in milliseconds.) There were 51 receivers located at half–offsets $0, 20, \ldots, 1000$ meters, and the recording interval was $[0, 1024]$ milliseconds. A contour plot of the objective function for this problem is shown in Figure 2.2. From this plot, we can see that the problem tends to be somewhat ill–conditioned and very nonconvex. In particular, outside of the valley which contains the solution, the objective function tends to be very flat. The one-dimensional scan of $\tilde{J}(c,d)$ in Figure 2.2 shows this clearly (here the depth is held fixed at the correct value and the velocity is varied).

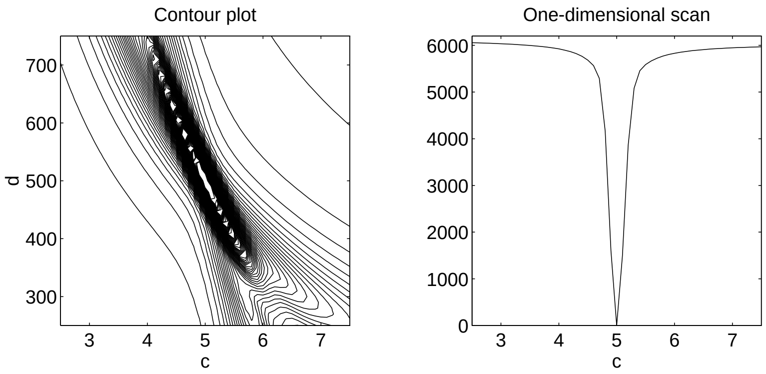

Figure 2.2: $\tilde{J}(c,d)$ for a Ricker source

The output least–squares approach thus can produce objective functions which dip very sharply down to the solution and which are otherwise rather flat. This leads to local, nonglobal minimizers when the data is noisy. For instance, we computed values of $\tilde{J}$ for the example above with random noise added to the data. A contour plot of the results is shown in Figure 2.3; this plot shows several stationary points besides the solution. A one–dimensional scan of $\tilde{J}$ is also shown in Figure 2.3.

The existence of local, nonglobal minimizers is troublesome because the most efficient algorithms for finding minimizers are local techniques which are in no way guaranteed to find the global minimizer as opposed to local, nonglobal minimizers. In fact, these algorithms tend to converge to the nearest (in some sense) minimizer, local or global.

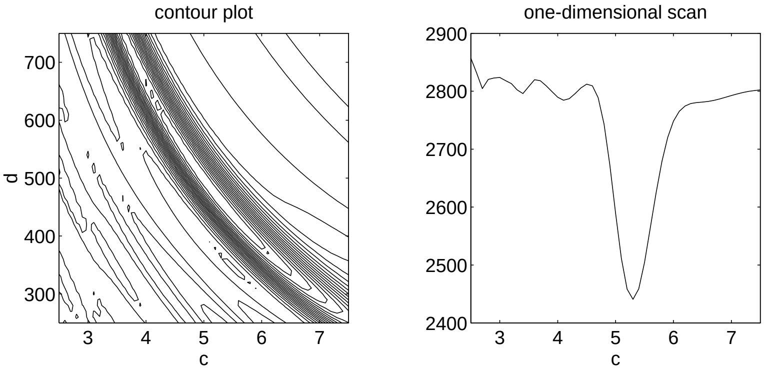

Figure 2.3: $\tilde{J}(c,d)$ for noisy data

In recent years, several algorithms intended to perform genuine global optimization have been designed and investigated. Prominent among these are the Simulated Annealing and Genetic Algorithms. These are algorithms which combine some type of random search with an acceptance criterion which allows some steps to be accepted which increase the value of the objective function. Applications of these methods to seismic inversion problems can be found in [8] and [9]. Although these algorithms have been applied with some success, the objective function must be evaluated a large number of times due to the inherent stochastic nature of the methods. This is very expensive for problems of realistic size, and it is therefore desirable to look at methods which allow the use of the more efficient local methods.

2.1.2 Differential Semblance Optimization

The undesirable characteristics of the least–squares objective function can be easily understood. For noise–free data, the different traces of the seismogram (that is, the functions $D(h,\cdot)$) are simply the source waveform shifted by the travel time $2\sqrt{d^2+h^2}/c$ from the source to the receiver. When the velocity and depth are wrong, these travel times cannot be correct across all of the traces. In fact, most of the traces of the predicted seismogram must be out of phase with those of the observed seismogram. When the source is oscillatory (as it is typically), a small error in the travel times can make the observed seismogram nearly orthogonal to the observed seismogram.

As a simple example of this, we computed values of

$$\frac{\langle f, T_\tau f\rangle_{L^2[0,T]}}{||f||^2_{L^2[0,t]}}$$

for f equal to the Ricker wavelet used in the above examples. Here $T_\tau f$ represents f time–shifted by $t=\tau$. The results are shown in Figure 2.4. In this case, when the wavelet is shifted by more than about 35 ms, the result is essentially orthogonal to the original.

This effect is shown in the examples, where it can be seen that the data misfit is on the order of the size of the data except when c and d are very close to the correct values. This line of reasoning

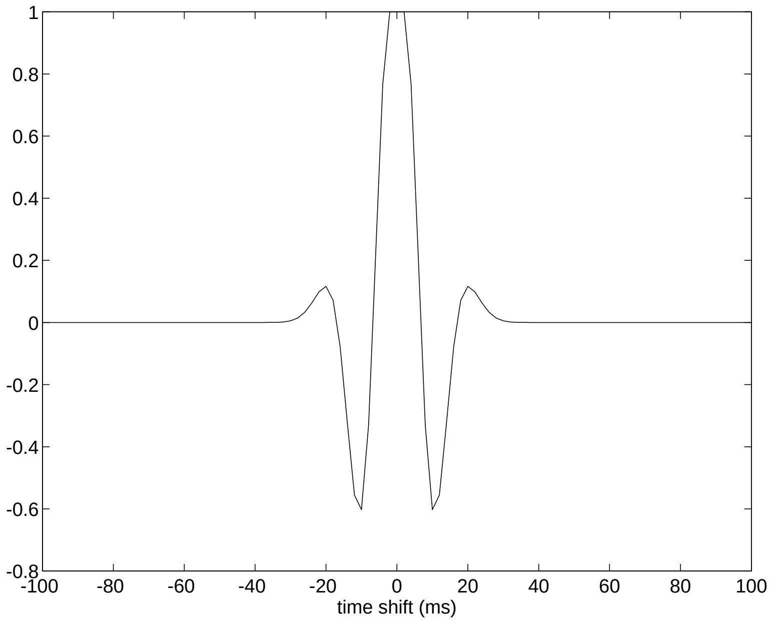

Figure 2.4: Cosine of angle between shifted and unshifted wavelets

is also supported by the fact that the undesirable characteristics of the least–squares objective function have been observed to be more pronounced for higher frequency sources (see, for example, [3]).

The idea behind Differential Semblance Optimization is to change the criterion by which the nearness of predicted and observed seismograms is measured. The DSO measure of misfit will grow gradually, rather than abruptly, as c and d move away from the exact values.

There are several steps involved in reformulating the objective function. The first step is to expand the space of model parameters to allow the source waveform to depend on the receiver location as well as on time. Mathematically, this means that the source is a function of two variables, $f = f(h,t)$, and the forward map becomes

$$\hat{S}(c,d)f(h,t) = f\left(h, t - \frac{2\sqrt{d^2+h^2}}{c}\right).$$

This addresses the root of the difficulty with Output Least–Squares, which was the difficulty in fitting the data when the velocity and depth were not close to the correct values. Within the expanded space of parameters, the observed seismogram D can be fitted exactly regardless of the value of c and d simply by taking

$$f(h,t) = D\left(h, t + \frac{2\sqrt{d^2+h^2}}{c}\right).$$

Of course, such a source is inadmissible on physical grounds, and this fact is taken into account by imposing a constraint which is satisfied only by those sources f which are constant with respect to h:

$$\frac{\partial f}{\partial h} = 0.$$

If the constraint is enforced at all times, then the problem has not been changed. In particular, if

$$\begin{aligned} \text{min} \quad & \|\hat{S}(c,d)f - D\|^2_{L^2(\Omega)} \\ \text{s.t.} \quad & \tfrac{\partial f}{\partial h} = 0 \end{aligned} \tag{2.4}$$

is solved by a *feasible point* method, then we are back to the output least–squares problem. However, the formulation of the problem as a constrained optimization problem in a larger space introduces the possibility of solving (2.4) by means of an *infeasible* point approach, that is, by an algorithm which approaches the solution through intermediate values of the parameters which do not satisfy the constraints. In Differential Semblance Optimization, the quadratic penalty function method is used.

Before writing the DSO problem, we must introduce a change of variables for technical reasons. We shall represent by f not the source waveform, but its primitive. This necessitates a change in the forward map:

$$S(c,d)f(h,t) = \frac{\partial f}{\partial t}\left(h, t - \frac{2\sqrt{d^2+h^2}}{c}\right).$$

We consider S to be a mapping from U into the space of bounded linear operators from H^1_b to $L^2(\Omega)$, where H^1_b is an appropriate subspace of $H^1(\Omega)$. We will show later that the derivative appearing in the forward map balances the derivative appearing in the constraint in an essential way.

We can now write the DSO formulation. With the above definitions,

$$\begin{aligned} \text{min} \quad & \|S(c,d)f - D\|^2_{L^2(\Omega)} \\ \text{s.t.} \quad & \tfrac{\partial f}{\partial h} = 0 \end{aligned} \tag{2.5}$$

is equivalent to the Output Least–Squares problem (bearing in mind the change of variables involving f), in the sense that the two problems have equivalent minimizers. The quadratic penalty function approach to solving this problem introduces the constraint into the objective function by adding the term

$$r^2\|\frac{\partial f}{\partial h}\|^2_{L^2(\Omega)}.$$

The idea behind this method is to replace the constrained problem by a sequence of unconstrained problems which converge to it (in some sense) as $r \uparrow \infty$.

As we did above, we reduce the objective function by first minimizing over f, taking advantage of the fact that this inner problem is a linear least–squares problem. The (reduced) DSO objective function is then

$$J_r(c,d) = \min_f \left\{ \frac{1}{2}\|S(c,d)f - D\|^2_{L^2(\Omega)} + \frac{1}{2}r^2\|\frac{\partial f}{\partial h}\|^2_{L^2(\Omega)} \right\}, \tag{2.6}$$

and we wish to minimize $J_r(c,d)$ over c and d.

The following facts about the DSO problem should be noted:

- For any $D \in L^2(\Omega)$, J_r is smooth (in fact, infinitely differentiable) as a function of c and d.
- The minimization problem which defines J_r is a linear least–squares problem which may be solved via the normal equations. These normal equations take the form of an elliptic boundary value problem.
- When the data D is exact, that is, when there exists f, c, and d such that $S(c,d)f = D$, $\frac{\partial f}{\partial h} = 0$, then the quadratic penalty function is an *exact* penalty function (meaning that the minimizer of the penalty function is the solution to the constrained problem). This is because both the data misfit and penalty terms are zero at the solution.
- The DSO objective function can be shown to have good global properties (in fact, to be nearly convex) for *small* values of the penalty parameter.

These facts will be proved in later sections.

The above claim that the DSO objective function is well–behaved for small values of the penalty parameter is somewhat vague — after all, what constitutes a small value? — and deserves comment. In making this distinction in the size of the penalty parameter, we are not thinking of the absolute size of r, but rather of the conventional thinking about penalty functions. The idea of the penalty method is to replace a constrained problem with a "nearby" unconstrained problem, or rather by a family of unconstrained problems which converge to the constrained problem. In the context of the quadratic penalty method, convergence is obtained as the penalty parameter tends to infinity. The only significance of smaller values of the parameter lies in the fact that the *local* conditioning of the penalty problem is improved. In the context of the DSO method, however, the numerical examples show that the *global* (or semiglobal) properties of the penalty function are better for values of r which are "small". In fact, as we shall see in the following sections, there is a special significance to the limiting behavior of the DSO function as $r \downarrow 0$. It is in this sense that we refer to the penalty parameter being small.

The DSO objective function was computed for the first example above (see Figure 2.2). The value of the penalty parameter in this example is $r = 1$. The result is displayed in Figure 2.5. As the plots show, the DSO objective function does not suffer from the poor global characteristics of the output least–squares function; in fact, the DSO objective function is nearly convex.

After presenting the abstract theory in the following sections, we shall return to this example and prove that these plots tell the right story; in other words, we shall prove that the DSO objective function has good global properties for minimization by local methods. In particular, we will discuss conditions under which there is a unique stationary point. We will also show how the solution of the DSO problem can lead to the solution of the OLS problem through path–following techniques.

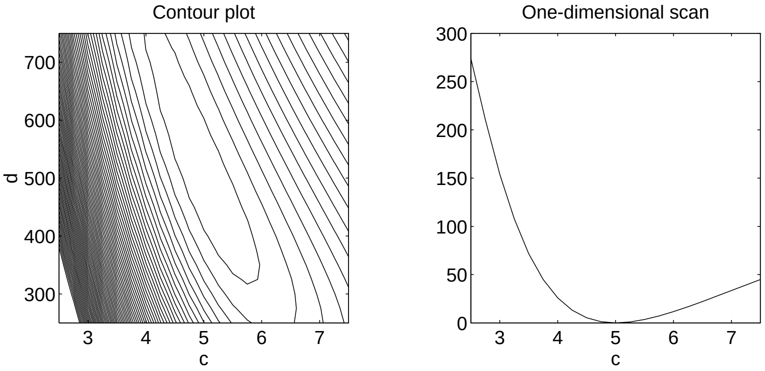

Figure 2.5: $J_1(c, d)$ for a Ricker source

2.2 Plane wave detection

The plane wave detection problem is another simple model problem for velocity inversion. The application of DSO to this problem has been investigated by Symes (see [12]). We wish to briefly discuss the plane wave detection problem in order to compare the predictions of our theory with known results.

A *plane wave* (moving at speed one) is a function $g : \mathcal{R}^3 \to \mathcal{R}$ of the form

$$g(x, y, t) = u(t + x\sin\theta + y\cos\theta),$$

where $u : \mathcal{R} \to \mathcal{R}$. Suppose that g is sampled along an interval $[-X, X]$ on the x-axis for $t \in [0, T]$, giving $z(x, t) = g(x, 0, t) = u(t + x\sin\theta)$ (again, in practice, the sampling would be done at a number of discrete locations in the interval [-X,X]). The aim of the plane wave detection problem is to recover the wave form u and the "slope" $s = \sin\theta$ from the data z.

2.2.1 Output Least–Squares

Define $\Sigma = [-X, X] \times [0, T]$ and assume that $z \in L^2(\Sigma)$. The forward map for this problem is $\tilde{\phi} : \mathcal{R} \to \mathcal{L}(L^2[0, T], L^2(\Sigma))$ defined by

$$\tilde{\phi}(s)u(x, t) = u(t + sx).$$

The OLS problem is to find a real number $s \in [-1, 1]$ and a function $u \in L^2[0, T]$ so that $\phi(s)u$ matches the data z as closely as possible in the L^2 norm. As in the previous problem, we reduce the objective function by eliminating the source u. The reduced least–squares objective function is

$$\tilde{J}(s) = \min_{u} \|\phi(s)u - z\|^2_{L(\Sigma)}; \tag{2.7}$$

as before, if the data z is smooth, then so is $\tilde{J}$.

We computed $\tilde{J}$ for both noise–free and noisy data. The source generating the data was a Ricker wavelet of frequency 20 Hz., there were 201 receivers equally spaced in the interval $[-0.5, 0.5]$, and the recording interval was $[0, 1]$. The results are shown in Figure 2.6. As in the case of the layer–over–half–space problem, the least–squares objective function is seen to have its minimizer at the bottom of a narrow valley and to be very flat outside of that valley. In the case of noisy data, many local minimizers appear on the plateau.

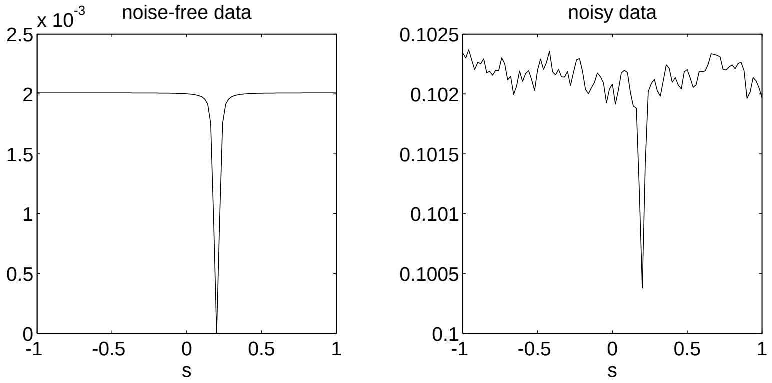

Figure 2.6: $\tilde{J}(s)$ for a Ricker source

2.2.2 Differential Semblance Optimization

The DSO formulation is derived just as in the layer–over–half–space problem:

- Change variables so that u represents the time–antiderivative of the source, and introduce a time–derivative into the forward map.
- Allow the source to depend on the receiver, so that u becomes a function of two variables, $u = u(x, t)$.
- Introduce the differential constraint
$$\frac{\partial u}{\partial x} = 0,$$
so that the original unconstrained problem is replaced by an equivalent constrained problem. Apply the quadratic penalty function method.

The reduced DSO objective function is

$$J_r(s) = \min_{u} \left\{ \|\phi(s)u - z\|^2_{L^2(\Sigma)} + r^2 \|\frac{\partial u}{\partial x}\|^2_{L^2(\Sigma)} \right\}, \tag{2.8}$$

where

$$\phi(s)u(x,t) = \frac{\partial u}{\partial t}(x, t + sx) \, .$$

For the purposes of comparison, the DSO objective function was computed for the same example. The graphs in Figure 2.7 should be compared to those in Figure 2.6. As in the layer–over–half–space problem, the DSO objective function has very good global properties when the penalty parameter is "small".

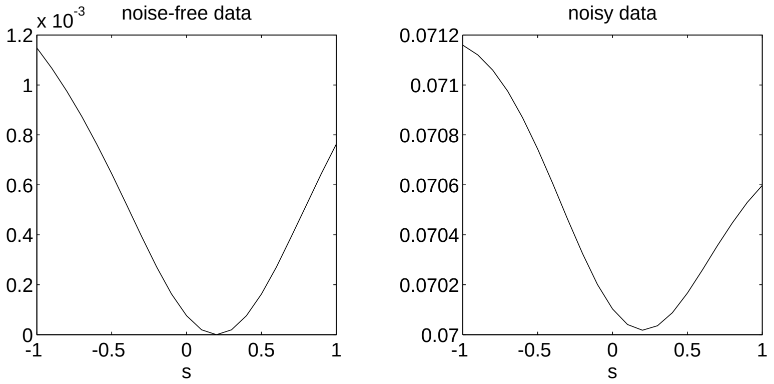

Figure 2.7: $J_1(s)$ for a Ricker source

In the next sections, we give an abstract framework for understanding the kind of results illustrated in these examples. We shall then return to these examples to put everything on a firm mathematical foundation. In particular, we shall apply the results of the abstract analysis to show that the numerical examples given in this section can be justified theoretically.

3 Abstract formulation: inner problem

We wish to define an abstract optimization problem that includes the preceding examples as special cases. Capturing the salient features requires a great deal of structure, particularly in the underlying Hilbert spaces. We shall introduce this structure as needed for the analysis.

The general framework of our problem is as follows. The forward map is of the form

$$A : U \to \mathcal{L}(Y, D),$$

where Y and D are Hilbert spaces and U is an open subset of another Hilbert space X. The model parameters are represented by $(x, y) \in X \times Y$, where x has a nonlinear influence on the predicted data and y has a linear influence. The data d lies in the space D.

Recall that in the examples, there were two main steps in arriving at the DSO formulation. The first was expanding the space of model parameters; the second was the imposition of a constraint to regain the "physics" of the problem. We are assuming that the first of these steps has already been carried out; the space Y represents the expanded model space. The operator $W \in \mathcal{L}(Y, E)$, where E is another Hilbert space, will represent the constraint.

We also follow the examples in that the objective functions we consider are all reduced; that is, the parameter y, which appears linearly in the forward map and constraint, is always eliminated through the "inner" minimization. We thus have the following problems: The "Output Least–Squares" problem is to find $x \in U$ to solve

$$\min_x \quad J_\infty(x),$$

where

$$J_\infty(x) = \min \left\{ \|A(x)y - d\|^2 : Wy = 0 \right\}.$$

The "Differential Semblance Optimization" problem is to find $x \in U$ to solve

$$\min_x \quad J_r(x),$$

where

$$J_r(x) \quad = \quad \min \left\{ \|A(x)y - d\|^2 + r^2\|Wy\|^2 : y \in Y \right\}.$$

In order to simplify the presentation, and to focus attention on the appropriate hypotheses, we divide our analysis into two parts. In this section and the next, we will analyze the DSO inner minimization problem, which turns out to be a regularization of a problem which is dual (in a sense to be defined later) to the OLS inner problem. This observation is perhaps the key insight contained in all of this research. In Sections 5 and 6, we will discuss the dependence of the problem on the nonlinear variable. This will require both stronger assumptions and more delicate analysis.

3.1 Assumptions

There are several assumptions (and nonassumptions) that are important in capturing the flavor of the examples. The basis for the DSO approach is that, although the data may be far from the range of A restricted to the null space of W, it can be fitted arbitrarily well when the constraint is no longer imposed. We will therefore be primarily interested in the case in which $d \in \overline{\mathcal{R}(A)}$.

One of the important nonassumptions is that the operator A is not assumed to be coercive. In other words, we allow for the possibility, which occurs in the examples, that there exists a sequence $\{y_n\}$ such that

$$\|y_n\|_Y = 1 \text{ for all n}$$

and

$$Ay_n \to 0 \text{ as } n \to \infty.$$

(Since A represents the forward map in the problems of interest, there is an important interpretation of this lack of coercivity. There are perturbations of the model parameters which do not lead to proportionally large perturbations of the resulting data.) We will assume, however, that A is injective, as this can be achieved by redefining the domain Y, if necessary.

Under these circumstances, in order for the DSO inner problem to be well–defined, the operators A and W must have spectra that are complementary in some sense. Specifically, we assume that there is a real number $a > 0$ such that

$$\|Ay\|_D^2 + \|Wy\|_E^2 \geq a^{-1}\|y\|_Y^2$$

for all $y \in Y$.

We will also assume that Y is separable. This is a mild assumption that is satisfied in the problems of interest, in which Y is a function space.

3.2 Summary

For convenience and easy reference, we list here the assumptions that we make:

(A1) Y, D, and E are Hilbert spaces.

(A2) Y is separable.

(A3) $A \in \mathcal{L}\{Y, D\}$.

(A4) A is injective.

(A5) $W \in \mathcal{L}\{Y, E\}$.

(A6) There exists $a > 0$ such that $\|Ay\|_D^2 + \|Wy\|_E^2 \geq a^{-1}\|y\|_Y^2$ for all $y \in Y$.

4 Analysis of the inner problem

4.1 Preliminaries

We begin our analysis of the DSO inner problem by noting that it can be written as a parameterized linear least–squares problem:

$$
\begin{aligned}
J_r &= \min\left\{\|Ay-d\|_D^2 + r^2\|Wy\|_E^2 : y \in Y\right\} \\
&= \min\left\{\left\|\begin{bmatrix} A \\ rW \end{bmatrix} y - \begin{bmatrix} d \\ 0 \end{bmatrix}\right\|_{D\times E}^2 : y \in Y\right\}.
\end{aligned} \tag{4.1}
$$

Problems of this and similar forms have been investigated by many authors in the context of regularizing ill–posed problems. This topic is surveyed by Engl [1], who gives an extensive bibliography. In particular, if the operator W were replaced by the identity operator I, we would have the classical formulation of Tikhonov regularization ([14]). The situation in which A is an integral operator and W is a densely defined unbounded differential operator is discussed in [4], where results similar to ours are derived (with, however, different assumptions on the operators). Another reference in which similar results are discussed (again with somewhat different hypotheses) is [6].

The relationship to regularization theory is very suggestive, for the following reason. We are interested in the problem

$$\min \quad \|Ay-d\|_D^2 \tag{4.2}$$

$$\text{s.t.} \quad Wy = 0, \tag{4.3}$$

whereas the regularization idea embodied in (4.1) is directed toward the solution of

$$\min \quad \|Wy\|_D^2 \tag{4.4}$$

$$\text{s.t.} \quad y \text{ is a least–squares solution to } Ay = d, \tag{4.5}$$

which reduces to

$$\begin{array}{ll} \min & \|Wy\|_D^2 \\ \text{s.t.} & Ay = d \end{array} \tag{4.6}$$

when $d \in \mathcal{R}(A)$. Convergence (when it occurs) is obtained as $r \downarrow 0$.

Problem (4.6) is therefore intrinsic in the DSO formulation. The surprising aspect of this is that, in the examples, we obtain good results when r is small, even though we set out to solve (4.2). This suggests that we should consider the significance of problem (4.6). Indeed, we will see, both from the theory and the examples, that it is the existence of this "dual" problem which is responsible for the success of the DSO method.

Incidentally, if we write J_0 for the optimal value of (4.6), then under the right conditions we will obtain

$$\frac{1}{r^2} J_r \longrightarrow J_0 \text{ as } r \downarrow 0,$$

whereas

$$J_r \longrightarrow 0 \text{ as } r \downarrow 0.$$

This is natural when we notice that J_r/r^2 can be regarded as the minimum of the quadratic penalty function for (4.6), where the penalty on the violation of the constraint $Ay = d$ is large when r is close to zero.

Our first task is to verify that J_∞ and J_r are well–defined.

Lemma 1 *The space $\{Ay : y \in \mathcal{N}(W)\}$ is a closed subspace of D.*

Proof: Note that by Assumption (A6), $\|Ay\|_D^2 \geq a^{-1}\|y\|_Y^2$ for all $y \in \mathcal{N}(W)$. Suppose that $\{y_n\}$ is a sequence in $\mathcal{N}(W)$ such that $Ay_n \to d \in D$. Then $\{Ay_n\}$ is a Cauchy sequence, and the coercivity of A on the null space of W implies that $\{y_n\}$ is also a Cauchy sequence. It follows that $y_n \to y$ for some $y \in Y$, and hence that $Ay_n \to Ay$. Thus $d = Ay$, and so $\{Ay : y \in \mathcal{N}(W)\}$ is closed.

□

Theorem 1 *For any $d \in D$, there exists a unique $y_\infty \in \mathcal{N}(W)$ such that*

$$\|Ay_\infty - d\|_D^2 = min\left\{\|Ay - d\|_D^2 : Wy = 0\right\}.$$

Moreover, y_∞ is characterized by the conditions

$$A^*Ay_\infty - A^*d \in \mathcal{N}(W)^\perp,$$

$$Wy_\infty = 0.$$

Proof: Choose a sequence $\{y_n\}$ in $\mathcal{N}(W)$ such that $\|Ay_n - d\|_D$ converges to $\min\{\|Ay - d\|_D : Wy = 0\}$. The coercivity of A on the null space of W implies that $\{y_n\}$ is bounded, so, without loss of generality, $\{y_n\}$ is weakly convergent, say $y_n \to y_\infty$ weakly. It follows that $Ay_n - d \to Ay_\infty - d$ weakly, which implies that

$$\begin{aligned} & \|Ay_\infty - d\|_D^2 \\ \leq\ & \liminf \|Ay_n - d\|_D^2 = \min\left\{\|Ay - d\|_D^2 : Wy = 0\right\}. \end{aligned}$$

Moreover, the null space of W is weakly closed, so $y_\infty \in \mathcal{N}(W)$. This shows that y_∞ is a solution. The fact that it is unique follows immediately from the fact that

$$y \to \frac{1}{2}\|Ay - d\|_D^2$$

is a strictly convex function. The characterizing equations are the first order necessary conditions for a solution of the minimization problem, which are also sufficient due to the convexity of the function and the constraint set.

□

The above results show that J_∞ is well–defined; next, we consider J_r.

Lemma 2 *For any $r \in (0,\infty)$, the operator $M_r : Y \to D \times E$ defined by*

$$M_r y = \begin{bmatrix} A \\ rW \end{bmatrix} y = \begin{bmatrix} Ay \\ rWy \end{bmatrix}$$

is coercive, that is, there exists $c_r > 0$ such that

$$\|M_r y\|_{D\times E} \geq c_r \|y\|_Y$$

for all $y \in Y$.

Proof: If $r \geq 1$, then, applying Assumption (A6), we have

$$\begin{aligned} \|M_r y\|^2_{D\times E} &= \|Ay\|^2_D + r^2\|Wy\|^2_E \\ &\geq \|Ay\|^2_D + \|Wy\|^2_E \\ &\geq a^{-1}\|y\|^2_Y, \end{aligned}$$

so c_r can be taken to be $a^{-1/2}$ for $r \geq 1$. If $r \in (0,1)$, then

$$\begin{aligned} \|M_r y\|^2_{D\times E} &= \|Ay\|^2_D + r^2\|Wy\|^2_E \\ &\geq r^2\|Ay\|^2_D + r^2\|Wy\|^2_E \\ &\geq \frac{r^2}{a}\|y\|^2_Y. \end{aligned}$$

Thus for $r \in (0,1)$, we can take $c_r = r/\sqrt{a}$.

□

Corollary 1 *The symmetric operator $N_r : Y \to Y$ given by $N_r = A^*A + r^2W^*W$ defines a coercive bilinear form for all $r \in (0,\infty)$. Moreover, N_r is invertible for all $r \in (0,\infty)$ and*

$$\|N_r^{-1}\| \leq \begin{cases} a/r^2 & , \quad r \in (0,1) \\ a & , \quad r \in [1,\infty) \end{cases} .$$

Proof: The proof follows immediately from the proof of the preceding lemma.

□

Theorem 2 *For any $d \in D$ and any $r \in (0,\infty)$, there exist a unique $y_r \in Y$ such that*

$$\|Ay_r - d\|^2_D + r^2\|Wy_r\|^2_E = min\left\{\|Ay - d\|^2_D + r^2\|Wy\|^2_E : y \in Y\right\},$$

namely,

$$y_r = N_r^{-1}A^*d.$$

Proof: The functional $F : Y \to \mathcal{R}$ defined by $F(y) = \|Ay\|_D^2 + r^2\|Wy\|_E^2$ is a convex quadratic whose gradient is

$$(A^*A + r^2W^*W)y - A^*d.$$

The coercivity of the normal operator $N_r = A^*A + r^2W^*W$ implies that there is a unique stationary point

$$y_r = (A^*A + r^2W^*W)^{-1}A^*d,$$

which by convexity is the global minimizer.

□

When the dependence on x or d needs to be emphasized, we will write $y_r = y_r(x,d)$ and $y_\infty = y_\infty(x,d)$.

4.2 Limiting behavior as $r \downarrow 0$

We now consider the behavior of y_r as r varies. The following results culminate in the result that, when $d \in \mathcal{R}(A)$, $y_r \to y_0$, where y_0 is the unique element of Y such that $Ay_0 = d$. We will also derive the asymptotic rate of convergence under certain conditions. We write J_0 for the optimal value of the dual problem:

$$J_0 = \min \left\{\|Wy\|_E^2 : Ay = d\right\} = \|Wy_0\|_E^2.$$

We first collect some basic facts about our parameterized linear least–squares problem. In what follows, we must distinguish two cases: (1) $d - Ay_\infty \in \mathcal{R}(A)^\perp$ (this condition implies that the solution y_∞ of

$$\begin{array}{ll} \min & \|Ay - d\|_D^2 \\ \text{s.t.} & Wy = 0 \end{array} \tag{4.7}$$

is the unconstrained minimizer of $\|Ay - d\|_D$); (2) $d - Ay \notin \mathcal{R}(A)^\perp$ for all $y \in \mathcal{N}(W)$.

Lemma 3 *Assume that $d - Ay_\infty \in \mathcal{R}(A)^\perp$. Then*

1. *$y_r = y_\infty\,[= y_0]$ for all $r \in (0,\infty)$ [if $d \in \mathcal{R}(A)$];*
2. *$J_r = J_\infty\,[= J_0] = 0$ for all $r \in (0,\infty)$ [if $d \in \mathcal{R}(A)$].*

Proof: The proof is straightforward.

□

Lemma 4 *Assume that $d - Ay_\infty \notin \mathcal{R}(A)^\perp$. Then*

1. *$Wy_r \neq 0$ for all $r \in (0,\infty)$;*

2. *J_r is an increasing function of r;*
3. *$J_r < J_\infty$ for all $r \in (0,\infty)$;*
4. *$y_{r_1} \neq y_{r_2}$ whenever $r_1 \neq r_2$;*
5. *$\|Wy_r\|_E \downarrow 0$ as $r \uparrow \infty$;*
6. *$\|Ay_r - d\|_D$ is an increasing function of r;*
7. *$\|Ay_r - d\|_D < \|Ay_\infty - d\|_D$ for all $r \in (0,\infty)$;*
8. *if $d \in \mathcal{R}(A)$, then $\|Wy_r\|_E < \|Wy_0\|_E$ for all $r > 0$.*

Proof:

1. Recall that y_r satisfies the normal equations:
$$(A^*A + r^2W^*W)y_r = A^*d.$$
If $Wy_r = 0$, then we have $A^*Ay_r = A^*d$, or $A^*(Ay_r - d) = 0$. This yields
$$\langle A^*(Ay_r - d), z\rangle_Y = 0 \text{ for all } z \in Y,$$
or
$$\langle Ay_r - d, Az\rangle_D = 0 \text{ for all } z \in Y.$$
This last condition, together with the fact that $y_r \in \mathcal{N}(W)$, implies that $y_r = y_\infty$ and $d - Ay_\infty \in \mathcal{R}(A)^\perp$.

2. Suppose $r_1 < r_2$. Then
$$\begin{aligned} J_{r_1} &= \|Ay_{r_1} - d\|_D^2 + r_1^2\|Wy_{r_1}\|_E^2 \\ &\le \|Ay_{r_2} - d\|_D^2 + r_1^2\|Wy_{r_2}\|_E^2 \\ &< \|Ay_{r_2} - d\|_D^2 + r_2^2\|Wy_{r_2}\|_E^2 \\ &= J_{r_2} \end{aligned}$$
(the strict inequality follows from the fact that $\|Wy_{r_2}\|_E \neq 0$).

3. We have
$$\begin{aligned} J_r = \min\left\{\|Ay - d\|_D^2 + r^2\|Wy\|_E^2 : y \in Y\right\} &< \|Ay_\infty - d\|_D^2 + r^2\|Wy_\infty\|_E^2 \\ &= \|Ay_\infty - d\|_D^2 = J_\infty. \end{aligned}$$

4. The normal equations yield
$$(A^*A + r_1^2W^*W)y_{r_1} = (A^*A + r_2^2W^*W)y_{r_2}$$
If $y_{r_1} = y_{r_2}$, then
$$r_1^2W^*Wy_{r_1} = r_2^2W^*Wy_{r_2},$$
which implies that
$$(r_1^2 - r_2^2)W^*Wy_{r_1} = 0.$$
Since $Wy_{r_1} \neq 0$, this yields $r_1 = r_2$.

5. Suppose $r_1 < r_2$. Then

$$\begin{aligned} \|Ay_{r_1} - d\|_D^2 + r_1^2\|Wy_{r_1}\|_E^2 &< \|Ay_{r_2} - d\|_D^2 + r_1^2\|Wy_{r_2}\|_E^2, \\ \|Ay_{r_2} - d\|_D^2 + r_2^2\|Wy_{r_2}\|_E^2 &< \|Ay_{r_1} - d\|_D^2 + r_2^2\|Wy_{r_1}\|_E^2. \end{aligned}$$

   Adding and simplifying yields

$$(r_1^2 - r_2^2)(\|Wy_{r_1}\|_E^2 - \|Wy_{r_2}\|_E^2) < 0.$$

   Since $r_1 < r_2$ by assumption, this shows that $\|Wy_{r_1}\|_E^2 > \|Wy_{r_2}\|_E^2$. Thus $\|Wy_r\|_E$ is strictly decreasing. If $\|Wy_r\|_E$ did not converge to zero, then $r^2\|Wy_r\|_E \to \infty$, contradicting the fact that $J_r < J_\infty$ for all r.

6. Suppose $r_1 < r_2$. Then

$$\|Ay_{r_1} - d\|_D^2 + r_1^2\|Wy_{r_1}\|_E^2 < \|Ay_{r_2} - d\|_D^2 + r_1^2\|Wy_{r_2}\|_E^2,$$

   which implies that

$$r_1^2(\|Wy_{r_1}\|_E^2 - \|Wy_{r_2}\|_E^2) < \|Ay_{r_2} - d\|_D^2 - \|Ay_{r_1} - d\|_D^2.$$

   The preceding result then yields

$$\|Ay_{r_2} - d\|_D^2 - \|Ay_{r_1} - d\|_D^2 > 0,$$

   as desired.

7. We have

$$\|Ay_r - d\|_D^2 < J_r < J_\infty = \|Ay_\infty - d\|_D^2,$$

   which is the desired inequality.

8. For all $r > 0$, we have

$$\begin{aligned} \|Ay_r - d\|_D^2 + r^2\|Wy_r\|_E^2 &\le \|Ay_0 - d\|_D^2 + r^2\|Wy_0\|_E^2 \\ &= r^2\|Wy_0\|_E^2, \end{aligned}$$

   since $Ay_0 = d$. This implies that

$$\|Wy_r\|_E^2 \le \|Wy_0\|_E^2,$$

   and the inequality must be strict since we have already shown that $\|Wy_r\|_E$ is strictly increasing as $r \downarrow 0$.

$\square$

Theorem 3 *If $d \in \overline{\mathcal{R}(A)}$, then $\|Ay_r - d\|_D \to 0$ as $r \downarrow 0$.*

Proof: We have shown that $||Ay_r - d||_D$ decreases as $r \downarrow 0$. Suppose, by way of contradiction, that $||Ay_r - d||_D \geq c > 0$ for all $r > 0$. Since d is in the closure of $\mathcal{R}(A)$, there exists $\tilde{y} \in Y$ such that $||A\tilde{y} - d||_D < c/\sqrt{2}$. But for $\tilde{r}$ sufficiently small, $\tilde{r}||W\tilde{y}||_E < c/\sqrt{2}$. Then

$$\begin{aligned} J_{\tilde{r}} &\leq ||A\tilde{y} - d||_D^2 + \tilde{r}^2||W\tilde{y}||_E^2 \\ &< \frac{c^2}{2} + \frac{c^2}{2} \\ &= c^2. \end{aligned}$$

From this we see that

$$||Ay_{\tilde{r}} - d||_D^2 \leq J_{\tilde{r}} < c^2,$$

or

$$||Ay_{\tilde{r}} - d||_D < c.$$

This is the desired contradiction.

□

Theorem 4 *Assume $d \in \mathcal{R}(A)$. Then $y_r \to y_0$ weakly as $r \downarrow 0$.*

Proof: We must show that, for all $y \in Y$,

$$\langle y_r, y\rangle_Y \to \langle y_0, y\rangle_Y \text{ as } r \downarrow 0.$$

Choose any $y \in Y$ and define

$$z_r = N_r^{-1}y.$$

Then

$$\begin{aligned} \langle y_r, y\rangle_Y &= \langle y_r, N_r z_r\rangle_Y \\ &= \langle N_r y_r, z_r\rangle_Y \\ &= \langle A^* d, z_r\rangle_Y \text{ (by the normal equations)} \\ &= \langle d, Az_r\rangle_D \\ &= \langle Ay_0, Az_r\rangle_D \\ &= \langle y_0, A^*Az_r\rangle_D \\ &= \langle y_0, N_r z_r\rangle_Y - r^2\langle y_0, W^*Wz_r\rangle_Y \\ &= \langle y_0, y\rangle_Y - r^2\langle y_0, W^*WN_r^{-1}y\rangle_Y. \end{aligned}$$

Therefore, if $0 < r \leq 1$, then

$$\begin{aligned} |\langle y_r, y\rangle_Y| &\leq ||y_0||_Y||y||_Y + r^2||W^*W||||N_r^{-1}||||y_0||_Y||y||_Y \\ &\leq ||y_0||_Y||y||_Y \left(1 + r^2\frac{a}{r^2}||W||^2\right) \text{ (by Corollary 1)} \\ &= ||y_0||_Y||y||_Y \left(1 + a||W||^2\right). \end{aligned}$$

So, for each $y \in Y$, $\{|\langle y_r, y\rangle_Y| : r \in (0,1]\}$ is bounded. By the Uniform Boundedness Principle, $\{y_r : r \in (0,1]\}$ must be bounded.

Since Y is separable, $\{y_r : r \in (0,1]\}$ has a weakly convergent subsequence. Let $\{y_{r_n}\}$ be any such subsequence, where $r_n \downarrow 0$ and $y_{r_n} \to \bar{y}$ weakly as $n \to \infty$. Then $Ay_{r_n} \to A\bar{y}$ weakly. But by the previous theorem, $Ay_{r_n} \to d = Ay_0$. Since A is injective, $\bar{y} = y_0$.

We have shown that any weakly convergent subsequence of $\{y_r : r \in (0,1]\}$ converges weakly to y_0; this suffices to show that $\{y_r : r \in (0,1]\}$ itself converges weakly to y_0.

□

Corollary 2 *If $d \in \mathcal{R}(A)$, then $Wy_r \to Wy_0$ as $r \downarrow 0$.*

Proof: By the previous theorem, $Wy_r \to Wy_0$ weakly, which implies that

$$\|Wy_0\|_E \le \liminf \|Wy_r\|_E.$$

But by Lemma 4,

$$\limsup \|Wy_r\|_E \le \|Wy_0\|_E.$$

This shows that $\|Wy_r\|_E \to \|Wy_0\|_E$ as $r \downarrow 0$; with the fact that $Wy_r \to Wy_0$ weakly, this suffices to prove that $Wy_r \to Wy_0$.

□

Theorem 5 *If $d \in \mathcal{R}(A)$, then $y_r \to y_0$ as $r \downarrow 0$.*

Proof: By Assumption (A6), it suffices to show that

$$(A^*A + W^*W)(y_r - y_0) \to 0 \text{ as } r \downarrow 0.$$

But we have shown (Theorem 3 and Corollary 2) that

$$A(y_r - y_0) \to 0 \text{ and } W(y_r - y_0) \to 0 \text{ as } r \downarrow 0.$$

The desired result follows immediately.

□

The preceding theorem, while interesting, does not answer the more important question of the limiting behavior of J_r/r^2. In particular, the term

$$\frac{\|Ay_r - d\|_D^2}{r^2}$$

may "blow up" as $r \downarrow 0$. The following theorem gives a condition under which we can make a positive statement about the rate of convergence of y_r to y_0. The condition appears rather obscure, but it will be seen in the examples to be related to the hypothesis of "exact" data. In the context of our problem, data $d \in D$ is exact if there exist $x^* \in X$, $y^* \in Y$ such that $A(x^*)y^* = d$ and $Wy^* = 0$. This understanding of what constitutes exact data is natural when the evolution of the problem is considered. The space of parameters represented by y was expanded and the constraint $Wy = 0$ was added to retrieve the original problem. Exact data is therefore data which can be fitted exactly in the original formulation, that is, by feasible data. We shall use the following result several times.

Lemma 5 *1. Let $y \in Y$. Then*

$$N_r^{-1}y$$

is bounded as $r \downarrow 0$ if and only if $y \in \mathcal{R}(A^*A)$.

2. Let $y \in Y$. Then

$$(r^{-2}A^*A + W^*W)^{-1}y$$

is bounded as $r \uparrow \infty$ if and only if $y \in \mathcal{R}(W^*W)$.

Proof:

1. If $y = A^*Au$, then

$$\begin{aligned} N_r^{-1}y &= N_r^{-1}A^*Au \\ &= N_r^{-1}(A^*A + r^2W^*W)u - r^2N_r^{-1}W^*Wu \\ &= u - r^2N_r^{-1}W^*Wu. \end{aligned}$$

Since

$$\|N_r^{-1}\| \leq \frac{a}{r^2}$$

for $r \in (0,1)$,we obtain

$$\|N_r^{-1}y\|_Y \leq (1 + a\|W\|^2)\|u\|_Y$$

for all $r \in (0,1)$.

Conversely, suppose that

$$z_r = N_r^{-1}y$$

is bounded as $r \downarrow 0$. Suppose that $r_k \downarrow 0$ and $z_r \rightharpoonup \overline{z}$ as $k \to \infty$. Then

$$y = (A^*A + r_k^2W^*W)z_{r_k} \rightharpoonup A^*A\overline{z} \text{ weakly },$$

which implies that $y = A^*Au \in \mathcal{R}(A^*A)$.

2. The proof of this part is exactly analogous to that of the preceding part.

□

We can now prove one of our main results.

Theorem 6 *Assume $d \in \mathcal{R}(A)$. Then $y_r - y_0 = O(r^2)$ if and only if $W^*Wy_0 \in \mathcal{R}(A^*A)$.*

Proof: We have

$$\begin{aligned} y_r - y_0 &= N_r^{-1}A^*d - y_0 \\ &= N_r^{-1}A^*Ay_0 - y_0 \\ &= N_r^{-1}(A^*A - A^*A - r^2W^*W)y_0 \\ &= -r^2N_r^{-1}W^*Wy_0. \end{aligned}$$

Thus we see that $\|y_r - y_0\|_Y = O(r^2)$ if and only if

$$N_r^{-1}W^*Wy_0$$

is bounded as $r \downarrow 0$. By Lemma 5, this occurs exactly when $W^*Wy_0 \in \mathcal{R}(A^*A)$.

□

The condition $W^*Wy_0 \in \mathcal{R}(A^*A)$ has been used in different circumstances to prove similar convergence results; see [4]. This result gives us a condition under which the penalty problem has a limit as $r \downarrow 0$.

Corollary 3 *Assume that $d \in \mathcal{R}(A)$ and $W^*Wy_0 \in \mathcal{R}(A^*A)$. Then*

$$\frac{1}{r^2}J_r - J_0 = O(r^2).$$

Proof: We have

$$\|Ay_r - d\|_D^2 = \|A(y_r - y_0)\|_D^2 = O(r^4)$$

and

$$\begin{aligned} 0 &\le \|Wy_0\|_E^2 - \|Wy_r\|_E^2 \\ &= (\|Wy_0\|_E - \|Wy_r\|_E)(\|Wy_0\|_E + \|Wy_r\|_E) \\ &\le \|W(y_0 - y_r)\|_E(\|Wy_0\|_E + \|Wy_r\|_E) \\ &= O(r^2). \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{1}{r^2}J_r - J_0 &= \frac{\|Ay_r - d\|_D^2}{r^2} + \|Wy_r\|_E^2 - \|Wy_0\|_E^2 \\ &= O(r^2). \end{aligned}$$

□

The preceding results will give a satisfactory picture for exact data, and a partial understanding of the situation for inexact data d satisfying $d \in \mathcal{R}(A)$ (we know only that $y_r \to y_0$ in this case). The following results give some information about the situation in which the data d lies outside of $\mathcal{R}(A)$.

Theorem 7 *If $d \in \overline{\mathcal{R}(A)} \setminus \mathcal{R}(A)$, then $\|y_r\|_Y \to \infty$ as $r \downarrow 0$.*

Proof: Note that if $\{y_n\}$ is a sequence in Y and $Ay_n \to d$, then $||y_n||_Y \to \infty$. For if not, then, without loss of generality, $\{y_n\}$ is bounded and $y_n \to \tilde{y}$ weakly. But then $||A\tilde{y} - d||_D \leq \liminf ||Ay_n - d||_D = 0$, which implies that $A\tilde{y} = d$, a contradiction.

The desired result now follows from Theorem 3.

□

The following result gives some information about the rate at which $||y_r||_Y$ "blows up" in the presence of noise in the data.

Theorem 8 *Assume* $d \in \overline{\mathcal{R}(A)} \setminus \mathcal{R}(A)$. *Let* $y_r = y_r(d)$. *Then:*

1. $||Wy_r||_E^2 \geq ||y_r||_Y^2/2a$ *for* r *sufficiently small;*
2. $r||Wy_r||_E^2 \to 0$ *as* $r \downarrow 0$;
3. *for all* r *sufficiently small,*
$$||y_r||_Y \leq \sqrt{2a}\frac{||d||_D}{r};$$
4. $r||y_r||_Y \to 0$ *as* $r \to 0$.

Proof:

1. By Assumption (A6),
$$\langle y_r, A^*Ay_r\rangle_Y + \langle y_r, W^*Wy_r\rangle_Y \geq a^{-1}||y_r||_Y^2,$$
which yields
$$\begin{aligned} \langle y_r, W^*Wy_r\rangle_Y &\geq \frac{||y_r||_Y^2}{a}\left(1 - \langle A\frac{y_r}{||y_r||_Y}, A\frac{y_r}{||y_r||_Y}\rangle_D\right) \\ &\geq \frac{||y_r||_Y^2}{2a} \end{aligned}$$
for r sufficiently small, since $Ay_r \to d$ and $||y_r||_Y \to \infty$.
2. From the normal equations, we obtain
$$r^2||Wy_r||_E^2 = \langle y_r, r^2W^*Wy_r\rangle_Y = \langle y_r, A^*(d - Ay_r\rangle_Y = \langle Ay_r, d - Ay_r\rangle_D.$$
Since $Ay_r \to d$, we obtain the desired result.
3. This follows immediately from the preceding results and the fact that
$$r^2||Wy_r||_E^2 \leq J_r \leq J_\infty \leq ||d||_D^2.$$
4. This follows from the first two results.

□

4.3 Limiting behavior as $r \uparrow \infty$

We now turn our attention to the limiting behavior of y_r and J_r as $r \uparrow \infty$.

Theorem 9 *For any $d \in D$, $J_r \uparrow J_\infty$ and $y_r \to y_\infty$ as $r \uparrow \infty$.*

Proof: We have seen that

$$\|N_r^{-1}\| \le a$$

for all $r \ge 1$. It follows that $\{y_r\}$ is bounded. Suppose that $r_k \uparrow \infty$ and $y_{r_k} \to \overline{y}$ weakly. Then

$$\|A\overline{y} - d\|_D \le \liminf \|Ay_{r_k} - d\|_D \le \|Ay_\infty - d\|_D.$$

Moreover, $\|Wy_{r_k}\|_E \downarrow 0$, so that

$$W\overline{y} = 0.$$

These facts imply that $\overline{y} = y_\infty$. By the usual argument, we see that, in fact, $y_r \to y_\infty$ weakly. In addition, we can conclude that

$$Ay_r \to Ay_\infty \text{ and } Wy_r \to Wy_\infty.$$

(if it were not the case that $Ay_r \to Ay_\infty$, then we would obtain

$$\|Ay_\infty - d\|_D < \liminf \|Ay_{r_k} - d\|_D \le \|Ay_\infty - d\|_D,$$

a contradiction.) It follows from the joint coercivity of A and W that $y_r \to y_\infty$.

□

An interesting question, given the infinite–dimensional character of the problem

$$\begin{array}{ll} \min & \|Ay - d\|_D^2 \\ \text{s.t.} & Wy = 0, \end{array}$$

is the existence of Lagrange multipliers. The standard theory for the existence of Lagrange multipliers (see [5]) requires that the operator W^* be coercive, which it is not in our examples. In the following result, we show that these is a connection between the existence of Lagrange multipliers and the rate of convergence of y_r to y_∞.

Theorem 10 *If there exists $\lambda \in E$ such that*

$$A^*Ay_\infty - A^*d = W^*\lambda,$$

then

$$\|y_r - y_\infty\|_Y = O(r^{-1}) \text{ as } r \uparrow \infty.$$

Proof: Write

$$\begin{aligned} z_r &= y_r - y_\infty \\ &= N_r^{-1}A^*d - y_\infty \\ &= N_r^{-1}(A^*d - A^*Ay_\infty) \\ &= N_r^{-1}W^*\lambda \\ &= r^{-1}N_r^{-1}rW^*\lambda. \end{aligned}$$

Write $z_r = N_r^{-1}rW^*\lambda$ and note that z_r is the solution of

$$\min_{y} \quad \left\{\|Ay\|_D^2 + \|rWy - \lambda\|_E^2\right\}.$$

It follows that

$$\|Az_r\|_D \leq \|\lambda\|_E$$

and

$$\|rWz_r\|_E \leq 2\|\lambda\|_E.$$

We then obtain

$$\begin{aligned} \|z_r\|_Y^2 &\leq a\left(\|Az_r\|_D^2 + r^2\|Wz_r\|_E^2\right) \\ &\leq 5a\|\lambda\|_E^2, \end{aligned}$$

from which we see that z_r is bounded as $r \uparrow \infty$. This gives the desired result.

□

The following result is the counterpart of Theorem 6.

Theorem 11 *For any $d \in D$, $\|y_r - y_\infty\|_Y = O(r^{-2})$ if and only if*

$$A^*Ay_\infty - A^*d \in \mathcal{R}(W^*W),$$

or, equivalently, there exists $\lambda \in \mathcal{R}(W)$ such

$$A^*Ay_\infty - A^*d = W^*\lambda.$$

Proof: Recall from the previous proof that

$$\begin{aligned} y_r - y_\infty &= N_r^{-1}(A^*d - A^*Ay_\infty) \\ &= r^{-2}(r^{-2}A^*A + W^*W)^{-1}(A^*d - A^*Ay_\infty). \end{aligned}$$

Therefore, the condition $\|y_r - y_\infty\|_Y = O(r^{-2})$ is equivalent to the quantity

$$(r^{-2}A^*A + W^*W)^{-1}(A^*d - A^*Ay_\infty)$$

being bounded as $r \uparrow \infty$, which is equivalent, by Lemma (5), to

$$A^*Ay_\infty - A^*d \in \mathcal{R}(W^*W).$$

□

5 Abstract formulation: outer problem

We now turn our attention to the dependence of the problem on the parameters having a nonlinear influence. This necessarily requires more specific hypotheses in order that the results of the analysis apply to the examples that we have in mind. (After all, specifying the linear influence of y in the forward map and the constraint defines almost all of the structure of the inner problem; on the other hand, postulating a nonlinear dependence of A on x is a negative statement that yields no information whatsoever.) We will make many specific and strong assumptions on the Hilbert space structure and the operators in this section and the next. These assumptions will be justified by two facts: they are satisfied in the examples, and they allow us to prove behavior of the DSO method which is significant and consistent with that seen in the examples.

One aspect of the problem for which we do not have a satisfactory understanding is the issue of smoothness of the DSO objective function. As was mentioned in Section 2, the DSO objective function is smooth, while the OLS objective function is not. It may very well be that this smoothness can be proven from certain generic assumptions which we describe below. However, we have been unable to decide this question one way or the other, and so we must assume that the necessary results can be proven for each individual example.

As the results of the previous section show, the problem dual to the OLS formulation is the key to the DSO method. If the dual problem has properties that make it amenable to local optimization methods, then we can, through the DSO formulation, take advantage of those properties to design an algorithm which will find the global minimizer of the OLS objective function in spite of its extreme nonconvexity. However, there is certainly nothing in our generic structure which states that the dual problem should have any particular properties, good or bad. We will have to assume that the dual problem has certain properties; we can then show that the DSO method leads to a successful algorithm.

5.1 Hilbert spaces

We will assume that

$$\cdots \subset Y_k \subset \cdots \subset Y_1 \subset Y_0$$

is a nested sequence of separable inner product spaces with the following properties:

- Y_0 is a Hilbert space with respect to the norm $\|\cdot\|_{Y_0}$, and each Y_k is dense in Y_0 with respect to this norm;
- for every k, there is a norm $\|\cdot\|_{Y_k}$ under which Y_k is a Hilbert space, and $\|\cdot\|_{Y_{k+1}}$ is stronger than $\|\cdot\|_{Y_k}$ for all k;

Moreover, we shall denote by Y_{-k} the dual space of Y_k under the (formal) pairing $\langle\cdot,\cdot\rangle_{Y_0}$. We then have the further nesting

$$\cdots \subset Y_k \subset \cdots \subset Y_1 \subset Y_0 \subset Y_{-1} \subset \cdots \subset Y_{-k} \subset \cdots,$$

and Y_0 is dense in Y_{-k} for all positive integers k.

The model parameters which appear linearly in the forward map (for example, the source waveform in our examples) will be represented by vectors y belonging to some Y_k. The sequence $\{D_n\}_{-\infty}^{\infty}$ will have the same properties as $\{Y_n\}_{-\infty}^{\infty}$ and will represent data spaces.

The spaces used in our analysis in the last two sections have thus been embedded in these Hilbert space structures. Specifically, we have the following correspondence:

Section 3	Section 5
Y	Y_1
D	D_0
E	Y_0

5.2 Operators

We now wish to define the concept of order for linear operators. Order is defined by the following properties:

- If L is an operator of order k, then L is a linear transformation from $Y_{-\infty}$ to $Y_{-\infty}$, where $Y_{-\infty} = \cup_{n=-\infty}^{\infty} Y_n$.
- If L is an operator of order k, then $L\mid_{Y_n} \in \mathcal{L}(Y_n, Y_{n-k})$ for all integers n.
- If L is an operator of order k and K is an operator of order j, then $[L, K]$ is an operator of order $k + j - 1$.
- If L is an operator of order k, then L^T is also an operator of order k, where $L^T \in \mathcal{L}(Y_0, Y_{-k})$ is the Banach space adjoint of $L \in \mathcal{L}(Y_k, Y_0)$.

For an operator from $Y_{-\infty}$ to $D_{-\infty}$, the concept of order will be defined in exactly the same fashion. This concept of order mimics the order of a differential operator, and will appear in an essential way in the calculations we carry out in the following analysis.

5.3 Regularity of the forward map

In the type of examples that we have in mind, the regularity of the forward map is a very delicate issue. We will illustrate this by considering the forward map S in the layer–over–half–space problem.

Recall that $S : U \to \mathcal{L}(H_b^1, L^2(\Omega))$ is defined by

$$S(c, d)f(h, t) = \frac{\partial f}{\partial t}\left(h, t - \frac{2\sqrt{d^2 + h^2}}{c}\right).$$

Because of the translation operation inherent in this forward map, the map is not smooth. In fact, S is not even continuous with respect to the uniform (operator) topology on $\mathcal{L}(H_b^1, L^2(\Omega))$ (it is, however, continuous with respect to the strong operator topology on $\mathcal{L}(H_b^1, L^2(\Omega))$). We note the following facts about S:

- If S were defined to be a map from U into $\mathcal{L}(H_b^3, L^2(\Omega))$, where

$$H_b^3 = H_b^1 \bigcap H^3(\Omega),$$

then S would be C^1 with respect to the uniform topology on $\mathcal{L}(H_b^3, L^2(\Omega))$.

- Viewing the same matter a different way, there exists a dense subspace of H_b^1, namely H_b^3, such that if $f \in H_b^3$, then

$$S(\cdot,\cdot)f,$$

(regarded as a map from U to $L^2(\Omega)$) is C^1.

- If $f \in H_b^3$, then

$$\begin{aligned} D_{(c,d)}S(c,d)(\delta c, \delta d)f(h,t) &= \frac{\partial^2 f}{\partial t^2}\left(h, t - \frac{2\sqrt{d^2+h^2}}{c}\right)\cdot \\ &\quad \left(\frac{2\sqrt{d^2+h^2}}{c^2}\delta c - \frac{2d}{c\sqrt{d^2+h^2}}\delta d\right) \\ &= S(c,d)Q(c,d)(\delta c, \delta d)f(h,t), \end{aligned}$$

where

$$Q(c,d)(\delta c, \delta d)f(h,t) = \frac{\partial f}{\partial t}(h,t)\left(\frac{2\sqrt{d^2+h^2}}{c^2}\delta c - \frac{2d}{c\sqrt{d^2+h^2}}\delta d\right).$$

- The operators $S(c,d)$ and $\partial/\partial x$ are both operators of order 1, as is $Q(c,d)(\delta c, \delta d)$.

- The operator $Q(c,d)(\delta c, \delta d)$ is antisymmetric, that is,

$$Q(c,d)(\delta c, \delta d)^T = -Q(c,d)(\delta c, \delta d).$$

Here the transpose symbol denotes the formal transpose operator,

$$Q(c,d)(\delta c, \delta d)^T : L^2(\Omega) \to H_b^{-1},$$

where H_b^{-1} is the dual space of H_b^1 under the L^2 pairing.

- $S(c,d)^T S(c,d) = -\partial^2/\partial t^2$ is smooth (in fact, constant) with respect to (c,d).

- Finally, if we compute the derivative of the DSO objective function J_r with respect to (c,d), carrying out the computations formally using the factorization $DS = SQ$, we find that the derivative is defined in terms of operators that extend continuously to H_b^1. Since it can be shown (by a straightforward adaptation of the argument in [12]) that J_r is in fact C^∞, these formal calculations must give the derivative of J_r.

This last point is extremely important to the analysis that we carry out in the abstract setting.

Guided by these properties of S, we make the following assumptions about our forward map A and its transpose $A^T : U \to \mathcal{L}(D_0, Y_{-1})$:

(B1) X is a finite–dimensional Hilbert space, and U is an open subset of X.

(B2) $A : U \to \mathcal{L}(Y_1, D_0)$ has the property that $A(x)$ extends to an operator of order 1 for all $x \in U$ (which implies that $A^T(x)$ is also an operator of order 1 for all $x \in U$).

(B3) $A(x)$ is injective for all $x \in U$.

(B4) $\|A(\cdot)\|$ is continuous.

(B5) $W \in \mathcal{L}(Y_1, Y_0)$ extends to an operator of order 1.

(B6) For some $n \geq 1$, $A(x)y$ is continuously differentiable as a function of x for each $y \in Y_n$, and $D_x A(x)\delta x = A(x)Q(x)\delta x$, where $Q : U \to \mathcal{L}(Y_1, Y_0)$ is of class C^1 and extends to an operator of order 1.

(B7) $A^T(x)d$ is continuously differentiable as a function of x for each $d \in D_{n-1}$.

(B8) The symmetric part $Q_0(x)\delta x$ of $Q(x)\delta x$ extends to an operator of order 0.

(B9) For each integer k, the operator $A^T(x)A(x) + W^T W$ is coercive as a map from Y_k to Y_{k-2}, with coercivity constant independent of $x \in U$ and k; that is,

$$\|(A^T(x)A(x) + W^T W)y\|_{Y_{k-2}} \geq a^{-1}\|y\|_{Y_k}$$

for all $x \in U$, $y \in Y_k$, and k.

(B10) The DSO objective function J_r is smooth.

We will write $N_r = N_r(x) = A^*(x)A(x) + r^2 W^* W$ and $\tilde{N}_r = \tilde{N}_r(x) = A^T(x)A(x) + r^2 W^T W$.

Note in particular Assumption (B9); this is analogous to assuming that the normal operator $A^T A + W^T W$ is elliptic. It is also important to realize that we are *assuming* that J_r is smooth; given this assumptions, we can use the structure of the forward map to compute derivatives formally and be assured that these computations are valid. It would be desirable to know whether the smoothness of J_r follows from the other assumptions; that remains an open question.

5.4 Assumptions on the dual problem

We are going to make quite liberal assumptions on the properties of the dual problem, which takes the form

$$\min_x \quad J_0(x),$$

where

$$\begin{aligned} J_0(x) &= \min \left\{\|Wy\|^2_{Y_0} : A(x)y = d\right\} \\ &= \|Wy_0(x, d)\|^2_{Y_0}, \end{aligned}$$

and $y_0(x,d)$ is the unique element of Y_1 satisfying $A(x)y = d$. When the dependence on the data d needs to be emphasized, we will write $J_0 = J_0(\cdot, d)$ and similarly for J_r and J_∞. Of course, this dual problem is only well–defined when $d \in \mathcal{R}(A(x))$. In fact, since x varies in the set U, the dual problem is really only well–defined when

$$d \in \bigcap_{x \in U} \mathcal{R}(A(x)).$$

Our strategy is to the derive properties of the DSO objective function for noisy data from the properties of the dual problem for nearby exact data.

Since, in the examples, the range of $A(x)$ can only be expected to be dense in the space of actual data, this means that the dual problem is ill–posed; it is only defined for a dense set of data, and when it is defined, $y_0(x,d)$ is not a continuous function of d. This explains why regularization is necessary.

We assume that there is a compact subset K of U, a point $x^* \in K$, a neighborhood N of x^*, and a point $y^* \in Y_1$ such that

(C1) $d^* = A(x^*)y^*$ and $Wy^* = 0$;

(C2) x^* is a strong local minimizer of $J_0(\cdot, d^*)$; that is, at x^* the gradient of $J_0(\cdot, d^*)$ is zero and the Hessian is positive definite;

(C3) $J_0(\cdot, d^*)$ is strongly convex on N (that is, $\nabla^2 J_0(\cdot, d^*)$ is positive definite on N);

(C4) the gradient of $J_0(\cdot, d^*)$ is bounded away from zero on $K \setminus N$.

Under these assumptions, we will be able to prove that there exists $\delta > 0$ and $r_0 > 0$ such that:

- for any $d \in B_\delta(d^*)$, $J_{r_0}(\cdot, d)$ has a unique minimizer $x_{r_0}(d)$ in K;
- the mapping $d \rightarrow x_{r_0}(d)$ can be extended to a continuous function $x : [r_0, \infty) \times B_\delta(d^*) \rightarrow X$ with the property that $x_r(d)$ is a strong local minimizer of $J_r(\cdot, d)$ for each r and d, and for any fixed d, $x_r(d) \rightarrow x_\infty(d)$, a local minimizer of $J_\infty(\cdot, d)$.

Obtaining the second of these results will, however, require that we restrict the data to have some degree of "smoothness", that is, to belong to D_k for some $k > 0$. We consider these results to be qualitatively very satisfying. In the applications we have in mind, we would like to find the global minimizer of J_∞ (the Output Least–Squares objective function). However, this function typically has multiple local minimizers, which leaves no efficient way to locate the global minimizer. The above theorems show that if the dual problem has a unique minimizer in the case of noise–free data, then the DSO objective function can be minimized by local optimization techniques provided that the level of noise is not too great and provided that the penalty parameter r is chosen correctly. Moreover, the existence of a parameterized path of minimizers of J_r means that we can go from the solution of the DSO problem for "small" r to the solution of the OLS problem. Roughly speaking, it is possible to get on the path for small r (because of the uniqueness of the minimizer) and to follow it to the desired (OLS) solution.

6 Analysis of the outer problem

We begin our analysis of the outer problem by computing the derivatives of the functions J_r and J_0. We then use the resulting formulas to prove that $J_r(\cdot,d^*)$ converges to $J_0(\cdot,d^*)$ in the C^2 sense as $r\downarrow 0$, and to derive C^2 bounds on $J_r(\cdot,d)-J_r(\cdot,d^*)$ in terms of $\|d-d^*\|_{D_0}$ and r. These will form the basis from which we will prove the theorems referred to in the previous section.

6.1 Derivatives

Lemma 6 *If $d\in D_{n-1}$, then, for all $x\in U$, $y_r(x,d)\in Y_n$ and $y_\infty(x,d)\in Y_n$.*

Proof: The fact that $y_r(x,d)\in Y_n$ follows immediately from Assumption (B9). Moreover, the same assumption shows that the operator $\tilde{A}(x)$, defined to be $A(x)$ restricted to the null space of W, has a normal operator which is coercive as an operator from Y_n to Y_{n-2}. The result now follows.

□

Lemma 7 *If $d\in D_{n-1}$, then the mappings*

$$x\longrightarrow y_\infty(x,d)$$

and

$$x\longrightarrow y_r(x,d)$$

are continuously differentiable.

Proof: The fact that $x\longrightarrow y_r(x,d)$ is C^1 follows immediately from the previous lemma and Assumption (B6). The analogous result for y_∞ can be deduced by considering the operator $\tilde{A}$ defined in the previous proof.

□

Theorem 12 *For any $d\in D_0$, the derivative of J_r with respect to x is given by*

$$DJ_r(x,d)\delta x=-r^2\langle V(x)\delta x y_r(x,d),Wy_r(x,d)\rangle_0,$$

where

$$V(x)\delta x=\frac{1}{2}Q_0(x)\delta xW+[W,Q(x)\delta x]$$

is an operator of order 1.

Proof: We assume for now that $d\in D_{n-1}$; by Assumption (B9), this implies that $y_r(x,d)\in Y_n$ for all $x\in U$, and hence, by Assumption (B6), that $x\longrightarrow A(x)y_r(x,d)$ is a C^1 map. We compute the derivative of J_r as follows (we write $A = A(x)$, $DA = DA(x)\delta x$, $Dy_r = Dy_r(x,d)\delta x$, and so forth):

$$\begin{aligned} J_r(x,d) &= \frac{1}{2}\langle Ay_r - d, Ay_r - d\rangle_0 + \frac{r^2}{2}\langle Wy_r, Wy_r\rangle_0 \\ DJ_r(x,d)\delta x &= \langle DAy_r + ADy_r, Ay_r - d\rangle_0 + r^2\langle WDy_r, Wy_r\rangle_0 \\ &= \langle AQy_r, Ay_r - d\rangle_0 + \langle Dy_r(x,d), A^*Ay_r - A^*d\rangle_1 + r^2\langle Dy_r, W^*Wy_r\rangle_1 \\ &= \langle Qy_r, A^*Ay_r - A^*d\rangle_1 + \langle Dy_r, N_r y_r - A^*d\rangle_1. \end{aligned}$$

By the normal equations, $N_r y_r - A^*d = 0$ and $A^*Ay_r - A^*d = -r^2W^*Wy_r$. Therefore,

$$\begin{aligned} DJ_r(x,d)\delta x &= -r^2\langle Qy_r, W^*Wy_r\rangle_1 \\ &= -r^2\langle WQy_r, Wy_r\rangle_0 \\ &= -r^2\langle QWy_r, Wy_r\rangle_0 - r^2\langle [W,Q]y_r, Wy_r\rangle_0 \\ &= -r^2\langle \frac{1}{2}Q_0Wy_r, Wy_r\rangle_0 - r^2\langle [W,Q]y_r, Wy_r\rangle_0 \end{aligned}$$

(note how we used the relationship $Q^T = -Q + Q_0$). Since this formula extends continuously to $d \in D_0$, and since J_r is assumed to be smooth, the proof is complete.

□

Theorem 13 *For any $d \in D_{n-1}$, the derivative of y_r with respect to x is given by*

$$\begin{aligned} Dy_r(x,d)\delta x &= -Q(x)\delta x y_r(x,d) \\ &+ r^2\tilde{N}_r(x)^{-1}\left\{Q_0(x)\delta x W^TW - [Q(x)\delta x, W^TW]\right\} y_r(x,d). \end{aligned}$$

Proof: We have

$$\tilde{N}_r y_r = A^T d,$$

whence

$$\begin{aligned} \tilde{N}_r Dy_r &= DA^Td - DA^TAy_r - A^TDAy_r \\ &= Q^TA^Td - Q^TA^TAy_r - A^TAQy_r \\ &= Q_0(A^Td - A^TAy_r) - Q(A^Td - A^TAy_r) - A^TAQy_r \\ &= r^2Q_0W^TWy_r - r^2QW^TWy_r - A^TAQy_r \\ &= r^2Q_0W^TWy_r - r^2[Q, W^TW]y_r - \tilde{N}_rQy_r. \end{aligned}$$

Thus

$$Dy_r = -Qy_r + r^2\tilde{N}_r^{-1}\left\{Q_0W^TWy_r - [Q, W^TW]y_r\right\}.$$

□

Although the above result holds only for "smooth" data, we can use it, together with the fact that J_r is smooth, to derive the formula for the second derivative of J_r.

Theorem 14 *For any $d \in D_0$, we have*

$$\begin{aligned}
D^2 J_r(x,d)(\delta x,\delta x) &= r^2\langle [W,Q(x)\delta x]y_r(x,d),[W,Q(x)\delta x]y_r(x,d)\rangle_0 \\
&+ r^2\langle Wy_r(x,d),B(x)(\delta x,\delta x)y_r(x,d)\rangle_0 \\
&- r^4\langle (Q_0(x)\delta x W^TW-[Q(x)\delta x,W^TW])y_r(x,d), \\
&\quad \tilde{N}_r(x)^{-1}\cdot \\
&\quad (Q_0(x)\delta x W^TW-[Q(x)\delta x,W^TW])y_r(x,d)\rangle_{-1,1},
\end{aligned}$$

where

$$\begin{aligned}
B(x)(\delta x,\delta x) &= \frac{1}{2}(Q_0(x)\delta x)^2W+\frac{1}{2}[Q_0(x)\delta x,Q(x)\delta x]W \\
&+ 2Q_0(x)\delta x[W,Q(x)\delta x]+[[W,Q(x)\delta x],Q(x)\delta x] \\
&- \frac{1}{2}DQ_0(x)(\delta x,\delta x)W-[W,DQ(x)(\delta x,\delta x)]
\end{aligned}$$

is an operator of order 1.

Proof: We assume for now that $d \in D_{n-1}$. The extension to all of D_0 will follow immediately from the assumption that J_r is smooth and the fact that the formula given in the theorem extends continuously to D_0. We begin with the formula

$$DJ_r(x,d)\delta x = -r^2\langle (\frac{1}{2}Q_0W+[W,Q])y_r,Wy_r\rangle_0.$$

Differentiating, we obtain

$$\begin{aligned}
D^2J_r(x,d)(\delta x,\delta x) &= -r^2\left\{\frac{1}{2}\langle DQ_0Wy_r,Wy_r\rangle_0-\frac{1}{2}\langle Q_0WQy_r,Wy_r\rangle_0\right. \\
&+ \frac{1}{2}r^2\langle Q_0WN_r^{-1}(Q_0W^TW-[Q,W^TW])y_r,Wy_r\rangle_0 \\
&- \frac{1}{2}\langle Q_0Wy_r,WQy_r\rangle_0 \\
&+ \frac{1}{2}r^2\langle Q_0Wy_r,WN_r^{-1}(Q_0W^TW-[Q,W^TW])y_r\rangle_0 \\
&+ \langle [W,DQ]y_r,Wy_r\rangle_0-\langle [W,Q]Qy_r,Wy_r\rangle_0 \\
&+ r^2\langle [W,Q]N_r^{-1}(Q_0W^TW-[Q,W^TW])y_r,Wy_r\rangle_0 \\
&- \langle [W,Q]y_r,WQy_r\rangle_0 \\
&+ \left. r^2\langle [W,Q]y_r,WN_r^{-1}(Q_0W^TW-[Q,W^TW])y_r\rangle_0\right\}.
\end{aligned}$$

We look at the $O(r^2)$ terms first:

$$\begin{aligned}
&\frac{1}{2}\langle DQ_0Wy_r,Wy_r\rangle_0-\frac{1}{2}\langle Q_0WQy_r,Wy_r\rangle_0-\frac{1}{2}\langle Q_0Wy_r,WQy_r\rangle_0+ \\
&\langle [W,DQ]y_r,Wy_r\rangle_0-\langle [W,Q]Qy_r,Wy_r\rangle_0-\langle [W,Q]y_r,WQy_r\rangle_0 \\
=\ &\frac{1}{2}\langle DQ_0Wy_r,Wy_r\rangle_0-\langle Q_0WQy_r,Wy_r\rangle_0-\langle [W,Q]y_r,[W,Q]y_r\rangle_0-
\end{aligned}$$

$$\begin{aligned}
& \langle [W,Q]y_r, QWy_r\rangle_0 - \langle [W,Q]Qy_r, Wy_r\rangle_0 + \langle [W,DQ]y_r, Wy_r\rangle_0 \\
= \; & \frac{1}{2}\langle DQ_0Wy_r, Wy_r\rangle_0 - \langle [W,Q]y_r, [W,Q]y_r\rangle_0 + \langle [W,DQ]y_r, Wy_r\rangle_0 - \\
& \langle Q[W,Q]y_r, Wy_r\rangle_0 - \langle [[W,Q],Q]y_r, Wy_r\rangle_0 + \langle Q[W,Q]y_r, Wy_r\rangle_0 - \\
& \langle Q_0[W,Q]Y_r, Wy_r\rangle_0 - \langle Q_0WQy_r, Wy_r\rangle_0 \\
= \; & \frac{1}{2}\langle DQ_0Wy_r, Wy_r\rangle_0 - \langle [W,Q]y_r, [W,Q]y_r\rangle_0 + \langle [W,DQ]y_r, Wy_r\rangle_0 - \\
& \langle [[W,Q],Q]y_r, Wy_r\rangle_0 - \langle 2Q_0[W,Q]y_r, Wy_r\rangle_0 - \\
& \langle Q_0QWy_r, Wy_r\rangle_0 .
\end{aligned}$$

Note that

$$\begin{aligned}
\langle Q_0QWy_r, Wy_r\rangle_0 &= \langle [Q_0,Q]Wy_r, Wy_r\rangle_0 + \langle QQ_0Wy_r, Wy_r\rangle_0 \\
&= \langle [Q_0,Q]Wy_r, Wy_r\rangle_0 - \langle Q_0Wy_r, QWy_r\rangle_0 + \\
&\quad \langle Q_0^2Wy_r, Wy_r\rangle_0 ,
\end{aligned}$$

so

$$\langle Q_0QWy_r, Wy_r\rangle_0 = \frac{1}{2}\langle [Q_0,Q]Wy_r, Wy_r\rangle_0 + \frac{1}{2}\langle Q_0^2Wy_r, Wy_r\rangle_0 .$$

We therefore have obtained

$$\begin{aligned}
D^2J_r(x,d)(\delta x,\delta x) &= -r^2\,\{\ \frac{1}{2}\langle DQ_0Wy_r, Wy_r\rangle_0 - \langle [W,Q]y_r, [W,Q]y_r\rangle_0 + \\
&\quad \langle [W,DQ]y_r, Wy_r\rangle_0 - \langle [[W,Q],Q]y_r, Wy_r\rangle_0 - \\
&\quad \langle 2Q_0[W,Q]y_r, Wy_r\rangle_0 - \frac{1}{2}\langle [Q_0,Q]Wy_r, Wy_r\rangle_0 - \\
&\quad \frac{1}{2}\langle Q_0^2Wy_r, Wy_r\rangle_0\ \} + O(r^4) \\
&= r^2\langle [W,Q]y_r, [W,Q]y_r\rangle_0 + r^2\langle By_r, Wy_r\rangle_0 + O(r^4).
\end{aligned}$$

We now gather the $O(r^4)$ terms:

$$\begin{aligned}
& \langle Q_0W\tilde{N}_r^{-1}(Q_0W^TW - [Q,W^TW])y_r, Wy_r\rangle_0 + \\
& \langle [W,Q]\tilde{N}_r^{-1}(Q_0W^TW - [Q,W^TW])y_r, Wy_r\rangle_0 + \\
& \langle [W,Q]y_r, W\tilde{N}_r^{-1}(Q_0W^TW - [Q,W^TW])y_r\rangle_0 \\
= \; & \langle \tilde{N}_r^{-1}(Q_0W^TW - [Q,W^TW])y_r, W^TQ_0Wy_r\rangle_0 + \\
& \langle \tilde{N}_r^{-1}(Q_0W^TW - [Q,W^TW])y_r, [Q^T,W^T]Wy_r\rangle_0 + \\
& \langle \tilde{N}_r^{-1}(Q_0W^TW - [Q,W^TW])y_r, W^T[W,Q]y_r\rangle_0 .
\end{aligned}$$

Since

$$\begin{aligned}
& W^TQ_0W + [Q^T,W^T]W + W^T[W,Q] \\
= \; & W^TQ_0W - [Q,W^T]W + [Q_0,W^T]W + W^T[W,Q] \\
= \; & W^TQ_0W - QW^TW + W^TQW + \\
& Q_0W^TW - W^TQ_0W + W^TWQ - W^TQW \\
= \; & Q_0W^TW - [Q,W^TW],
\end{aligned}$$

we have derived the desired result.

□

We are not assuming that J_0 is smooth. However, we proceed to compute the first and second derivatives of J_0 formally. Later in this section, we will show that under certain conditions, the differentiability of J_0 may be deduced from that of J_r.

Theorem 15 *If J_0 is well–defined and of class C^1, and if the mapping $x \to y_0(x,d)$ is well–defined and of class C^1, then*

$$DJ_0(x,d)\delta x = -\langle V(x)\delta x y_0(x,d), W y_0(x,d)\rangle_0,$$

where V is the operator in Theorem 12.

Proof: Note that

$$Ay_0 = d,$$

so

$$DAy_0 + ADy_0 = 0,$$

or

$$ADy_0 = -AQy_0.$$

Since A is assumed to be injective, we obtain $Dy_0 = -Qy_0$. Therefore,

$$\begin{aligned} DJ_0(x,d)\delta x &= \langle WDy_0, Wy_0\rangle_0 \\ &= -\langle WQy_0, Wy_0\rangle_0 \\ &= -\langle [W,Q]y_0, Wy_0\rangle_0 - \langle QWy_0, Wy_0\rangle_0 \\ &= -\langle \left(\frac{1}{2}Q_0 W + [W,Q]\right) y_0, Wy_0\rangle_0 \end{aligned}$$

(where we used, as before, the relation $Q^T = -Q + Q_0$). This gives the desired result.

□

Theorem 16 *If J_0 is well–defined and of class C^2, and if the mapping $x \to y_0(x,d)$ is well–defined and of class C^1, then*

$$\begin{aligned} & D^2 J_0(x,d)(\delta x, \delta x) \\ = \ & \|[W, Q(x)\delta x y_0(x,d)\|_0^2 + \langle W y_0(x,d), B(x)(\delta x,\delta x) y_0(x,d)\rangle_0, \end{aligned}$$

where B is the operator defined in Theorem 14.

Proof: We have

$$\begin{aligned} D^2 J_0(x,d)(\delta x,\delta x) &= -\frac{1}{2}\langle DQ_0 W y_0, W y_0\rangle_0 - \langle [W, DQ] y_0, W y_0\rangle_0 + \\ & \quad \frac{1}{2}\langle Q_0 W Q y_0, W y_0\rangle_0 + \langle [W,Q] Q y_0, W y_0\rangle_0 + \end{aligned}$$

$$
\begin{aligned}
& \frac{1}{2}\langle Q_0 W y_0, W Q y_0\rangle_0 + \langle [W,Q] y_0, W Q y_0\rangle_0 \\
= \;& -\frac{1}{2}\langle D Q_0 W y_0, W y_0\rangle_0 - \langle [W, DQ] y_0, W y_0\rangle_0 + \\
& \frac{1}{2}\langle Q_0 W y_0, Q W y_0\rangle_0 + \frac{1}{2}\langle Q_0 W y_0, [W,Q] y_0\rangle_0 + \\
& \langle [W,Q] y_0, Q W y_0\rangle_0 + \langle [W,Q] y_0, [W,Q] y_0\rangle_0 + \\
& \frac{1}{2}\langle Q_0 Q W y_0, W y_0\rangle_0 + \frac{1}{2}\langle Q_0 [W,Q] y_0, W y_0\rangle_0 + \\
& \langle Q[W,Q] y_0, W y_0\rangle_0 + \langle [[W,Q],Q] y_0, W y_0\rangle_0 \\
= \;& \langle [W,Q] y_0, [W,Q] y_0\rangle_0 - \frac{1}{2}\langle D Q_0 W y_0, W y_0\rangle_0 - \\
& \langle [W, DQ] y_0, W y_0\rangle_0 + \frac{1}{2}\langle Q_0^2 W y_0, W y_0\rangle_0 + \\
& \frac{1}{2}\langle [Q_0, Q] W y_0, W y_0\rangle_0 + 2\langle Q_0 [W,Q] y_0, W y_0\rangle_0 + \\
& \langle [[W,Q],Q] y_0, W y_0\rangle_0 \\
= \;& \langle [W,Q] y_0, [W,Q] y_0\rangle_0 + \langle B(x)(\delta x, \delta x) y_0, W y_0\rangle_0 .
\end{aligned}
$$

□

6.2 Convergence of J_r/r^2 to J_0

We next wish to show that, under certain conditions, J_r/r^2 converges to J_0 in the C^2 sense. We will do this by showing that J_r/r^2 converges uniformly to J_0 on a compact set, and that the first and second derivatives of J_r/r^2 also converge uniformly on that set. This will suffice to show that J_0 is twice–continuously differentiable, and to verify the C^2 convergence.

Theorem 17 *Let K be a compact subset of $U \subset X$, assume $d \in \bigcap_{x\in K} \mathcal{R}(A(x))$, and assume that for each $x \in K$, there exists $u_0(x,d) \in Y_1$ such that $W^*Wy_0(x,d) = A^*(x)A(x)u_0(x,d)$. Assume further that the maps $x \to y_0(x,d)$ and $x \to u_0(x,d)$ are continuous. Then*

$$\|\frac{1}{r^2} J_r(\cdot,d) - J_0(\cdot,d)\|_{C^0(K)} = O(r^2).$$

Proof: For the sake of brevity, we omit explicit reference to the dependence of the various quantities on x and d. The proof of Lemma 5 shows that

$$\|y_r - y_0\|_1 \le r^2 \|u_0\|_1 (1 + a\|W\|^2).$$

We thus obtain

$$
\begin{aligned}
|\frac{1}{r^2} J_r - J_0| &= \frac{1}{2} |r^{-2}\|Ay_r - d\|_{D_0}^2 + \|Wy_r\|_0^2 - \|Wy_0\|_0^2| \\
&\le \frac{1}{2} r^{-2} \|A(y_r - y_0)\|_0^2 + \frac{1}{2}(\|Wy_r\|_0 + \|Wy_0\|_0) \cdot \|W(y_r - y_0)\|_0
\end{aligned}
$$

$$\begin{aligned}
&\le \frac{1}{2}r^{-2}\|A\|^2\|y_r - y_0\|_1^2 + \|Wy_0\|_0\|W\|\|y_r - y_0\|_1 \\
&\le r^2\|u_0\|(1 + a\|W\|^2)\left(\frac{1}{2}\|A\|\|u_0\|(1 + a\|W\|^2) + \|W\|\|Wy_0\|_0\right).
\end{aligned}$$

The compactness of K and the continuity of the various operations involved gives a bound that is uniform for $x \in K$. This completes the proof.

□

Theorem 18 *Under the hypotheses of Theorem 17, $J_0(\cdot, d)$ is of class C^1, and*

$$\|\frac{1}{r^2}\nabla J_r(x,d) - \nabla J_0(x,d)\|_{C^0(K)} = O(r^2).$$

Proof: We prove the result by showing that the difference between the first derivative of J_r/r^2 and the (formal) derivative of J_0 is $O(r^2)$ uniformly over K as $r \downarrow 0$. This will both show that J_0 is of class C^1 and give the desired rate of convergence. We have

$$\begin{aligned}
|\frac{1}{r^2}DJ_r(x,d)\delta x - DJ_0(x,d)\delta x| &= |\langle Vy_r, Wy_r\rangle_0 - \langle Vy_0, Wy_0\rangle_0| \\
&\le |\langle Vy_r, Wy_r\rangle_0 - \langle Vy_0, Wy_r\rangle_0| + \\
&\quad\; |\langle Vy_0, Wy_r\rangle_0 - \langle Vy_0, Wy_0\rangle_0| \\
&\le \|V\|\|\delta x\|_X\|W\|\|y_r\|_1\|y_r - y_0\|_1 + \\
&\quad\; \|V\|\|\delta x\|_X\|W\|\|y_0\|_1\|y_r - y_0\|_1 \\
&\le \|V\|\|\delta x\|_X\|W\|(\|y_r\|_1 + \|y_0\|_1)\|y_r - y_0\|_1 \\
&\le 2\|V\|\|\delta x\|_X\|W\|\|y_0\|_1\|y_r - y_0\|_1
\end{aligned}$$

for r sufficiently small (since $y_r \to y_0$). Recalling our bound for $\|y_r - y_0\|_1$, we obtain

$$|\frac{1}{r^2}\nabla J_r(x,d) - \nabla J_0(x,d)| \le 2r^2\|u_0\|_1\left(1 + a\|W\|^2\right)\|y_0\|_1\|V\|\|W\|.$$

Since, by hypothesis, all of the relevant operations are continuous over the compact set K, we obtain the necessary uniform bound.

□

The fundamental result which leads to the above convergence rates is Theorem 6, which gives the condition under which $y_r - y_0 = O(r^2)$ as $r \downarrow 0$. However, this condition is not strong enough to yield $O(r^2)$ convergence of Dy_r to Dy_0, even when these derivatives exist. In order to obtain this rate of convergence, and from it the C^2 convergence of J_r/r^2 to J_0, we need the following results.

Lemma 8 *For $r \in (0,1)$,*

$$\|\tilde{N}_r^{-1}A^TA\| = \|N_r^{-1}A^*A\| \le 1 + a\|W\|^2.$$

Proof: We have

$$\begin{aligned}||N_r^{-1}A^*A|| &= ||N_r^{-1}(A^*A + r^2W^*W - r^2W^*W)|| \\ &= ||I - r^2N_r^{-1}W^*W|| \\ &\le 1 + r^2\frac{a}{r^2}||W||^2 \\ &= 1 + a||W||^2.\end{aligned}$$

Note also that, by the Riesz representation theorem, $(A^TA+r^2W^TW)^{-1}A^TA = (A^*A+r^2W^*W)^{-1}A^*A$.

□

Lemma 9 *Assume that $d \in D_0$ and*

$$W^*Wy_0(x,d) = A^*(x)A(x)u_0(x,d).$$

Then

$$\tilde{N}_r(x)^{-1}(Q_0(x)\delta x W^TW - [Q(x)\delta x, W^TW])y_r(x,d)$$

is bounded as $r \downarrow 0$ if and only if

$$(Q_0(x)\delta x W^TW - [Q(x)\delta x, W^TW])y_0(x,d) \in \mathcal{R}(A^T(x)A(x)).$$

Proof: Write

$$L = Q_0W^TW - [Q, W^TW].$$

We have

$$\begin{aligned}&\tilde{N}_r^{-1}Ly_r \\ =\;& \tilde{N}_r^{-1}L\tilde{N}_r^{-1}A^Td \\ =\;& \tilde{N}_r^{-1}L\tilde{N}_r^{-1}A^TAy_0 \\ =\;& \tilde{N}_r^{-1}L\tilde{N}_r^{-1}(A^TA + r^2W^TW - r^2W^TW)y_0 \\ =\;& \tilde{N}_r^{-1}L\left\{y_0 - r^2\tilde{N}_r^{-1}W^TWy_0\right\} \\ =\;& \tilde{N}_r^{-1}Ly_0 - r^2\tilde{N}_r^{-1}L\tilde{N}_r^{-1}A^TAu_0.\end{aligned}$$

By virtue of the previous lemma and our bound on the norm of $||\tilde{N}_r^{-1}||$, the second term is bounded as $r \downarrow 0$. Therefore it follows from Lemma 5 that $\tilde{N}_r^{-1}Ly_r$ is bounded as $r \downarrow 0$ if and only if

$$Ly_0 \in \mathcal{R}(A^TA).$$

□

The condition in the preceding lemma can be regarded as an abstract "smoothness" condition. In the following section, we will show that this is a reasonable assumption in our examples.

Theorem 19 *Let the hypotheses of Theorem 17 hold, and assume that for each $x \in K$ and for each $\delta x \in X$, there exists $v_0(x, d, \delta x) \in Y_1$ such that*

$$\left\{ Q_0(x)\delta x W^T W - [Q(x)\delta x, W^T W] \right\} y_0(x, d) = A^T(x)A(x)v_0(x, d, \delta x).$$

Assume further that $\|v_0(x, d, \delta x)\|_1 \leq C(x)\|\delta x\|_X$, where $C : X \to \mathcal{R}$ is a continuous function. Then

$$\|\frac{1}{r^2}\nabla^2 J_r(x, d) - \nabla^2 J_0(x, d)\|_{C^0(K)} = O(r^2).$$

Proof: We have

$$\begin{aligned}
& |\frac{1}{r^2}D^2 J_r(x, d)(\delta x, \delta x) - D^2 J_0(x, d)(\delta x, \delta x)| \\
\leq\ & |\langle [W, Q]y_r, [W, Q]y_r\rangle_0 + \langle W y_r, B y_r\rangle_0 + \\
& r^2\langle (Q_0 W^T W - [Q, W^T W])y_r, \tilde{N}_r^{-1}(Q_0 W^T W - [Q, W^T W])y_r\rangle_{-1,1} - \\
& \langle [W, Q]y_0, [W, Q]y_0\rangle_0 - \langle W y_0, B y_0\rangle_0| \\
\leq\ & |\langle [W, Q]y_r, [W, Q]y_r\rangle_0 - \langle [W, Q]y_0, [W, Q]y_r\rangle_0| + \\
& |\langle [W, Q]y_0, [W, Q]y_r\rangle_0 - \langle [W, Q]y_0, [W, Q]y_0\rangle_0| + \\
& |\langle W y_r, B y_r\rangle_0 - \langle W y_0, B y_r\rangle_0| + \\
& |\langle W y_0, B y_r\rangle_0 - \langle W y_0, B y_0\rangle_0| + \\
& r^2\|Q_0 W^T W - [Q, W^T W]\|\|y_r\|_1\|\tilde{N}_r^{-1}(Q_0 W^T W - [Q, W^T W])y_r\|_{-1}.
\end{aligned}$$

The $O(r^2)$ convergence of y_r to y_0 and Lemma 9, together with our various continuity assumptions, now yield the desired rate of convergence.

□

Corollary 4 *Under the hypotheses of the preceding theorem, for r sufficiently small, $J_r(\cdot, d^*)$ has a unique stationary point in K, namely the global minimizer x^*.*

Proof: We have seen that J_r/r^2 converges to J_0 in the C^2 sense as $r \downarrow 0$. Therefore, for r sufficiently small, the gradient of J_r/r^2 will be bounded away from zero on $K \setminus N$, and J_r/r^2 will be strongly convex on N. This suffices to give the desired result.

□

6.3 Effect of noise on J_r

We now prove the main results which show the effectiveness of the DSO method. Our reasoning is based on two ideas: the existence of a well–behaved dual objective function J_0, and the existence of a path of minimizers of the penalty functions J_r. By referring to the dual problem, we can show that there exist values of r for which it is possible to find the global minimizer of J_r. The existence of a path converging to the minimizer of J_∞ then shows that the dual problem indeed enables us to find the desired minimizer of J_∞. The first result relies on the following theorem.

Theorem 20 *Assume that there exists $x^* \in K$, $y^* \in Y_1$, and $d^* \in D_0$ such that $A(x^*)y^* = d^*, Wy^* = 0$, and let $d \in D_0$. Assume further that for each $x \in K$, there exists $u_0(x, d^*)$, depending continuously on x, such that $W^*Wy_0(x, d^*) = A^*(x)A(x)u_0(x, d^*)$. Then there exist positive constants c_1, c_2 such that for all r sufficiently small,*

$$\|\frac{1}{r^2}J_r(\cdot, d) - \frac{1}{r^2}J_r(\cdot, d^*)\|_{C^2(K)} \leq c_1\frac{\|d - d^*\|_0}{r} + c_2\frac{\|d - d^*\|_0^2}{r^2}.$$

Proof: By the normal equations, $y_r(x, d)$ is linear in d. It therefore follows from Theorem 8 that

$$\|y_r(x, d) - y_r(x, d^*)\|_1 = \|y_r(x, d - d^*)\|_1 \leq \sqrt{2a}\frac{\|d - d^*\|_0}{r}.$$

Note that

$$\begin{aligned} J_r(x, d) &= \frac{1}{2}\langle A(x)y_r(x, d) - d, A(x)y_r(x, d)\rangle_0 + \frac{r^2}{2}\langle Wy_r(x, d), Wy_r(x, d)\rangle_0 \\ &= \frac{1}{2}\langle y_r(x, d), N_r(x)y_r(x, d) - A^*(x)d\rangle_1 + \frac{1}{2}\langle d, d - A(x)y_r(x, d)\rangle_0 \\ &= \frac{1}{2}\langle d, d - A(x)y_r(x, d)\rangle_0 \end{aligned}$$

by the normal equations. Also,

$$\begin{aligned} \langle d_1, d_2 - A(x)y_r(x, d_2)\rangle_0 &= \langle d_1, (I - A(x)N_r(x)^{-1}A^*(x))d_2\rangle_0 \\ &= \langle (I - A(x)N_r(x)^{-1}A^*(x))d_1, d_2\rangle_0 \\ &= \langle d_1 - A(x)y_r(x, d_1), d_2\rangle_0. \end{aligned}$$

With these relationships, we compute as follows:

$$\begin{aligned} & J_r(x, d) - J_r(x, d^*) \\ = & \frac{1}{2}\left\{\langle d, d - A(x)y_r(x, d)\rangle_0 - \langle d^*, d^* - A(x)y_r(x, d^*)\rangle_0\right\} \\ = & \frac{1}{2}\left\{\langle d - d^*, d - A(x)y_r(x, d)\rangle_0 + \langle d^*, d - d^* - A(x)y_r(x, d - d^*)\rangle_0\right\} \\ = & \frac{1}{2}\{\ \langle d - d^*, d - A(x)y_r(x, d)\rangle_0 - \langle d - d^*, d - d^* - A(x)y_r(x, d - d^*)\rangle_0 + \\ & \langle d, d - d^* - A(x)y_r(x, d - d^*)\rangle_0\ \} \\ = & \frac{1}{2}\left\{2\langle d - d^*, d - A(x)y_r(x, d)\rangle_0 - \langle d - d^*, d - d^* - A(x)y_r(x, d - d^*)\rangle_0\right\}. \end{aligned}$$

Therefore, bearing in mind that $\|y_r(x, d^*) - y_0(x, d^*)\|_1 = O(r^2)$ uniformly in x, we obtain

$$\begin{aligned} & |\frac{1}{r^2}J_r(x, d) - \frac{1}{r^2}J_r(x, d^*)| \\ \leq & 2\|d - d^*\|_0\frac{\|A(x)y_r(x, d^*) - A(x)y_0(x, d^*)\|_0}{r^2} + \\ & \frac{1}{r^2}\|d - d^*\|_0\|d - d^* - A(x)y_r(x, d - d^*)\|_0 \\ \leq & \text{const}\|d - d^*\|_0 + \frac{\|d - d^*\|_0^2}{r^2}, \end{aligned}$$

and thus

$$\|\frac{1}{r^2}J_r(x,d) - \frac{1}{r^2}J_r(x,d^*)\|_{C^0(K)} \leq \text{const}\|d-d^*\|_0 + \frac{\|d-d^*\|_0^2}{r^2}.$$

Next, we have

$$\begin{aligned}
& \frac{1}{r^2}DJ_r(x,d)\delta x - \frac{1}{r^2}DJ_r(x,d^*)\delta x \\
= \quad & \langle V(x)\delta x y_r(x,d^*), Wy_r(x,d^*)\rangle_0 - \langle V(x)\delta x y_r(x,d), Wy_r(x,d)\rangle_0 \\
= \quad & \langle V(x)\delta x y_r(x,d^*-d), Wy_r(d^*\rangle_0 + \langle V(x)\delta x y_r(x,d^*), Wy_r(x,d^*-d)\rangle_0 + \\
& \langle V(x)\delta x y_r(x,d-d^*), Wy_r(x,d^*-d)\rangle_0.
\end{aligned}$$

From this, we obtain

$$\begin{aligned}
& |\frac{1}{r^2}DJ_r(x,d)\delta x - \frac{1}{r^2}DJ_r(x,d^*)\delta x| \\
\leq \quad & \|V(x)\|\|\delta x\|_X\|W\|\left(2\|y_r(x,d^*-d)\|_1\|y_r(x,d^*)\|_1 + \|y_r(x,d^*-d)\|_1^2\right),
\end{aligned}$$

which yields

$$\begin{aligned}
& |\frac{1}{r^2}\nabla J_r(x,d) - \frac{1}{r^2}\nabla J_r(x,d^*)| \\
\leq \quad & \text{const}\frac{\|d-d^*\|_0}{r} + \text{const}\frac{\|d-d^*\|_0^2}{r^2}.
\end{aligned}$$

Finally, we look at the second derivative:

$$\begin{aligned}
& \frac{1}{r^2}D^2J_r(x,d)(\delta x,\delta x) - \frac{1}{r^2}D^2J_r(x,d^*)(\delta x,\delta x) \\
= \quad & \langle [W,Q(x)\delta x]y_r(x,d), [W,Q(x)\delta x]y_r(x,d)\rangle_0 + \\
& \langle Wy_r(x,d), B(x)(\delta x,\delta x)y_r(x,d)\rangle_0 + \\
& r^2\langle (Q_0(x)\delta x W^TW - [Q(x)\delta x, W^TW])y_r(x,d), \\
& \tilde{N}_r(x)^{-1}((Q_0(x)\delta x W^TW - [Q(x)\delta x, W^TW])y_r(x,d)\rangle_0 - \\
& \langle [W,Q(x)\delta x]y_r(x,d^*), [W,Q(x)\delta x]y_r(x,d^*)\rangle_0 - \\
& \langle Wy_r(x,d^*), B(x)(\delta x,\delta x)y_r(x,d^*)\rangle_0 - \\
& r^2\langle (Q_0(x)\delta x W^TW - [Q(x)\delta x, W^TW])y_r(x,d^*), \\
& \tilde{N}_r(x)^{-1}((Q_0(x)\delta x W^TW - [Q(x)\delta x, W^TW])y_r(x,d^*)\rangle_0 \\
= \quad & 2\langle [W,Q(x)\delta x]y_r(x,d^*), [W,Q(x)\delta x]y_r(x,d-d^*)\rangle_0 \\
& \langle [W,Q(x)\delta x]y_r(x,d-d^*), [W,Q(x)\delta x]y_r(x,d-d^*)\rangle_0 + \\
& \langle Wy_r(x,d), B(x)(\delta x,\delta x)y_r(x,d-d^*)\rangle_0 + \\
& \langle Wy_r(x,d-d^*), B(x)(\delta x,\delta x)y_r(x,d)\rangle_0 + \\
& \langle Wy_r(x,d-d^*), B(x)(\delta x,\delta x)y_r(x,d-d^*)\rangle_0 + \\
& 2r^2\langle (Q_0(x)\delta x W^TW - [Q(x)\delta x, W^TW])y_r(x,d^*),
\end{aligned}$$

$$\tilde{N}_r(x)^{-1}\langle (Q_0(x)\delta x W^T W - [Q(x)\delta x, W^T W]) y_r(x, d-d^*)\rangle_0 +$$
$$r^2\langle (Q_0(x)\delta x W^T W - [Q(x)\delta x, W^T W]) y_r(x, d-d^*),$$
$$\tilde{N}_r(x)^{-1}\langle (Q_0(x)\delta x W^T W - [Q(x)\delta x, W^T W]) y_r(x, d-d^*)\rangle_0.$$

Recalling that $\|r^2\tilde{N}_r(x)^{-1}\|$ is bounded as $r \downarrow 0$, we obtain

$$\|\frac{1}{r^2}\nabla^2 J_r(x,d) - \frac{1}{r^2}\nabla^2 J_r(x,d^*)\|_{C^0(K)} \quad \leq \quad \text{const}\frac{\|d-d^*\|_0}{r} + \text{const}\frac{\|d-d^*\|_0^2}{r^2}.$$

This gives the desired result.

□

With this result, we can prove the following theorem.

Theorem 21 *Assume that the hypotheses of Theorem 19 hold. Then there exists $r_0 > 0$ and $\delta > 0$ such that for each*

$$d \in B_\delta(d^*),$$

$J_{r_0}(\cdot, d)$ has a unique stationary point, which corresponds to a minimum, in K.

Proof: Recall that, by assumption, $J_0(\cdot, d^*)$ has a unique stationary point x^* in the set K; moreover, $x^* \in N \subset K$, J_0 is strongly convex on N, and the gradient of $J_0(\cdot, d^*)$ is bounded away from zero on $K \setminus N$. Clearly, there exists $\gamma > 0$ sufficiently small that if

$$\|\frac{1}{r^2}J_r(\cdot, d) - J_0(\cdot, d^*)\|_{C^2(K)} \leq \gamma,$$

then $J_r(\cdot, d)/r^2$ will have a minimizer at a unique stationary point lying in N, and the gradient of $J_r(\cdot, d)/r^2$ will be bounded away from zero on $K \setminus N$. But our preceding theorems show that for $r_0 > 0$ sufficiently small,

$$\|\frac{1}{r^2}J_{r_0}(\cdot, d^*) - J_0(\cdot, d^*)\|_{C^2(K)} \leq \frac{\gamma}{2},$$

and for $\delta > 0$ sufficiently small (relative to r_0),

$$\|\frac{1}{r^2}J_{r_0}(\cdot, d) - \frac{1}{r^2}J_{r_0}(\cdot, d^*)\|_{C^2(K)} \leq \frac{\gamma}{2}.$$

This completes the proof.

□

It is clear that, at the expense of decreasing the level δ of allowable noise, we can conclude that there exists an interval $[r_0, r_1]$ such that for all $r \in [r_0, r_1]$ and all $d \in B_\delta(d^*)$, $J_r(\cdot, d)/r^2$ has a unique stationary point in K.

6.4 Existence of a parameterized path of minimizers

We now turn to our final result, namely, that there exists a path of minimizers of J_r, parameterized by r, which exists from arbitrarily small r and converges to a minimizer of J_∞ as $r \uparrow \infty$. The existence of such a path, indeed, the existence of a global minimizer of J_∞ for noisy data, is problematical unless we restrict the allowable data perturbations. Recall that, in the motivating examples, J_∞ is not differentiable for general data. Of course, when the data is noise–free, that is, when $d^* = A(x^*)y^*$ for some $y^* \in \mathcal{N}(W)$, then x^* is necessarily a global minimizer of $J_\infty(\cdot, d^*)$ (since it yields zero residual). However, if d represents data perturbed from d^* by arbitrary noise, then we cannot (because of the nonsmoothness of J_∞) use the usual implicit function theorem argument to conclude that there is a minimizer of $J_\infty(\cdot, d)$ near x^*.

We address the above difficulties by restricting the data to have some minimal degree of regularity. We will require that the data be smooth enough to allow the application of the implicit function theorem to J_∞ and to prove the existence of a continuous path of minimizers of J_r. In order to be able to prove that the desired path exists, we will need the following version of the implicit function theorem.

Theorem 22 *Let X be a Hilbert space, U an open subset of X, D a normed linear space, and I an interval of real numbers. Suppose that $G : U \times D \times I \to X$ has the following properties:*

1. *G is twice differentiable with respect to $x \in X$ and $d \in D$;*
2. *there exist $x^* \in U$ and $d^* \in D$ such that*
$$G(x^*, d^*, r) = 0 \text{ for all } r \in I;$$
3. *$D_x G(x^*, d^*, r)$ is invertible for all $r \in I$ and there exists $\alpha > 0$ such that*
$$\|D_x G(x^*, d^*, r)^{-1}\| \leq \alpha \text{ for all } r \in I;$$
4. *there exist $\beta > 0$ and $\gamma > 0$ and neighborhoods N of x^* and M of d^* such that*
$$\|D_{xx} G(x, d^*, r)(\delta x, \delta x)\| \leq \beta \|\delta x\|_X^2 \text{ for all } x \in N \text{ and } r \in I,$$
$$\|D_{xd} G(x, d, r)(\delta x, \delta d)\| \leq \gamma \|dx\|_X \|\delta d\|_D^2 \text{ for all } x \in N, d \in M, r \in I;$$
5. *G and $D_x G$ are continuous functions of r.*

Then there exist $\epsilon > 0$ and $\delta > 0$ and a continuous function $x : I \times B_\delta(d^*) \to B_\epsilon(x^*)$ such that
$$G(x_r(d), d, r) = 0 \text{ for all } d \in B_\delta(d^*) \text{ and } r \in I,$$
and
$$G(x, d, r) \neq 0 \text{ whenever } x \in B_\epsilon(x^*) \text{ and } x \neq x_r(d)$$
(where we write $x_r(d) = x(r, d)$).

Proof: Define $H_{d,r} : B_\epsilon(x^*) \to B_\epsilon(x^*)$ by

$$H_{d,r}(x) = x - D_x G(x^*, d^*, r)^{-1} G(x, d, r).$$

Using the uniform bounds α,β, and γ, it is straightforward to show that there exist δ and ϵ such that $H_{d,r}$ is well–defined for all $d \in B_\delta(d^*)$ and all $r \in I$, and that $H_{d,r}$ is a contractive mapping with a constant independent of $r \in I$. The result follows.

□

We now give the main theorem of this section.

Theorem 23 *Assume that $x^* \in U$, $y^* \in Y_{m+1}$, and $d^* \in D_m$ satisfy $A(x^*)y^* = d^*$, $Wy^* = 0$, where $m = max\{n, 3\}$. Assume further that for each x in a neighborhood of x^*, there exists $u_\infty(x, d^*) \in Y_4$ such that*

$$A^*(x)A(x)y_\infty(x, d^*) - A^*(x)d^* = W^*Wu_\infty(x, d^*).$$

Finally, assume that the mapping $x \to u_\infty(x, d^*)$ is continuous. Then for any $r_0 > 0$, there exist $\epsilon > 0$ and $\delta > 0$ with the following property: if $d \in D_m$ and $\|d - d^*\|_2 < \delta$, then there exists a path $x_r(d)$, continuously parameterized by $r \in [r_0, \infty)$, such that $x_r(d)$ is a stationary point of $\nabla J_r(\cdot, d)$ and $x_r(d) \in B_\epsilon(x^*)$ for all $r \in [r_0, \infty)$.

Proof: We apply the previous theorem to the function $G(x, d, r) = \nabla J_r(x, d)$. We need merely to verify that the hypotheses of the above theorem are satisfied, where we take D to be the space D_n under the D_2 inner product and $I = [r_0, \infty)$.

Recall that the mapping $x \to A(x)y$ is of class C^1 provided $y \in Y_n$, and that the mapping $x \to y_r(x, d)$ is of class C^1 provided $d \in D_{n-1}$. So under the assumptions of this theorem we can compute as follows (suppressing the dependence of the various quantities on x and δx):

$$\begin{aligned}
J_r(x,d) &= \frac{1}{2}\langle Ay_r - d, Ay_r - d\rangle_0 + \frac{r^2}{2}\langle Wy_r, Wy_r\rangle_0 \\
D_xJ_r(x,d)\delta x &= \langle DAy_r + ADy_r, Ay_r - d\rangle_0 + r^2\langle WDy_r, Wy_r\rangle_0 \\
&= \langle AQy_r, Ay_r - d\rangle_0 + \langle Dy_r, A^*A + r^2W^*W - A^*d\rangle_1 \\
&= \langle Qy_r, A^T(Ay_r - d)\rangle_{1,-1} \\
D^2_{xx}J_r(x,d)(\delta x, \delta x) &= \langle DQy_r + QDy_r, A^T(Ay_r - d)\rangle_{1,-1} + \\
&\quad \langle Qy_r, Q^TA^T(Ay_r - d) + A^T(AQy_r + ADy_r)\rangle_{1,-1}.
\end{aligned}$$

Note that

$$\tilde{N}_r y_r = A^T d$$

yields

$$\tilde{N}_r D_x y_r = Q^T A^T d - Q^T A^T A y_r - A^T A Q y_r,$$

or

$$D_x y_r = \tilde{N}_r^{-1}(Q^T A^T (d - Ay_r) - A^T A Q y_r).$$

Recall from Corollary 1 that $\|\tilde{N}_r^{-1}\|$ is uniformly bounded on I. Upon simplifying, we obtain

$$\begin{aligned}
&D^2_{xx}J_r(x,d)(\delta x,\delta x)\\
=\;&\langle D_xQy_r+Q\tilde{N}_r^{-1}(Q^TA^T(d-Ay_r)-A^TAQy_r),A^T(Ay_r-d)\rangle_{1,-1}+\\
&\langle Qy_r,(I-A^TA\tilde{N}_r^{-1}(Q^TA^T(Ay_r-d)+A^TAQy_r)\rangle_{1,-1}.
\end{aligned}$$

Note that our assumptions allow $D^2_{xx}J_r$ to be differentiated again; in particular, note that $d\in D_n$ implies that $y_r\in Y_{n+1}$ and hence that $Qy_r\in Y_n$. Without actually computing the next derivative, it is easy to see that it will be bounded (uniformly with respect to r going to ∞ and x in a compact set) if $\|y_r-y_\infty\|_4=O(r^{-2})$ (the use of the stronger norm is necessary because of the order of the operators involved). This rate of convergence of y_r to y_∞ is a consequence of the fact that u_∞ belongs to Y_4. Similarly, the partial derivative of $D^2_{xx}J_r$ with respect to d is easily seen to be uniformly bounded with respect to the D_2 norm on d. The continuity with respect to r is also immediate.

It remains only to verify the uniform upper bound on the norm of $\nabla^2_{xx}J_r(x^*,d^*)^{-1}$. For this, we return to the expression for $D_{xx}J_r$ which was derived in Theorem 14. Since $y_r(x^*,d^*)=y^*$ and $Wy^*=0$, we have

$$\begin{aligned}
D^2_{xx}J_r(x^*,d^*)(\delta x,\delta x) &= r^2\langle[W,Q]y^*,[W,Q]y^*\rangle_0-r^4\langle[Q,W^TW],\tilde{N}_r^{-1}[Q,W^TW]\rangle_0\\
&= r^2\langle[W,Q]y^*,(I-r^2W\tilde{N}_r^{-1}W^*)[W,Q]y^*\rangle_0.
\end{aligned}$$

By differentiating this expression with respect to r, it is straightforward to show that $D^2_{xx}J_r(x^*,d^*)(\delta x,\delta x)$ is positive and increasing as a function of r. It follows from the finite–dimensionality of X that $\nabla^2J_r(x^*,d^*)$ is positive definite with a uniform lower bound. This completes the proof.

□

We now examine the perturbation theory for J_∞, and show that the limit of the path in the previous theorem is indeed a minimizer of J_∞.

Theorem 24 *Assume that $d\in D_n$. Then $J_\infty(\cdot,d)$ is of class C^2.*

Proof: By Lemma 7, the mapping $x\longrightarrow y_\infty(x,d)$ is of class C^1, and $y_\infty(x,d)\in Y_{n+1}$ for all $x\in U$. It follows from Leibniz's rule that the map $x\longrightarrow A(x)y_\infty(x,d)$ is continuously differentiable, and hence so is $J_\infty(\cdot,d)$. We have

$$\begin{aligned}
DJ_\infty(x,d)\delta x &= \langle DA(x)\delta x y_\infty(x,d)+A(x)Dy_\infty(x,d)\delta x,A(x)y_\infty(x,d)-d\rangle_0\\
&= \langle Q(x)\delta x y_\infty(x,d)+Dy_\infty(x,d)\delta x,A^*(x)A(x)y_\infty(x,d)-A^*(x)d\rangle_1.
\end{aligned}$$

By the optimality conditions for the inner problem (Theorem 1),

$$A^*(x)A(x)y_\infty(x,d)-A^*(x)d\in\mathcal{N}(W)^\perp$$

and

$$Dy_\infty(x,d)\delta x\in\mathcal{N}(W).$$

Hence,

$$\begin{aligned} DJ_\infty(x,d)\delta x &= \langle Q(x)\delta x y_\infty(x,d), A^*(x)A(x)y_\infty(x,d) - A^*(x)d\rangle_1 \\ &= \langle A(x)Q(x)\delta x y_\infty(x,d), A(x)y_\infty(x,d) - d\rangle_0. \end{aligned}$$

It follows from this formula and the above reasoning that $J_\infty(\cdot,d)$ can be differentiated a second time and, in fact, that it is of class C^2.

□

Theorem 25 *Assume that the hypotheses of Theorem 23 hold. Assume further that $\nabla^2 J_\infty(x^*,d^*)$ is positive definite. Then there exist $\delta > 0$ and $\epsilon > 0$ such that, for each $d \in D_m$ satisfying $\|d - d^*\|_2 < \delta$, the path of Theorem 23 converges to the unique minimizer of $J_\infty(\cdot,d)$ in $B_\epsilon(x^*)$ as $r \uparrow \infty$.*

Proof: From the results of Theorem 23 and the implicit function theorem applied to ∇J_∞, we can conclude that there exist $\delta > 0$ and $\epsilon > 0$ such that for any $d \in D_m$ satisfying $\|d - d^*\|_2 < \delta$:

1. There exists a path $x_r(d)$ in $B_\epsilon(x^*)$ such that $x_r(d)$ is the unique stationary point of $J_r(\cdot,d)$ in $B_\epsilon(x^*)$;
2. There exists a point $x_\infty(d) \in B_\epsilon(x^*)$ such that $x_\infty(d)$ is the unique stationary point of $J_\infty(\cdot,d)$ in $B_\epsilon(x^*)$.

Moreover, by possibly decreasing ϵ and δ, we can ensure that these points are actually strong local minimizers, and that all of the objective functions are convex on $B_\epsilon(x^*)$ (this is possible because of the uniform bounds derived in Theorem 23). It therefore follows that $x_r(d) \to x_\infty(d)$ as $r \uparrow \infty$.

□

7 The examples revisited

In this section, we will show that the results of Sections 4 and 6 are applicable to the examples we discussed in Section 2. We also present additional numerical experiments which illustrate the theory.

7.1 The plane wave detection problem

As mentioned in Section 2, the plane wave detection problem has been investigated by Symes in [12]. In this section, we will summarize his results, and compare them with the predictions of our theory.

7.2 Existing theory

Define

$$H_b^1 \quad = \quad \left\{ u \in H^1(\Sigma) : u(x,0) \equiv u(x,T), \int_0^T \int_{-X}^X u(x,t)\mathrm{dxdt} \equiv 0 \right\},$$

where $\Sigma = [-X, X] \times [0, T]$, and

$$\phi : \mathcal{R} \to \mathcal{L}(H_b^1, L^2(\Sigma))$$

by

$$\phi(s)u(x,t) = \frac{\partial u}{\partial t}(x, t + sx).$$

The function ϕ corresponds to the function we have called A in the previous sections, and the partial differential operator $\partial/\partial x$ will play the role of the operator W. The DSO objective function J_r is defined by

$$J_r(s) = \min \left\{ \frac{1}{2} \|\phi(s)u - z\|^2_{L^2(\Sigma)} + \frac{r^2}{2} \|\frac{\partial u}{\partial x}\|^2_{L^2(\Sigma)} : u \in H_b^1 \right\},$$

where $z \in L^2(\Sigma)$ comprises the data.

The main results concerning the behavior of J_r for this problem are given by the following two theorems. The symbol $R_{s,t}$ will represent the ratio of the $H^s(\Sigma)$ norm of u^* to the $H^t(\Sigma)$ norm of u^*. Note that if $s < t$ and $R_{s,t}$ is very small, then the function u^* is *oscillatory*.

Theorem 26 ([12],Theorem 2.9) *Suppose that the data z is consistent, that is, $z = \phi(s^*)u^*$ for some $s^* \in [-1, 1]$, $u^* \in H_b^1$, where $\partial u^*/\partial x = 0$. Then:*

1. *If, in addition, $u^* \in H^{3/2}(\Sigma)$, then*

$$\left(\|z\|_{L^2(\Sigma)} - const \left\{ \frac{R_{0,1}^2}{|s - s^*|} + \frac{R_{3/2,1}^2}{r} \right\}^{1/2} \|u^*\|_{H^1(\Sigma)} \right)^2$$

$$\leq \quad J_r(s) \leq \left(\|z\|_{L^2(\Sigma)} + const \left\{ \frac{R_{0,1}^2}{|s - s^*|} + \frac{R_{3/2,1}^2}{r} \right\}^{1/2} \|u^*\|_{H^1(\Sigma)} \right)^2.$$

2.

$$\left|\frac{dJ_r}{ds}(s)\right| \geq r^2|s-s^*|\|u^*\|^2_{H^1(\Sigma)}\left\{\frac{*}{\pi} - constR^2_{0,1} - const\sqrt{r}R_{1/2,1}\right\}.$$

3. If $\sqrt{r}R_{1/2,1}$ and $R_{0,1}$ are sufficiently small, then J_r is quasiconvex over $[-1,1]$.

The first part the Theorem 26 shows that, as $r \uparrow \infty$, the DSO objective function $J_r(s)$ flattens out for $|s-s^*|$ sufficiently large compared to the frequency content of the signal u^*. This, among other things, means that the shape of the OLS objective function, discussed in connection with the numerical examples in Section 2, is qualitatively correct. Parts 2 and 3 of this theorem give a result similar to the conclusion of our Corollary 4. However, Theorem 26 requires that the source waveform u^* be sufficiently oscillatory in order for this conclusion to be valid. We shall see below that such a restriction is not necessary.

Theorem 27 ([12], Theorem 2.10) *Assume that $r > 0$ and that $u^* \in H^1_b$ satisfying $\partial u^*/\partial x = 0$ is sufficiently oscillatory relative to r, as in Theorem 26. Let $z \in L^2(\Sigma)$ and define e by*

$$e = \frac{\|z - z^*\|_{L^2(\Sigma)}}{\|z^*\|_{L^2(\Sigma)}},$$

where $z^* = \phi(s^*)u^*$. If s is a critical point of J_r, then

$$|s-s^*| \leq const\frac{e}{r}.$$

Theorem 27 should be compared to Theorems 23 and 25. Theorem 27 is stronger in that it gives a Lipschitz bound on the perturbation of the minimizer; it also does not require the use of a stronger norm to measure the data perturbation. On the other hand, Theorem 27 only applies to data generated by a source which is sufficiently oscillatory compared to r, and gives no information about what happens when r increases without bound.

7.3 New results

We will now formulate the plane wave detection problem in such a way that we can apply the results of our theory, and compare those results with those given above. We need to make one change in the definitions given in the previous section; in order that the forward map be injective, we define

$$H^1_b(\Sigma) = \left\{u \in H^1(\Sigma) : u(x,0) \equiv u(x,T), \int_0^T u(x,t)\mathrm{dt} \equiv 0\right\}.$$

This space will correspond to the space Y_1 in the previous sections; the space

$$H^k_b(\Sigma) = \left\{u \in H^k(\Sigma) : \int_0^T u(x,t)\mathrm{dt} \equiv 0, \frac{\partial^i u}{\partial t^i}(x,0) \equiv \frac{\partial^i u}{\partial t^i}(x,T), i = 0,\ldots,k-1\right\}$$

will correspond to the space Y_k. The data space D_0 will be represented by

$$\left\{ z \in L^2(\Sigma) : \int_0^T z(x,t)\mathrm{dt} \equiv 0 \right\},$$

with the other spaces D_k defined analogously. We will still denote the forward map (defined by the same formula) by ϕ. Note that the range of $\phi(s)$ is actually constant with respect to s, and is in fact dense in the space D_0.

One of the most important hypotheses that we made in Section 5 is that the normal operator satisfies an elliptic regularity condition. This is the reason that we must include the time derivative in the definition of the forward map. The result of doing so is that the normal equation for solving the inner minimization problem (that is, for computing $u_r[s]$) is formally equivalent to the following elliptic boundary value problem:

$$\frac{\partial^2 u}{\partial t^2} + r^2 \frac{\partial^2 u}{\partial x^2} = \frac{\partial z}{\partial t}(x, t + sx)$$

$$u(x,0) \equiv u(x,T)$$

$$\frac{\partial u}{\partial t}(x,0) \equiv \frac{\partial u}{\partial t}(x,T)$$

$$\frac{\partial u}{\partial x}(-X,t) \equiv \frac{\partial u}{\partial x}(X,t) \equiv 0$$

$$\int_0^T u(x,t)\ \mathrm{dt} = 0$$

Thus the necessary elliptic regularity follows from standard results in the theory of partial differential equations.

We must compute the function J_0 for this problem, and show that the various assumptions made in the preceding sections are realistic for this example. Let z^* represent exact data, that is, suppose there exist $s^* \in (-1,1)$ and $u^* \in H_b^1(\Sigma)$ such that

$$\frac{\partial u^*}{\partial x} = 0$$

and $z^* = \phi(s^*)u^*$. Then it is easy to verify that

$$u_0[s](x,t) = u^*(t + (s^* - s)x)$$

satisfies

$$\phi(s)u_0[s] = z^*.$$

Since

$$\frac{\partial u_0[s]}{\partial x}(x,t) = (s^* - s)u^{*\prime}(t + (s^* - s)x),$$

we obtain

$$J_0(s) = X\|u^{*\prime}\|_{L^2[0,T])}(s - s^*)^2.$$

Thus the dual objective function in this case is quadratic.

In order to verify the C^2 convergence of $J_r(\cdot, z^*)$ to $J_0(\cdot, z^*)$ we must consider the conditions $W^*Wu_0[s] \in \mathcal{R}(\phi^*(s)\phi(s))$. We can simplify the calculations by considering the equivalent condition $W^TWu_0[s] \in \mathcal{R}(\phi^T(s)\phi(s))$. We have

$$W^TWu_0[s](x,t) = -u^{*\prime\prime}(t+(s^*-s)x)(s-s^*)^2$$

and

$$\phi^T(s)\phi(s)u(x,t) = -\frac{\partial^2 u}{\partial t^2}(x,t);$$

it is easy to check that

$$v_0[s](x,t) = (s-s^*)^2u^*(t+(s^*-s)x)$$

satisfies

$$W^TWu_0[s] = \phi^T(s)\phi(s)v_0[s].$$

We also see that the map $s \longrightarrow v_0[s]$ is continuous.

We next must check that the condition

$$\left\{Q_0(s)\delta s W^TW - [Q(s)\delta s, W^TW]\right\} u_0[s] \in \mathcal{R}(\phi^T(s)\phi(s))$$

is satisfied, which requires that we compute the operators Q and Q_0. If u is smooth, then

$$\begin{aligned} D_s\phi(s)\delta s u(x,t) &= \frac{\partial^2 u}{\partial t^2}(x,t+sx)x\delta s \\ &= \phi(s)Q(s)\delta s u(x,t), \end{aligned}$$

where

$$Q(s)\delta s u(x,t) = \frac{\partial u}{\partial t}(x,t)x\delta s.$$

The operator Q is in fact antisymmetric, so that $Q_0 = 0$. It is also easy to compute

$$[Q(s)\delta s, W^TW]u = -2\delta s\frac{\partial^2 u}{\partial x \partial t}.$$

It follows that

$$w_0[s](x,t) = -2\delta s(s^*-s)u^*(t+(s^*-s)x)$$

satisfies

$$-[Q(s)\delta s, W^TW]u_0 = \phi^T(s)\phi(s)w_0.$$

The map $s \longrightarrow w_0[s]$ is continuous.

We have thus verified that $J_r(\cdot, z^*)/r^2$ converges to $J_0(\cdot, z^*)$ in $C^2[-1,1]$. It follows that for r sufficiently small, $J_r(\cdot, z^*)$ is convex and is minimized at s^*. In contrast to Theorem 26, the nature of the source does not affect this conclusion.

We next would like to apply Theorem 23 to the plane wave detection problem, which would represent a partial improvement over Theorem 27. In order to verify the hypotheses of Theorem

23, we must compute the solution $u_\infty[s]$ to the inner "infinity" problem. It may be verified by the optimality conditions given in Theorem 1 that $u_\infty[s]$ is given by

$$u_\infty[s](x,t) = \frac{1}{2X(s^*-s)}\left(U^*(t+(s^*-s)X) - U^*(t-(s^*-s)X)\right),$$

where U^* is the primitive of u^* having mean zero. We have

$$\begin{aligned} & \left(\phi^T(s)\phi(s)u_\infty - \phi^T(s)z^*\right)(x,t) \\ = & -\frac{1}{2X(s^*-s)}\left(u^{*\prime}(t+(s^*-s)X) - u^{*\prime}(t-(s^*-s)X)\right) + u^{*\prime\prime}(t+(s^*-s)x). \end{aligned}$$

We conclude that

$$\phi^T(s)\phi(s)u_\infty - \phi^T(s)z^* = W^TWv_\infty[s],$$

where

$$\begin{aligned} v_\infty[s](x,t) &= \frac{1}{4X(s^*-s)}x^2\left(u^{*\prime}(t+(s^*-s)X) - u^{*\prime}(t-(s^*-s)X)\right) - \\ & \frac{1}{(s^*-s)^2}u^*(t+(s^*-s)x)). \end{aligned}$$

It is clear that $v_\infty[s]$ belongs to $H_b^4(\Sigma)$ provided $u^* \in H_b^5(\Sigma)$. So under these conditions on the source u^*, we can apply Theorems 23 and 25. As discussed above, this gives us a perturbation result that is in some ways stronger than Theorem 27. In particular, it shows that, except for the problem of choosing a value of r_0 as in Theorem 21, there is a deterministic way of finding the minimizer of J_∞ which is closest to x^* provided the noise in the data is small enough in the H^2 norm. (Note that we cannot conclude that the the limit of the path of minimizers of $J_r(\cdot,z)$ is the global minimizer of $J_\infty(\cdot,z)$.)

7.4 The layer–over–half–space problem

We now turn to the layer–over–half–space problem with the goal of verifying the various conditions which allow the application of the results in Section 6. We will then illustrate these results with a variety of numerical experiments.

7.5 Verification of hypotheses

We proceed to compute the various quantities as we did for the plane wave detection problem. As in Section 2, we define $\Omega = [0,H]\times[0,T]$. Corresponding to the space Y_1 in Sections 5 and 6 will be the space

$$H_b^1(\Omega) = \left\{ f \in H^1(\Omega) : f(h,0) \equiv f(h,T), \int_0^T f(h,t)\mathrm{d}t \equiv 0 \right\},$$

and the ladder of spaces $\left\{H_b^k(\Omega)\right\}_{k=-\infty}^{\infty}$ will be the usual Sobolev spaces with the appropriate boundary conditions. The data space D_0 will be given by

$$D_0 = \left\{ f \in L^2(\Omega) : \int_0^T f(h,t)\mathrm{d}t \equiv 0 \right\},$$

and the ladder $\{D_k\}_{k=-\infty}^{\infty}$ will also be the appropriate Sobolev spaces.

We define $U = \{(c,d) \in \mathcal{R}^2 : c > 0, d > 0\}$, and the forward map $S : U \rightarrow \mathcal{L}(H_b^1, D_0)$ is defined by

$$S(c,d)f(h,t) = \frac{\partial f}{\partial t}\left(h, t - \frac{2\sqrt{d^2+h^2}}{c}\right).$$

As in the plane wave detection problem, the inclusion of the time derivative in the forward map leads to an elliptic normal operator.

Now consider exact data (the source waveform is independent of h):

$$D^*(h,t) = f^{*\prime}\left(t - \frac{2\sqrt{{d^*}^2+h^2}}{c^*}\right).$$

Let (c,d) be fixed. The function $f_0[c,d]$ given by

$$f_0[c,d](h,t) = f^*\left(t + \frac{2\sqrt{d^2+h^2}}{c} - \frac{2\sqrt{{d^*}^2+h^2}}{c^*}\right)$$

is easily seen to satisfy $S(c,d)f_0[c,d] = D$. Therefore,

$$\begin{aligned} J_0(c,d) &= \|\frac{\partial f_0[c,d]}{\partial h}\|^2_{L^2(\Omega)} \\ &= \|f^{*\prime}\|^2_{L^2[0,T]} \int_0^H \left(\frac{2h}{c\sqrt{d^2+h^2}} - \frac{2h}{c^*\sqrt{{d^*}^2+h^2}}\right)^2 \text{dh}. \end{aligned}$$

We have

$$S^T(c,d)S(c,d)f(h,t) = -\frac{\partial^2 f}{\partial t^2}(h,t)$$

and

$$\left(\frac{\partial}{\partial h}\right)^T \left(\frac{\partial}{\partial h}\right) f = -\frac{\partial^2 f}{\partial h^2}.$$

We thus have

$$\begin{aligned} &\left(\frac{\partial}{\partial h}\right)^T \left(\frac{\partial}{\partial h}\right) f_0[c,d](h,t) \\ = \;& f^{*\prime\prime}\left(t - \frac{2\sqrt{{d^*}^2+h^2}}{c^*} + \frac{2\sqrt{d^2+h^2}}{c}\right) \cdot \\ & \left(-\frac{2h}{\sqrt{{d^*}^2+h^2}c^*} + \frac{2h}{\sqrt{d^2+h^2}c}\right)^2 + \\ & f^{*\prime}\left(t - \frac{2\sqrt{{d^*}^2+h^2}}{c^*} + \frac{2\sqrt{d^2+h^2}}{c}\right) \cdot \end{aligned}$$

$$\left(\frac{2\,h^2}{\left({d^*}^2+h^2\right)^{3/2} c^*} - \frac{2}{\sqrt{{d^*}^2+h^2}c^*} - \frac{2\,h^2}{\left(d^2+h^2\right)^{3/2} c} + \frac{2}{\sqrt{d^2+h^2}c} \right)$$

It can be seen that this function is in the range of $S^T S$ provided only that the primitive of f^* is periodic with mean 0. Under these restrictions on the source, we see that $J_r(\cdot, D^*)/r^2$ converges to $J_0(\cdot, D^*)$ in the sense of C^1.

In order to verify the convergence of the second derivative, we must compute the operator Q and check the condition in Theorem 19. If f is smooth, then we have

$$\begin{aligned} D_{(c,d)}S(c,d)f(h,t) &= \frac{\partial^2 f}{\partial t^2}\left(h, t - \frac{2\sqrt{d^2+h^2}}{c}\right) \cdot \\ &\quad \left(\frac{2\sqrt{d^2+h^2}}{c^2}\delta c - \frac{2d}{c\sqrt{d^2+h^2}}\delta d\right) \\ &= S(c,d)Q(c,d)(\delta c, \delta d)f(h,t), \end{aligned}$$

where

$$\begin{aligned} & Q(c,d)(\delta c, \delta d)f(h,t) \\ = \;& \frac{\partial f}{\partial t}(h,t)\left(\frac{2\sqrt{d^2+h^2}}{c^2}\delta c - \frac{2d}{c\sqrt{d^2+h^2}}\delta d\right). \end{aligned}$$

It may be verified that Q is antisymmetric, so that $Q_0 = 0$ for this problem as well.

We now compute

$$\begin{aligned} & [Q(c,d)(\delta c, \delta d), W^T W]f(h,t) \\ = \;& 2\frac{\partial^2 f}{\partial h \partial t}(h,t)\left(\frac{2h}{c^2\sqrt{d^2+h^2}}\delta c + \frac{2dh}{c\left(d^2+h^2\right)^{3/2}}\delta d\right) + \\ & \frac{\partial f}{\partial t}(h,t)\left(\frac{2d^2}{c^2\left(d^2+h^2\right)^{3/2}}\delta c + \frac{2d\left(2h^2+d^2\right)}{c\left(d^2+h^2\right)^{5/2}}\delta d\right). \end{aligned}$$

Therefore,

$$\begin{aligned} & [Q(c,d)(\delta c, \delta d), W^T W]f_0[c,d](h,t) \\ = \;& 2{f^*}''\left(t + \frac{2\sqrt{d^2+h^2}}{c} - \frac{2\sqrt{{d^*}^2+h^2}}{c^*}\right) \cdot \\ & \left(\frac{2h}{c^2\sqrt{d^2+h^2}}\delta c + \frac{2dh}{c\left(d^2+h^2\right)^{3/2}}\delta d\right)\frac{2h}{c\sqrt{d^2+h^2}} + \\ & {f^*}'\left(t + \frac{2\sqrt{d^2+h^2}}{c} - \frac{2\sqrt{{d^*}^2+h^2}}{c^*}\right)\left(\frac{2d^2}{c^2\left(d^2+h^2\right)^{3/2}}\delta c + \frac{2d\left(2h^2+d^2\right)}{c\left(d^2+h^2\right)^{5/2}}\delta d\right). \end{aligned}$$

Under the same conditions discussed above, we have that $[Q(c,d)(\delta c,\delta d),W^TW]f_0[c,d]$ belongs to $\mathcal{R}(S^T(c,d)S(c,d))$. We therefore have C^2 convergence of $J_r(\cdot,D^*)/r^2$ to $J_0(\cdot,D^*)$, as guaranteed by Theorem 19.

As in the plane wave detection problem, the dual objective function can be easily minimized using local optimization (although in this case, $J_0(\cdot,D^*)$ fails to fit exactly into any of the simple categories such as quadratic, convex, or quasiconvex). We computed the values of $J_0(c,d,D^*)$ (scaled by a constant) using the same geometry as in the earlier examples. The quadrature was done by a two–point Gaussian quadrature rule. The results are show in Figure 7.8, which should be compared to Figure 2.5. The similarity of $J_1(\cdot,D^*)$ to $J_0(\cdot,D^*)$ is clearly seen.

We now verify the hypothesis that allows us to apply Theorem 23. This requires that we compute $f_\infty[c,d]$, the solution to the inner "infinity" problem. It may be verified that

$$f_\infty[c,d](h,t) = \frac{1}{H}\int_0^H f^*\left(t + \frac{2\sqrt{d^2+h^2}}{c} - \frac{2\sqrt{{d^*}^2+h^2}}{c^*}\right)\,\mathrm{dh}$$

satisfies the optimality conditions given in Theorem 1. We then have

$$\begin{aligned}
&\left(S^T(c,d)S(c,d)f_\infty[c,d] - S^T(c,d)D^*\right)(h,t)\\
= \;& -\frac{1}{H}\int_0^H f^{*\prime}\left(t + \frac{2\sqrt{d^2+h^2}}{c} - \frac{2\sqrt{{d^*}^2+h^2}}{c^*}\right)\,\mathrm{dh} \; +\\
& f^{*\prime\prime}\left(t + \frac{2\sqrt{d^2+h^2}}{c} - \frac{2\sqrt{{d^*}^2+h^2}}{c^*}\right).
\end{aligned}$$

This function is the image under W^TW of its second h-antiderivative, which we shall denote by $g_\infty[c,d]$. It is clear that $g_\infty[c,d]$ belongs to $H_b^4(\Omega)$ provided the source f^* belongs to H^6 and satisfies the periodic boundary conditions. Under these conditions, we can apply Theorems 23 and 25. We conclude that if D represents the result of perturbing exact data D^* by H^3 noise, which is small in the H^2 sense, then there is a path of minimizers of $J_r(\cdot,D)$, parameterized by r, which converges to a minimizer of $J_\infty(\cdot,D)$. (The various regularity conditions — $f^*\in H_b^6(\Omega)$, the data perturbation belonging to H^3 and measured in the H^2 norm – seem rather puzzling, but in fact are just what is required to apply the form of the implicit function theorem given in Section 6. We require $f^*\in H_b^6(\Omega)$ so that $g_\infty[c,d]$ lies in $H_b^4(\Omega)$ and hence that $f_r[c,d]$ converges to $f_\infty[c,d]$ in the H^4 norm. This yields the necessary bound on the x-derivative of the Hessian of J_r in the case of exact data. The data perturbation must lie in H^3 so that the Hessian of J_r is continuously differentiable in x. Finally, the data perturbation is measured in the H^2 norm so that the Hessian of J_r will be uniformly Lipschitz in D.)

7.6 Numerical experiments

We now proceed to illustrate the results from Section 6 applied to the layer–over–half–space problem. In all of the graphs shown below, we have generated exact data, possibly added noise, as specified, and then sampled the value of J_r on a grid. We then display the objective function with contour plots and one–dimensional scans. The geometry of the seismic experiments which we are simulating is the same as we used in Section 2. There are 51 receivers located at half–offsets $0, 20, \ldots, 1000$; the source is a 50 Hz. Ricker wavelet. The time interval is 1024 milliseconds. The exact data is generated using $c = 5$ m/ms and $d = 500$ m.

7.6.1 Example 1

For the first example, we used exact data D^* from the experiment described above, and computed the functions $J_0(\cdot, D^*)$, $J_r(\cdot, D^*)/r^2$ for $r = 0.1$, $J_r(\cdot, D^*)$ for $r = 1, 10, 50, 100, 1000$, and $J_\infty(\cdot, D^*)$. These plots are displayed below, together with the seismic data D^*. These plots illustrate the convergence of $J_r(\cdot, D^*)$ to $J_\infty(\cdot, D^*)$ as $r \uparrow \infty$ and $J_r(\cdot, D^*)/r^2$ to $J_0(\cdot, D^*)$ as $r \downarrow 0$. With each contour plot, we include a one–dimensional scan of the objective function, in which d is held fixed at $d = 5$ and c is varied.

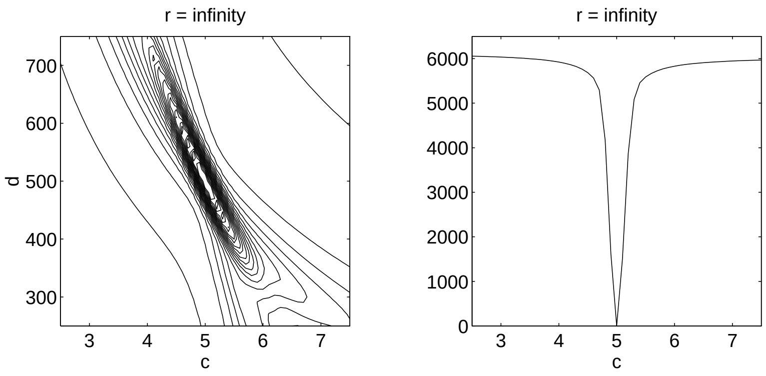

Figure 7.1: Example 1: $J_\infty(c, d)$

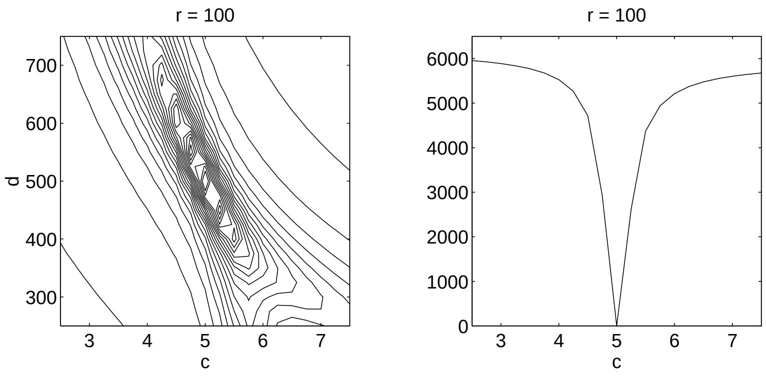

Figure 7.2: Example 1: $J_{100}(c, d)$

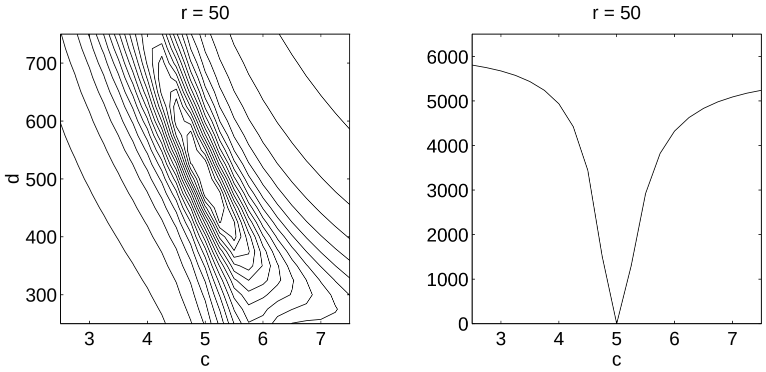

Figure 7.3: Example 1: $J_{50}(c,d)$

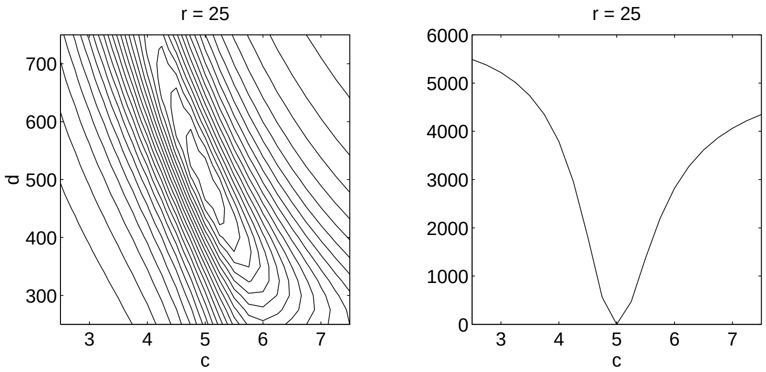

Figure 7.4: Example 1: $J_{25}(c,d)$

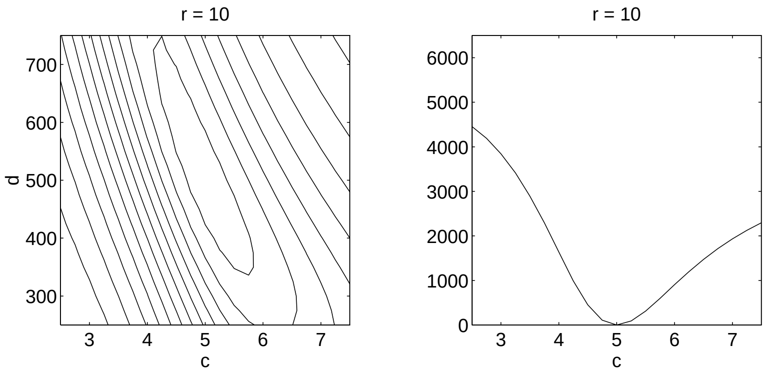

Figure 7.5: Example 1: $J_{10}(c,d)$

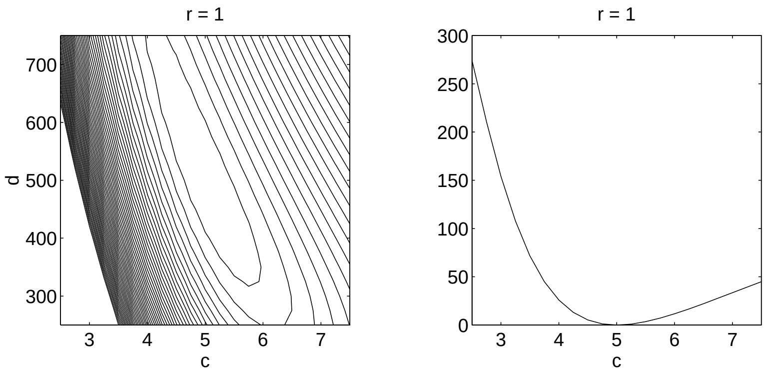

Figure 7.6: Example 1: $J_1(c,d)$

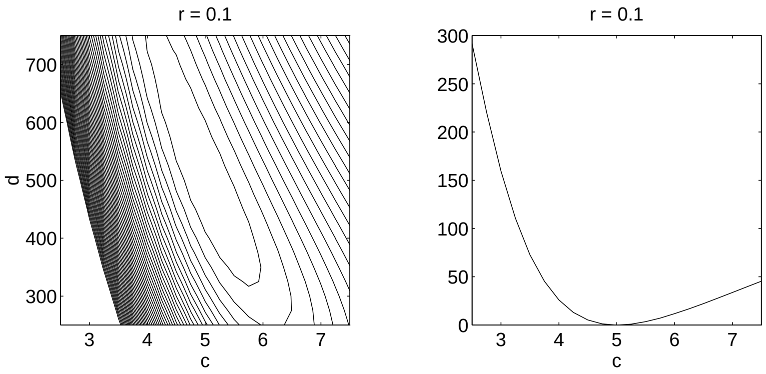

Figure 7.7: Example 1: $\frac{1}{0.1^2} J_{0.1}(c,d)$

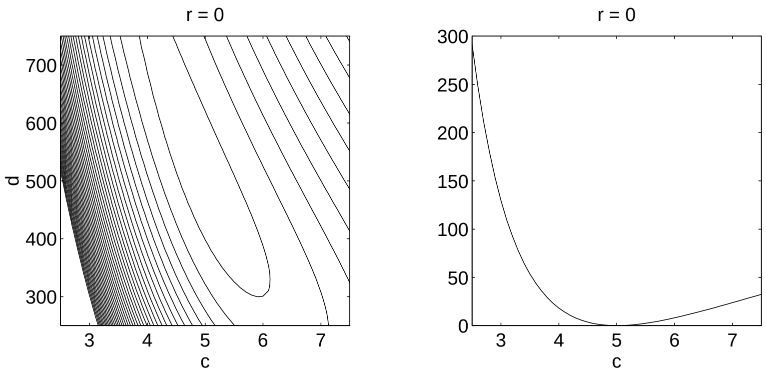

Figure 7.8: Example 1: $J_0(c,d)$ (scaled)

7.6.2 Example 2

In this example, we take the exact data D^* from the preceding example and add 25% filtered Gaussian noise. We display the same plots as in the previous examples.

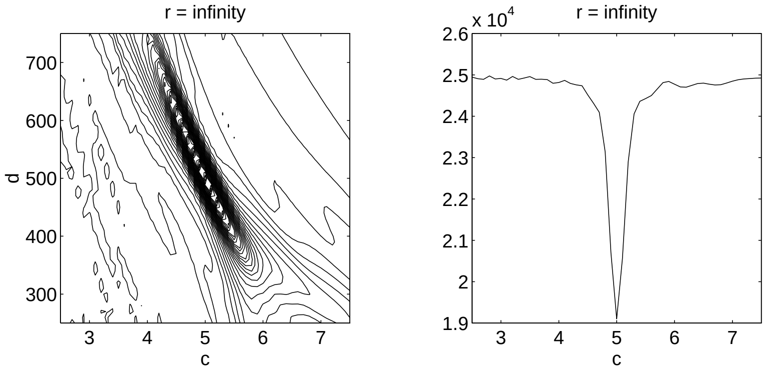

Figure 7.9: Example 2: $J_\infty(c, d)$

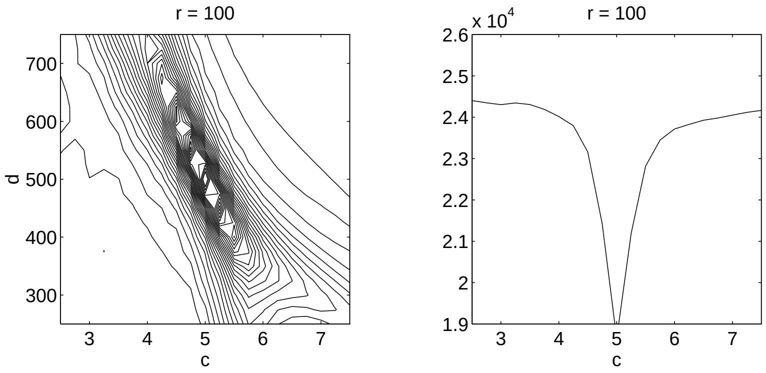

Figure 7.10: Example 2: $J_{100}(c, d)$

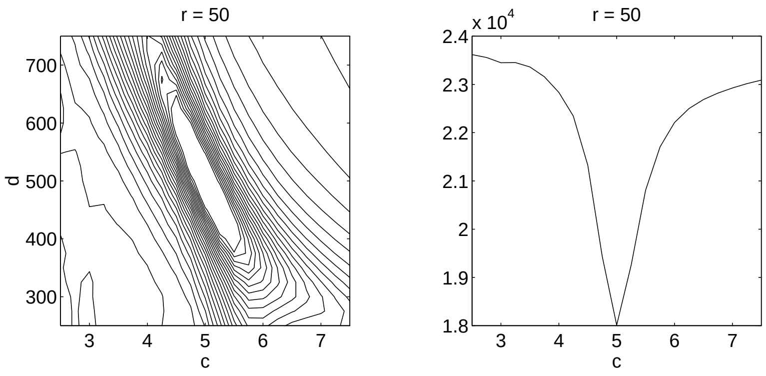

Figure 7.11: Example 2: $J_{50}(c, d)$

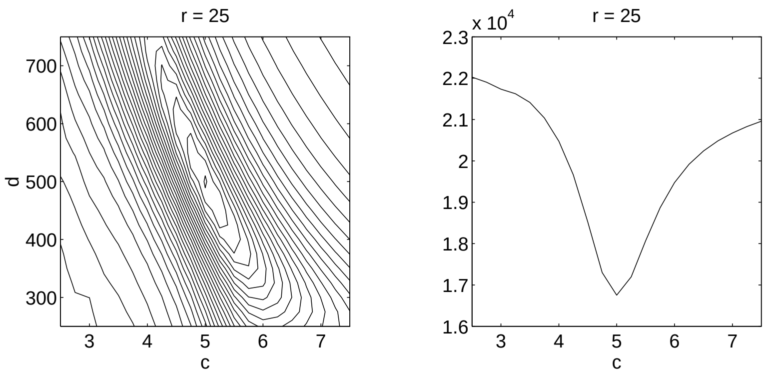

Figure 7.12: Example 2: $J_{25}(c, d)$

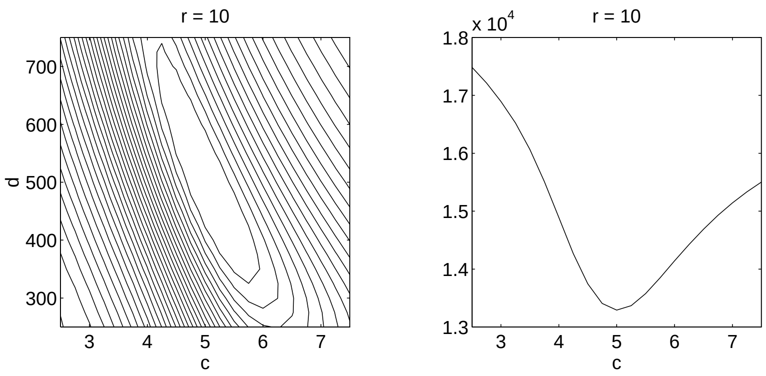

Figure 7.13: Example 2: $J_{10}(c, d)$

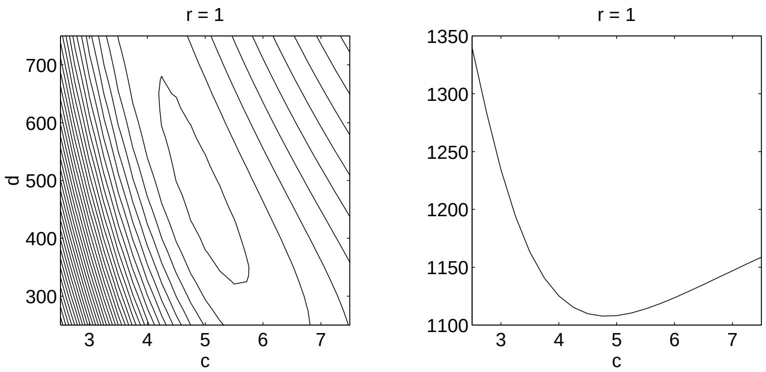

Figure 7.14: Example 2: $J_1(c, d)$

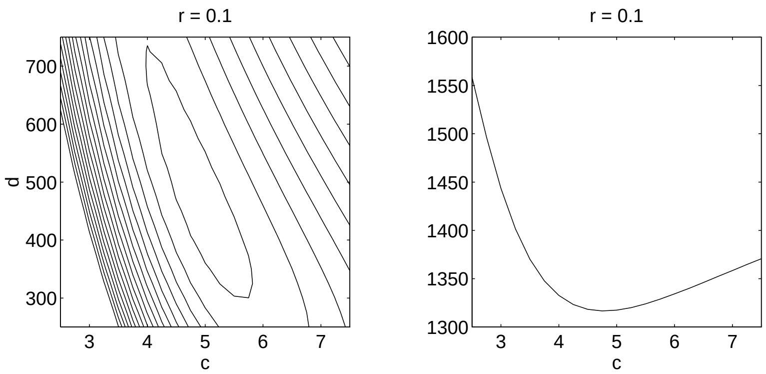

Figure 7.15: Example 2: $\frac{1}{0.1^2} J_{0.1}(c, d)$

In order to illustrate Theorems 23 and 25, we computed the minimizers of $J_\infty(\cdot, D)$ and $J_r(\cdot, D)$ for various values of r. The results are shown in Figure 7.16, where the curve represents the path $x_r(D)$, $\times$ marks $x_\infty(D)$, and $*$ marks the unperturbed solution.

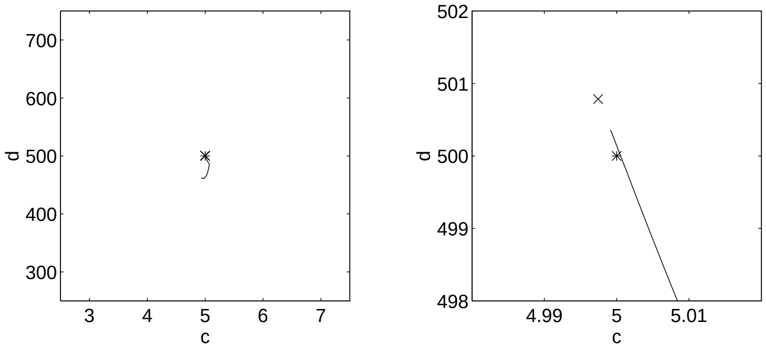

Figure 7.16: Example 2: Convergence of $x_r(D)$ to $x_\infty(D)$

7.6.3 Example 3

This example is exactly the same as Example 2, except that the noise level is 50%. The character of the family of objective functions is very similar to what is shown in Example 2, so we display only the convergence of $x_r(D)$ to $x_\infty(D)$.

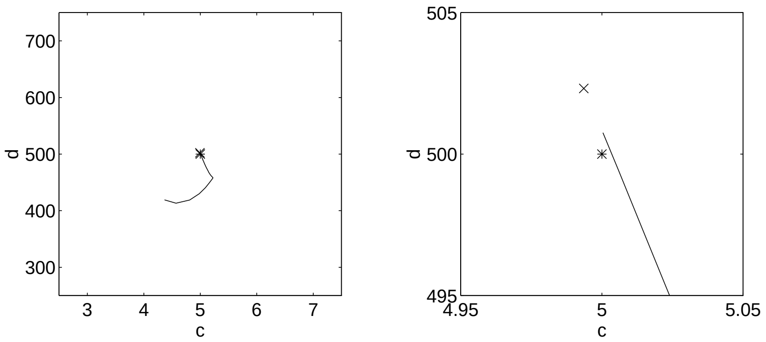

Figure 7.17: Example 3: Convergence of $x_r(D)$ to $x_\infty(D)$

7.6.4 Example 4

To generate the data for this example, we took the data from Example 1, introduced a random time–shift into each trace, and then added 10% random noise. The purpose of the time–shifts is to simulate a common type of noise found in seismic data. The objective functions and the path of minimizers are displayed as in Example 2.

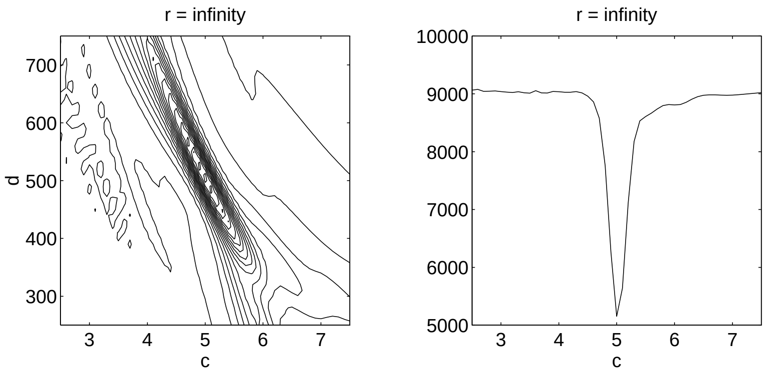

Figure 7.18: Example 4: $J_\infty(c,d)$

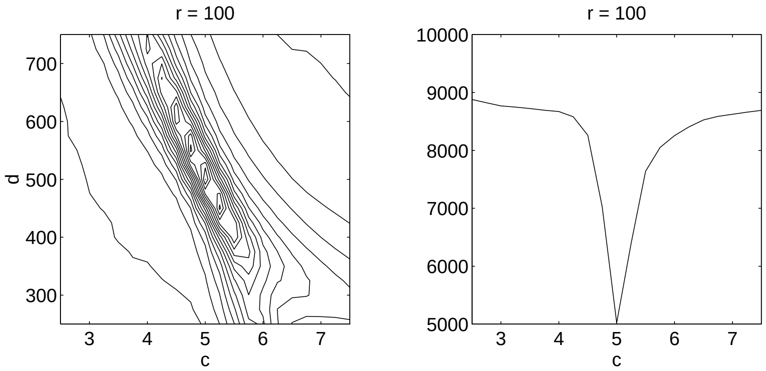

Figure 7.19: Example 4: $J_{100}(c,d)$

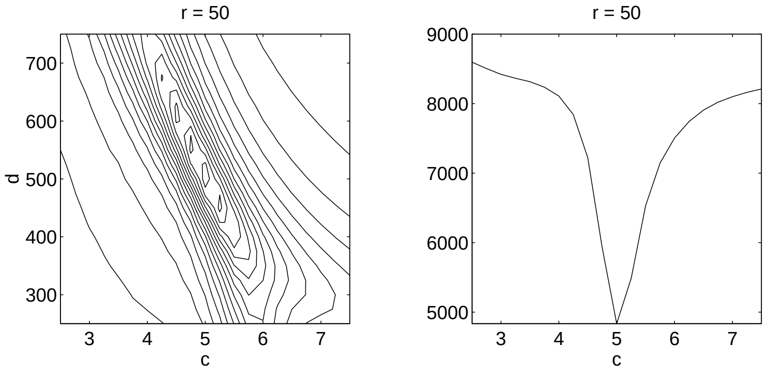

Figure 7.20: Example 4: $J_{50}(c, d)$

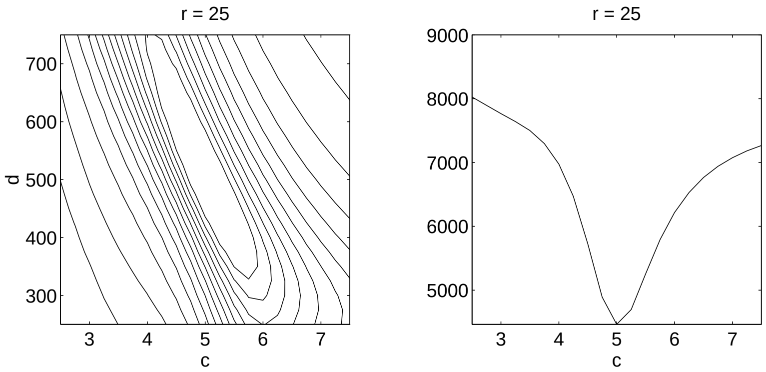

Figure 7.21: Example 4: $J_{25}(c, d)$

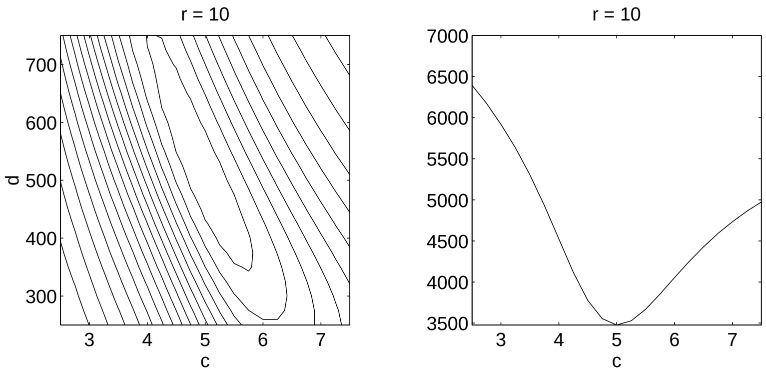

Figure 7.22: Example 4: $J_{10}(c, d)$

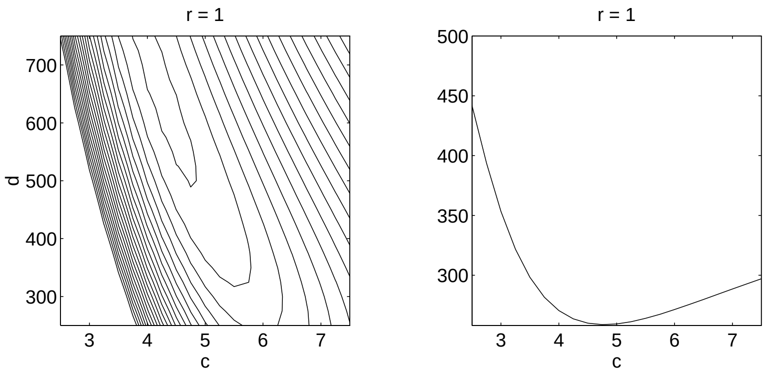

Figure 7.23: Example 4: $J_1(c, d)$

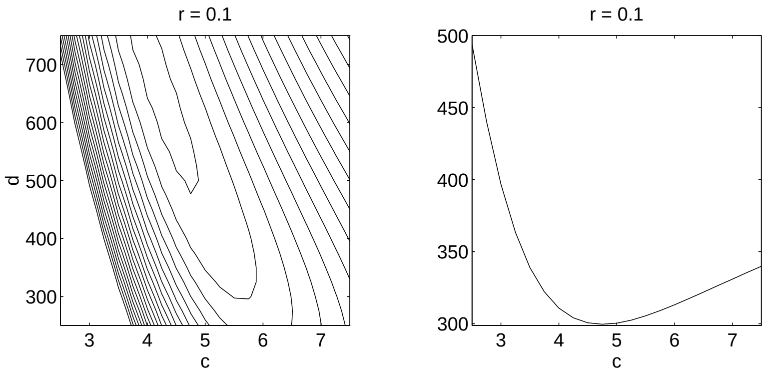

Figure 7.24: Example 4: $\frac{1}{0.1^2} J_{0.1}(c, d)$

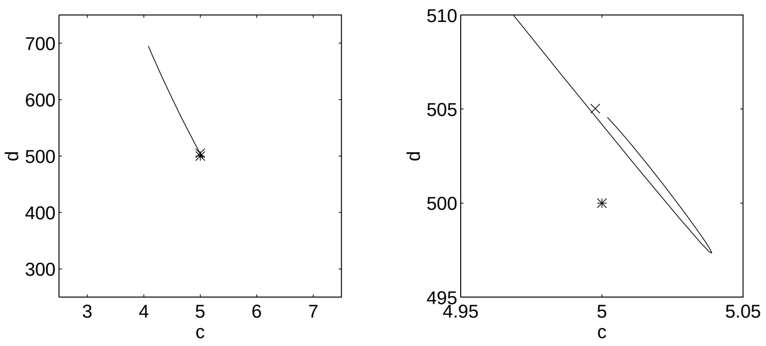

Figure 7.25: Example 4: Convergence of $x_r(D)$ to $x_\infty(D)$

References

[1] ENGL, H. W. [1993]. Regularization methods for the stable solution of inverse problems, *Surv. Math. Ind. 3*, 71–143.

[2] FIACCO, A. V. and McCORMICK, G. P. [1968]. Nonlinear Programming: Sequential Unconstrained Minimization Techniques, Wiley, New York.

[3] KOLB, P., COLLINO, F., and LAILLY, P. [1986]. Prestack Inversion of a 1D Medium, *Proceedings of IEEE 74*, 498–506.

[4] LOCKER, J. and PRENTER, P. M. [1980]. Regularization with Differential Operators, *J. Math. Anal. Appl. 74*, 504–529.

[5] LUENBERGER, D. G. [1973]. Introduction to Linear and Nonlinear Programming, Addison–Wesley, Reading, Massachusetts.

[6] NASHED, M. Z. [1974]. Approximate regularized solutions to improperly posed linear integral and operator equations, in *Constructive and Computational Methods for Differential and Integral Equations* (Colton, D. and Gilbert, R. P., Eds.), Lecture Notes in Mathematics No. 430, Springer–Verlag, Berlin/ Heidelberg/New York, 289–332.

[7] POLYAK, B. T. [1971]. The Convergence Rate of the Penalty Function Method, *Z. Vycisl. Mat. i Mat. Fiz., Vol. 11*, pp. 3–11.

[8] SEN, M. K. and STOFFA, P. L. [1991]. Nonlinear one-dimensional seismic waveform inversion using simulated annealing, *Geophysics 56*, 1624–1628.

[9] STOFFA, P. L. and SEN, M. K. [1991]. Nonlinear Multiparameter Optimization Using Genetic Algorithms: Inversion of Plane Wave Seismograms, *Geophysics 56*, 1794–1810.

[10] SYMES, W. W. and CARAZZONE, J. [1992]. Velocity Inversion by Coherency Optimization, *Proc. of Workshop in Geophysical Inversion*, J.B. Bednar, ed., SIAM.

[11] SYMES, W. W. [1990]. Velocity Inversion: a Case Study in Infinite-dimensional Optimization, *Mathematical Programming 48*, pp. 71–102.

[12] SYMES, W. W. [1992]. The Plane-Wave Detection Problem, Technical Report 92-22, Department of Computational and Applied Mathematics, Rice Universi ty, Houston, TX 77251

[13] SYMES, W. W. and CARAZZONE, J. [1991]. Velocity Inversion by Differential Semblance Optimization, *Geophysics 56*, 654–663.

[14] TIKHONOV, A. and ARSENIN, V. [1977]. Solutions of ill-posed problems, Wiley, New York.