

# A Primal Dual - Interior Point Framework for Using the L1-Norm or the L2-Norm on the Data and Regularization Terms of Inverse Problems

A. Borsic<sup>1</sup>, A. Adler<sup>2</sup>

<sup>1</sup>Thayer School Of Engineering, Dartmouth College, 8000 Cummings Hall Hanover, NH 03755, US

<sup>2</sup>Department of Systems and Computer Engineering Carleton University, 1125 Colonel By Drive Ottawa, Ontario, K1S 5B6, Canada

E-mail: [andrea.borsic@dartmouth.edu](mailto:andrea.borsic@dartmouth.edu), [adler@sce.carleton.ca](mailto:adler@sce.carleton.ca)

Submitted to: *Inverse Problems*

**Abstract.** Maximum A Posteriori (MAP) estimates in inverse problems are often based on quadratic formulations, corresponding to a Least Squares fitting of the data and to the use of the L2 norm on the regularization term. While the implementation of this estimation is straightforward and usually based on the Gauss Newton method, resulting estimates are sensitive to outliers, and spatial distributions of the estimates that are smooth. As an alternative, use of the L1 norm on the data term renders the estimation robust to outliers, and use of the L1 norm on the regularization term allows reconstructing sharp spatial profiles. The ability therefore of using the L1 norm either on the data term, on the regularization term, or on both is desirable. Use of this norm results though in non-smooth objective functions which require more sophisticated implementations compared to quadratic algorithms. Methods for L1 norm minimization have been studied in a number of contexts, including in the recently popular Total Variation regularization. Different approaches has been used and methods based on Primal Dual - Interior Point Methods (PD-IPM) have been shown to be particularly efficient. In the present manuscript we derive a PD-IPM framework for using the L1 norm indifferently on the two terms of an inverse problem. We use Electrical Impedance Tomography as an example inverse problem to demonstrate the implementation of the algorithms we derive, and the effect of choosing the L2 or the L1 norm on the two terms of the inverse problem. Pseudo codes for the algorithms and a public domain implementation are provided.

*Keywords:* L1-Norm, Least Absolute Values, Robust Estimation, Regularization, Total Variation, Primal Dual, Interior Point, Electrical Impedance Tomography

## 1. Introduction

In the present manuscript we derive a framework, based on a Primal Dual - Interior Point Method (PD-IPM), that allows choosing independently to use the L1-norm or the L2-norm on the data and on the regularization terms of discrete inverse problems. The use of the L1-norm on the data term leads to robust estimation in the presence of outliers in the data, and the use of the L1-norm on the regularization term leads to sharper spatial transitions in the estimated parameters. In the following we briefly introduce these choices, on the norms to be used, from the Bayesian standpoint.

We indicate with  $m$  the vector of model parameters, which belongs to  $\mathcal{M} \in \mathbb{R}^M$ ; we indicate with  $d$ , the vector of data observations, which belongs to  $\mathcal{D} \in \mathbb{R}^D$ , the space of data values; and we indicate with  $d = h(m)$  the forward model, that links the data  $d$  to the model parameters  $m$ .

Assuming now that the measurement and model uncertainties are additive and follow a generalized Gaussian distribution (Tarantola 2005), the likelihood of the difference between measured data and model predictions can be expressed as

$$\Theta(d | m) \propto \exp \left\{ - \sum_{i=1}^D \frac{|h(m)_i - d_i|^{n_d}}{(\sigma_{n_d})_i^{n_d}} \right\} \quad (1)$$

where  $\Theta(d | m)$  is the likelihood,  $n_d$  is the order of the generalized Gaussian distribution, and  $\sigma_{n_d}$  the generalized standard deviation. Assuming that the priors on the model parameters can be described by a generalized Gaussian distribution too, as

$$\Pi(m) \propto \exp \left\{ - \sum_{i=1}^M |L_i(m - m_0)|^{n_m} \right\} \quad (2)$$

where  $\Pi(m)$  is the distribution of the model parameters,  $m_0$  is a prior model state,  $L$  is a matrix describing the prior distribution,  $n_m$  is the order of the generalized Gaussian distribution. In the specific case of  $n_m = 2$  the matrix  $L^T L$  represents the covariance matrix of the model. The maximum a posteriori (MAP) estimate of the model parameters can be written as

$$m_{MAP} = \operatorname{argmin} \left\{ \sum_{i=1}^D |W_i(h(m)_i - d_i)|^{n_d} + \alpha \sum_{i=1}^M |L_i(m - m_0)|^{n_m} \right\} \quad (3)$$

where we indicate with  $W$  the diagonal matrix with the entries  $\frac{1}{(\sigma_{n_d})_i^{n_d}}$  on the diagonal. The matrices  $W$  and  $L$  can also be interpreted under the deterministic point of view to be respectively a data weighting matrix, and a regularization matrix corresponding to a Tikhonov style regularization of the inverse problem, and the scalar  $\alpha$  to be the Tikhonov parameter (Tikhonov & Arsenin 1977).

The choice of the model orders  $n_d$  and  $n_m$  should be made independently one from the other, and in such a way that each distribution models appropriately the respective parameters. In many practical cases model orders are chosen to be  $n_d = n_m = 2$ , as this

leads to least squares (LS) estimation problems where the Gauss-Newton method results in straightforward and efficient implementations. Furthermore least squares problems can be studied by singular valued decomposition (SVD) and by generalized SVD (GSVD) decomposition (Hansen 1998), which is a desirable property.

Use of the L1-norm in inverse problems has been gaining popularity in the last decade. Use of  $n_m = 1$  has assumed an important role in signal/image restoration and tomographic image reconstruction. Specifically, if  $L$  is a discretized representation of the gradient operator, using the L1-norm corresponds to Total Variation regularization (Rudin et al. 1992, Chan & Mulet 1996, Borsic 2002, Borsic, Graham, Adler & Lionheart 2010), which allows restoration/reconstruction of signals/images presenting step changes. Using  $n_d = 1$  on the likelihood corresponds to a long-tailed uncertainties distribution (Tarantola 2005), which results in estimates that are robust to data outliers. In general L1-norm minimization has been gaining momentum also in the context of Compressive Sensing (Candès et al. 2006), a method for reconstructing signals/images from incomplete frequency samples.

In this manuscript we derive a framework for solving (3) with either  $n_d = 1$ ,  $n_m = 1$ , or  $n_d = n_m = 1$ . In other words using the L1-norm on the data term, on the regularization term, or on both terms. The choice  $n_d = 2$ , the least squares fitting of the data, remains optimal for uncertainties that are normally distributed (Tarantola 2005), and the choice  $n_m = 2$ , remains a good choice if the model parameters have a smooth spatial variations. It is therefore worthwhile to consider all the four choices resulting from  $n_d = 1, 2$  and  $n_m = 1, 2$ , which correspond to the following MAP estimates

$$m_{MAP} = \operatorname{argmin} \left\{ \sum_{i=1}^D |W_i(h(m)_i - d_i)|^{n_d} + \alpha \sum_{i=1}^M |L_i(m - m_0)|^{n_m} \right\} \quad (4)$$

The MAP estimate for  $n_d = 2$  and  $n_m = 2$  is found trivially with well known methods for least squares problems. Gauss-Newton iterative algorithms are used typically, for example, in electrical, electromagnetic, and optical tomography problems (Polydorides & Lionheart 2002), (De Zaeytjijd et al. 2007), (Schweiger et al. 2005). For  $n_d = 1$  or  $n_m = 1$  the objective function involves sums of absolute values, and it is not differentiable at points where their arguments evaluate to zero: the MAP estimation is a non-smooth optimization problem. Points of non-differentiability typically occur in the very proximity of the sought solution, and so smooth optimization methods cannot be used even in a small region around the solution. In imaging applications, for example, the regularization matrix  $L$  takes differences of neighboring image elements, and  $m_0$  is the zero vector. The argument of  $|L_i(m)|$  is therefore zero when two neighboring image elements have the same value (or, in general, when the discrete differential operator evaluates locally to zero). In the image reconstruction process therefore, every time two image elements switch values with respect to each other, one becoming greater than the other, a kink in the objective function is crossed. It is therefore unfeasible to start from a initial guess and operate a smooth optimization without crossing kinks in the objective function. These optimization problems are therefore complex compared to the

quadratic case.

A body of literature is concerned with L1 estimation,  $n_d = 1$ , and with L1 regularization,  $n_m = 1$ . Algorithms used in the literature can be classified in two major types: Bayesian and deterministic. Bayesian approaches have been proposed, for example, by Kaipio *et. al.* in the context of Total Variation regularization in Electrical Impedance Tomography (Kaipio et al. 2000), where a Markov Chain - Montecarlo Method (MCMC) is used to provide MAP estimates. MCMC requires sampling the posterior probability density of the inverse problem, which consists in solving the forward problem for a large number of model configurations. This approach does not require differentiation of an objective function, and so no particular care is required if  $n_d = 1$  or  $n_m = 1$ . MCMC is however not attractive for non-linear problems, in applications such soft field tomography, as it requires solving the forward problem several thousand times, and therefore is very expensive computationally.

A plethora of approaches has instead been proposed from the deterministic point of view. Some algorithms transform the non-smooth optimization problem into a smooth one. In the context of L1 regularization for example (Acar & Vogel 1994, Chan et al. 1995, Vogel & Oman 1996, Dobson & Vogel 1997) use

$$m_{MAP} = \operatorname{argmin} \left\{ \|W(h(m) - d)\|^2 + \alpha \sum_{i=1}^M \sqrt{(L_i(m - m_0))^2 + \beta} \right\} \quad (5)$$

where the absolute value function  $|\cdot|$  is approximated with  $\sqrt{(\cdot)^2 + \beta}$ , which tends to the absolute value of the argument as  $\beta$  tends to zero, and  $\sqrt{(\cdot)^2 + \beta}$  is differentiable everywhere for  $\beta > 0$ . Traditional algorithms for smooth optimization (e.g. Steepest Descent, Newton Method) can therefore be used to solve an inverse problem with L1 norms on the data/regularization term with this approximation. In applications of L1 regularization it was shown that the Steepest Gradient method has an extremely slow convergence, and that the Newton method requires a very careful and adaptive control of  $\beta$ : solving with large values of  $\beta$  results in a poor approximation of  $|\cdot|$ , while solving with small values of  $\beta$  results in non-convergence of the method (Chan et al. 1995, Borsic 2002).

An other approach for L1 estimation is the Iteratively Re-weighted Least Squares (IRLS) algorithm (Weiszfeld 1937). IRLS is used, for example, for robust estimation in problems of the type

$$m_{MAP} = \operatorname{argmin} \left\{ \sum_{i=1}^D |W_i(h(m)_i - d_i)|^{n_d} + \alpha \|L(m - m_0)\|^2 \right\} \quad (6)$$

with values of  $n_d$  which are close to 1. IRLS casts (6) in an equivalent problem, which is

$$m_{MAP} = \operatorname{argmin} \left\{ \|D(h(m) - d)\|^2 + \alpha \|L(m - m_0)\|^2 \right\} \quad (7)$$

where  $D$  is a diagonal weighting matrix, whose  $i$ -th entry is:  $D_i = |W_i(h(m)_i - d_i)|^{n_d-2}$ . Equation (7) expresses (6) as least squares system, with special weights  $D$  that depend

on the current value of the estimated parameters  $m$ . Osborne (Osborne 1985) has proved linear convergence of the method for  $1 < n_p < 3$ . For L1-norm minimization though  $n_p = 1$ , resulting in weights which are

$$D_i = \frac{1}{|W_i(h(m)_i - d_i)|} \quad (8)$$

and anytime the argument  $(h(m)_i - d_i)$  tends to zero the weights tend to infinity, with  $n_p = 1$ , or close to 1. Similar tricks to (5) are used, for example by re-defining the weights to be

$$D_i = \frac{1}{\sqrt{W_i(h(m)_i - d_i)^2 + \beta}} \quad (9)$$

with  $\beta$  being a small positive scalar value that keeps the denominator from approaching zero. More complex reweighting strategies are considered by Daubechies in (Daubechies et al. 2010). The same considerations as before apply: a large  $\beta$  allows a smooth solution of the problem, but not an accurate one, a small value of  $\beta$  can cause convergence problems. A second aspect of IRLS that is not satisfactory is that its convergence is generally linear, while in certain cases it has been demonstrated to be super-linear (Daubechies et al. 2010), but in general the convergence is slow, and complicated by the choice of  $\beta$ , which should be potentially adaptive to be problem, and controlled iteration by iteration.

A third possibility contemplated in literature is that of transforming the original non-smooth optimization problem in a smooth convex program with linear constraints. For example, in the case of L1 regularization, the original MAP estimate can be modified in the equivalent constrained problem (Koh et al. 2007)

$$m_{MAP} = \operatorname{argmin} \left\{ \|W(h(m) - d)\|^2 + \alpha \sum_{i=1}^M u_i \right\} \quad (10)$$

subject to :  $-u_i \leq L_i(m - m_0) \leq u_i$

where  $u$  is a vector of auxiliary variables of length  $M$ . The problem (10) can be solved with the Sequential Quadratic Programming method (Bonnans 2006), with the Augmented Lagrangian methods (Bertsekas 1982), or with Interior Point methods (Koh et al. 2007). A review of optimization methods for L1-penalty problems is offered by (Loris 2009), where ulterior methods to the ones above are considered. Andersen *et. al.* (Andersen et al. 2000) have shown that Primal Dual - Interior Point methods can be very efficient at minimizing a sum of norms (or absolute values), compared to Interior Point methods and other classical methods. Chan *et. al.* (Chan et al. 1996) developed a PD-IPM framework for solving inverse problems in image restoration using Total Variation regularization and showed significantly superior performance compared to Gauss Newton methods such (5). Borsic *et. al.* (Borsic 2002) derived a PDIP framework, based on the developments of Andersen and Chan for L1-norm / Total Variation regularization in Electrical Impedance Tomography (EIT). The framework has been shown to be more efficient, over other standard methods, in minimizing objective functions

with  $n_m = 1$ , and to work well on experimental data. Primal Dual - Interior Point methods seem therefore to be an efficient and practical choice for minimizing the non smooth problems arising from the use of the L1 norm in the context of inverse problems.

In the present manuscript we derive a general PDIP framework for the efficient solution of the problem (4) with  $n_d = 1$ ,  $n_m = 1$ , or both  $n_d = n_m = 1$ . The PD-IPM framework we derive can be used on linear and non-linear problems, we do not prove convergence, but we use Electrical Impedance Tomography as a case problem and show successful results in terms of improving robustness to outliers in the data, when the L1 norm is used on the data term, and in terms of reconstructing sharp images, when the L1 norm is used on the regularization term.

The contents of the manuscript are organized as follows: in Section 2.1 the Primal Dual Interior Point framework is introduced, and a derivation for the L1-L2 problem ( $n_d = 1$ ,  $n_m = 1$ ) is given; in Section 2.2 the same framework is derived and applied to the L2-L1 problem ( $n_d = 2$ ,  $n_m = 1$ ); in Section 2.3 the PD-IPM framework is derived and applied to the L1-L1 problem ( $n_d = 1$ ,  $n_m = 1$ ). In Section 3 Electrical Impedance Tomography (EIT) is briefly introduced and used as a case problem; in Section 4 we discuss the implementation of the PD-IPM framework in the context of EIT and we provide pseudo-code diagrams that facilitate the implementation in the context of other inverse problems. Lastly in Section 5 we show numerical experiments using Gaussian noise and data outliers to highlight the effect of choosing different norms in the formulation of inverse problems, and to demonstrate that the PD-IPM framework we derive can be successfully used in practical applications.

## 2. Primal Dual - Interior Point Framework

In this section we derive methods for solving (4) based on Primal Dual - Interior Point methods. The case  $n_d = 2$ ,  $n_m = 2$  is not treated as it is a common least squares estimation problem, for which solutions are well known and typically based on the Gauss Newton method.

### 2.1. PD-IPM for $n_d = 1$ and $n_m = 2$

We will label as primal (P) the problem:

$$(P) = m_{MAP} = \operatorname{argmin} \left\{ \sum_{i=1}^D |W_i(h(m) - d)| + \alpha \|L(m - m_0)\|^2 \right\} \quad (11)$$

noting that for each  $i$

$$|W_i(h(m)_i - d_i)| = \max_{x_i: |x_i| \leq 1} x_i(W_i(h(m)_i - d_i)) \quad (12)$$

as the auxiliary variable  $x_i$ , in the range  $[-1, 1]$ , will take the value either 1, or  $-1$ , depending on the absolute value of  $W_i(d-h(m))$ . Using this consideration, and applying it to (P), it is possible to derive a second problem, a maximization problem, that is labeled dual (D)

$$(D) = \min_m \left\{ \max_x x^T W(h(m) - d) + \alpha \|L(m - m_0)\|^2 \right\}, \quad \text{with } \|x_i\| \leq 1 \quad (13)$$

in this context the auxiliary variables  $x$  are called dual variables, and the model parameters  $m$  primal variables. Interchanging the max and the min, which is possible if the forward operator results in a convex objective function (Rockefeller 1970), we obtain

$$(D) = \max_x \min_m \left\{ x^T W(h(m) - d) + \alpha \|L(m - m_0)\|^2 \right\}, \quad \text{with } \|x_i\| \leq 1 \quad (14)$$

and imposing the first order conditions on the primal variables

$$\begin{aligned} \frac{\partial}{\partial m} \left\{ x^T W(h(m) - d) + \alpha \|L(m - m_0)\|^2 \right\} = \\ J^T(m)Wx + 2L^T L(m - m_0) = 0 \end{aligned} \quad (15)$$

we can simplify the dual, removing the inner minimization, obtaining

$$\begin{aligned} (D) = \max_x \left\{ x^T W(h(m) - d) + \alpha \|L(m - m_0)\|^2 \right\} \\ |x_i| \leq 1 \\ J^T(m)Wx + 2\alpha L^T L(m - m_0) = 0 \end{aligned} \quad (16)$$

where  $J(m)$  is the Jacobian matrix of the forward model  $h(m)$ .

For feasible points of the dual problem (i.e.  $|x_i| \leq 1$ ) the objective function of (P) can be shown to take greater values than the objective function of the dual problem (D) (Andersen et al. 2000). The two objective functions take the same value at a single point, which is the optimal point both for (P) and for (D) (Andersen et al. 2000). Such optimal point can therefore be sought by nulling the difference between the two objective functions, called the *primal-dual gap*

$$\begin{aligned} G_{PD} &= \sum_{i=1}^D |W_i(h(m)_i - d_i)| + \alpha \|L(m - m_0)\|^2 - \\ &x^T W(h(m) - d) + \alpha \|L(m - m_0)\|^2 = \\ &\sum_{i=1}^D \{|W_i(h(m)_i - d_i)| - x_i W_i(h(m) - d)\} \end{aligned} \quad (17)$$

The primal-dual  $G_{PD}$  is null if, for each  $i$ , either  $W_i(h(m) - d) = 0$  or  $x_i = W_i(h(m) - d)/|W_i(h(m) - d)|$ . The condition that nulls the PD gap is therefore

$$|W_i(h(m)_i - d_i)| - x_i W_i(h(m) - d) = 0 \quad \forall i \quad (18)$$

and is called *complementarity condition* (Andersen et al. 2000), which captures the optimality of both (P) and (D). The strategy on which Primal Dual methods are based

is, rather to minimizing (P) or maximizing (D), to null the PD gap, by enforcing the complementarity condition, and obtaining the optimality. This, with the feasibility conditions of above, results in

$$\begin{aligned} \sum_{i=1}^M |W_i(h(m) - d_i)| - x^T W(h(m) - d) &= 0 \\ \|x_i\| &\leq 1 \\ J^T(m)Wx + 2\alpha L^T L(m - m_0) &= 0 \end{aligned} \quad (19)$$

which constitutes the Primal Dual method applied to the problem (4) with  $n_d = 1$  and  $n_m = 2$ . Equations (19) need to be solved jointly on  $m$  and  $x$ , for example with the Newton method. The absolute value appearing in (19) needs to be smoothed in order to obtain differentiability. The smoothing is obtained by replacing  $|W_i(h(m) - d_i)|$  by  $\sqrt{(W_i(h(m) - d_i))^2 + \beta}$ , with  $\beta > 0$ , similarly to (5), though the meaning, and the effect, of this smoothing is different from (5) in the context of the PD framework. The smoothed feasibility condition is called the *centering condition* (Andersen et al. 2000) and it leads to a smooth pair of optimization problems ( $P_\beta$ ) and ( $D_\beta$ ). The effect of such condition is that solutions are approached, as  $\beta$  is decreased, from points away from the boundary  $\|x_i\| \leq 1$  of the feasible region, from which the name of centering condition and the notion of interior point methods. In practical terms PD Interior Point Methods (PD-IPM) methods don't pose convergence issues related to the value of  $\beta$  as (5) does (Chan et al. 1996) (Borsic 2002), in our implementation, for example, we work with a small and fixed value of  $\beta$  (see Section 5), obtaining good converge rates. The choice of  $\beta$  instead in methods like (5) strongly affects convergence, and it is hard to operate a control on  $\beta$  that guarantees stability, let alone a fast convergence.

With the centering condition the PD framework becomes

$$(h_i(m) - d_i) - x_i \sqrt{(h(m) - d)^2 + \beta} = 0 \quad \forall i \quad (20)$$

$$\|x_i\| \leq 1 \quad (21)$$

$$J^T(m)Wx + 2\alpha L^T L(m - m_0) = 0 \quad (22)$$

and the Gauss Newton method is applied to solve it. In this context one wants to find the derivatives of (20) and (22) with respect to  $\partial m$  and  $\partial x$  and to impose first order conditions. For (20) we have

$$\frac{\partial}{\partial m} \left[ (h(m) - d) - x \sqrt{(h(m) - d)^2 + \beta} \right] = (I - XE^{-1}F)J(m) \quad (23)$$

where we have defined

- $f_i = h_i(m) - d_i$
- $F = \text{diag}(f_i)$
- $X = \text{diag}(x_i)$
- $\eta_i = \sqrt{f_i^2 + \beta}$

- $E = \text{diag}(\eta_i)$

and for the partial derivatives of (20) with respect to  $x$

$$\frac{\partial}{\partial x} \left[ (h(m) - d) - x\sqrt{(h(m) - d)^2 + \beta} \right] = -E \quad (24)$$

while for (22) we have

$$\frac{\partial}{\partial m} [J^T(m)Wx + 2\alpha L^T L(m - m^*)] = 2\alpha L^T L \quad (25)$$

and

$$\frac{\partial}{\partial x} [J^T(m)Wx + 2\alpha L^T L(m - m^*)] = J^T(m)W \quad (26)$$

yielding the Newton system for solving the PD-IPM problem as

$$\begin{bmatrix} 2\alpha L^T L & J^T W \\ (I - XE^{-1}F)J & -E \end{bmatrix} \begin{bmatrix} \delta m \\ \delta x \end{bmatrix} = - \begin{bmatrix} J^T Wx + 2\alpha L^T L(m - m^*) \\ f - Ex \end{bmatrix} \quad (27)$$

where we have dropped the explicit indication that  $J$  depends on  $m$ . The above system of equations can therefore be solved in an iterative fashion computing jointly the updates  $\delta m$  for the primal variables and  $\delta x$  for the dual variables. By substituting the second equation into the first it is possible to express the updates for the primal and dual variables separately as

$$\delta m = -[J^T W E^{-1}(I - XE^{-1}F)J + 2\alpha L^T L]^{-1}[J^T W E^{-1}f - 2\alpha L^T L(m - m^*)] \quad (28a)$$

$$\delta x = E^{-1}(f - Ex) + E^{-1}(I - XE^{-1}F)J\delta m; \quad (28b)$$

As  $\delta m$  is a descent direction for the primal problem (Andersen et al. 2000) a traditional line search procedure (Nocedal & Wright 1999) can be applied on  $(P_\beta)$ , finding an appropriate step length  $\lambda_m$  resulting in the update  $m^{(k+1)} = m^{(k)} + \lambda_m \delta m^{(k)}$ , where  $k$  is the iteration number.

Some care must be taken on the dual variable update, to maintain dual feasibility. A traditional line search with feasibility checks is not suitable as the dual update direction is not guaranteed to be an ascent direction for the modified dual objective function  $(D_\beta)$ . The system (27) is designed to minimize the primal-dual gap, but not necessarily to increase  $(D_\beta)$ . The update  $\delta m$  happens to be a descent direction for  $(P_\beta)$ , but  $\delta x$  is instead not guaranteed to be an ascent direction for  $(D_\beta)$ .

The simplest way therefore to compute the update is called the *scaling rule* (Andersen et al. 2000)

$$x^{(k+1)} = \varphi^* (x^{(k)} + \delta x^{(k)}) \quad (29)$$

where  $\varphi^*$  is a scalar value such that

$$\varphi^* = \sup \left\{ \varphi : \varphi \left| x_i^{(k)} + \delta x_i^{(k)} \right| \leq 1, \quad i = 1, \dots, n \right\} \quad (30)$$

An alternative way is to calculate the exact step length to the boundary, applying what is called the *step length rule* (Andersen et al. 2000)

$$x^{(k+1)} = x^{(k)} + \min(1, \varphi^*) \delta x^{(k)} \quad (31)$$

where  $\varphi^*$  is a scalar value such that

$$\varphi^* = \sup \left\{ \varphi : \left| x_i^{(k)} + \varphi \delta x_i^{(k)} \right| \leq 1, \quad i = 1, \dots, n \right\} \quad (32)$$

Usually the computation involved in calculating the exact step length to the boundary of the dual feasibility region is negligible compared to an iteration of the algorithm. It is preferable therefore to adopt the step length rule over the scaling rule, as the second places  $x$  always on the boundary of the feasible region, which prevents the algorithm from following the central path (sequence of points in the interior of the feasibility region that converges to the optimal point) - defeating effectively the purpose of the centering condition.

One important observation regarding the system (27) is that the matrices  $E$  and  $F$  depend on  $f$ , the discrepancy between forward solutions and measured data, and therefore on  $m$ . Even for linear problems therefore equations (27) must be iterated, until convergence is reached.

## 2.2. PD-IPM for $n_d = 2$ and $n_m = 1$

Formulating the inverse problem using the L2-norm on the data term and the L1-norm on the regularization term corresponds to the following formulation for the primal

$$(P) = \min_m \left[ \|W(h(m) - d)\|^2 + \alpha \sum_j |L_j(m - m_0)| \right] \quad (33)$$

using now a dual auxiliary variable  $y$ , we can derive a dual formulation ( $D$ ) with a similar procedure to (12) and (13)

$$(D) = \min_m \left[ \|W(h(m) - d)\|^2 + \max_y [\alpha y^T L(m - m_0)] \right], \quad \text{with } \|y_j\| \leq 1 \quad (34)$$

using min-max theory we can exchange the min and max (Rockefeller 1970)

$$(D) = \max_y \left[ \min_m [\|W(h(m) - d)\|^2 + \alpha y^T L(m - m_0)] \right], \quad \text{with } \|y_j\| \leq 1 \quad (35)$$

to perform the inner minimization on  $m$ , we impose the first order conditions on the objective function, nulling the derivative w.r.t to  $m$ .

$$\frac{\partial}{\partial m} \|W(h(m) - d)\|^2 = 2 J^T(m) W^T W (h(m) - d) \quad (36)$$

and

$$\frac{\partial}{\partial m} y^T L(m - m_0) = L^T y \quad (37)$$

so the first order conditions for the minimization in the dual problem are

$$2 J^T(m)W^TW(h(m) - d) + \alpha L^T y = 0 \quad (38)$$

and the primal dual gap is

$$G_{PD} = \|W(h(m) - d)\|^2 + \alpha \sum_j |L_j(m - m_0)| - \|W(h(m) - d)\|^2 - \alpha y^T L(m - m_0) \quad (39)$$

the PD framework is therefore formed by nulling the PD gap and by enforcing the feasibility conditions

$$\sum_j |L_j(m - m_0)| - y^T L(m - m_0) = 0 \quad (40)$$

$$\|y_j\| \leq 1 \quad (41)$$

$$2 J^T(m)W^TW(h(m) - d) + \alpha L^T y = 0 \quad (42)$$

the primal dual-gap will be null if and only if

$$L_j(m - m_0) - y_j |L_j(m - m_0)| = 0 \quad \forall j \quad (43)$$

the above condition is not differentiable when  $L_j(m - m_0) = 0$ , so using a centering condition

$$L_j(m - m_0) - y_j \sqrt{(L_j(m - m_0))^2 + \beta} = 0 \quad \forall j \quad (44)$$

with  $\beta > 0$ . The PD-IPM framework can then be formulated as

$$L_j(m - m_0) - y_j \sqrt{(L_j(m - m_0))^2 + \beta} = 0 \quad \forall j \quad (45)$$

$$\|y_j\| \leq 1 \quad (46)$$

$$2 J^T(m)W^TW(h(m) - d) + \alpha L^T y = 0 \quad (47)$$

The Newton method can now be applied to the above system. The partial derivatives of (47) are as follows

$$\frac{\partial}{\partial m} [2 J^T(m)W^TW(h(m) - d) + \alpha L^T y] = 2 J(m)^T W^T W J(m) \quad (48)$$

$$\frac{\partial}{\partial y} [2 J^T(m)W^TW(h(m) - d) + \alpha L^T y] = \alpha L^T \quad (49)$$

in deriving (48)  $J(m)$  has been considered constant with  $m$ , as this is one step of the Gauss-Newton algorithm, where the problem is linearized. Defining now

- $g_j = L_j m$

- $G = \text{diag}(g_j)$
- $Y = \text{diag}(y_j)$
- $s_j = \sqrt{(L_j m)^2 + \beta}$
- $S = \text{diag}(s_j)$

expressing now in matrix form the partial derivatives of (45) one obtains

$$\frac{\partial}{\partial m} \left[ L_j(m - m_0) - y_j \sqrt{(L_j(m - m_0))^2 + \beta} \right] = (I - YS^{-1}G)L \quad (50)$$

$$\frac{\partial}{\partial y} \left[ L_j(m - m_0) - y_j \sqrt{(L_j(m - m_0))^2 + \beta} \right] = -S \quad (51)$$

which, dropping the explicit indication of dependance of  $J$  on  $m$ , allows formulating a system for the primal and dual updates

$$\begin{bmatrix} 2 J^T W^T W J & \alpha L^T \\ (I - YS^{-1}G)L & -S \end{bmatrix} \begin{bmatrix} \delta m \\ \delta y \end{bmatrix} = - \begin{bmatrix} 2 J^T W^T W (h(m) - d) + \alpha L^T y \\ g - Sy \end{bmatrix} \quad (52)$$

where, similarly to 27, it is possible to work out separately the primal and dual updates as

$$\delta m = -[2J^T W^T W J + \alpha L^T S^{-1}(I - YS^{-1}G)L]^{-1} [2J^T W^T W (h(m) - d) - \alpha L^T S^{-1}g] \quad (53a)$$

$$\delta y = S^{-1}(g - Sy) + S^{-1}(I - YS^{-1}G)L\delta m \quad (53b)$$

As for (27) a line search can be used on the primal updates, a step length procedure on the dual updates, and the above equations iterated to convergence.

### 2.3. PD-IPM for $n_d = 1$ and $n_m = 1$

Formulating the inverse problem using the L1-norm on the data term and on the regularization term corresponds to the following formulation for the primal

$$(P) = \min_m \left[ \sum_i W_i |h_i(m) - d_i| + \alpha \sum_j |L_j(m - m_0)| \right] \quad (54)$$

and, similarly to (12) and (13), the dual formulation is

$$(D) = \min_m \left[ \max_x [x^T W (h(m) - d)] + \max_y [\alpha y^T L (m - m_0)] \right], \text{ with } \|x_i\| \leq 1, \|y_i\| \leq 1 \quad (55)$$

Interchanging the max and the min, which is possible if the forward operator results in a convex objective function (Rockefeller 1970), we obtain

$$(D) = \max_x \max_y \left[ \min_m x^T W(h(m) - d) + \alpha y^T L(m - m_0) \right], \text{ with } \|x_i\| \leq 1, \|y_i\| \leq 1 \quad (56)$$

in order to perform the inner minimization on  $m$ , the first order conditions need to be imposed. The derivative w.r.t  $m$  of the dual is

$$\frac{\partial}{\partial m} [x^T W(h(m) - d) + \alpha y^T L(m - m_0)] = J^T(m)Wx + \alpha L^T y \quad (57)$$

and therefore the first order conditions for the inner minimization in the dual problem are

$$J^T(m)Wx + \alpha L^T y = 0 \quad (58)$$

and the primal dual gap is

$$G_{PD} = \sum_i W_i |h_i(m) - d_i| + \alpha \sum_j |L_j(m - m_0)| - x^T W(h(m) - d) - \alpha y^T L(m - m_0) \quad (59)$$

nulling the PD gap and enforcing primal and dual feasibility results in the primal-dual framework

$$\sum_i W_i |h_i(m) - d_i| + \alpha \sum_j |L_j m - m_0| - x^T W(h(m) - d) - \alpha y^T L(m - m_0) = 0 \quad (60)$$

$$\|x_i\| \leq 1, \|y_j\| \leq 1 \quad (61)$$

$$J^T(m)Wx + \alpha L^T y = 0 \quad (62)$$

the primal dual-gap is null if and only if

$$(h_i(m) - d_i) - x_i |h_i(m) - d_i| = 0, \forall i \quad \text{and} \quad (L_j(m - m_0)) - y_j |L_j m - m_0| = 0, \forall j \quad (63)$$

the above two conditions are not differentiable when either  $(h_i(m) - d_i) = 0$  or  $L_j(m - m_0) = 0$ , and a centering condition can be used, rewriting them as

$$(h_i(m) - d_i) - x_i \sqrt{(h_i(m) - d_i)^2 + \beta} = 0, \forall i \quad \text{and} \quad (64)$$

$$L_j(m - m_0) - y_j \sqrt{(L_j(m - m_0))^2 + \beta} = 0, \forall j \quad (65)$$

with  $\beta > 0$ . The PD-IPM framework can be formulated as

$$(h_i(m) - d_i) - x_i \sqrt{(h_i(m) - d_i)^2 + \beta} = 0, \quad \forall i \quad (66)$$

$$L_j(m - m^*) - y_j \sqrt{(L_j(m - m^*))^2 + \beta} = 0, \quad \forall j \quad (67)$$

$$\|x_i\| \leq 1, \|y_j\| \leq 1 \quad (68)$$

$$J^T(m)Wx + \alpha L^T y = 0 \quad (69)$$

One wants now to apply the Gauss Newton method to the above system on non-linear equations. As before, we compute the partial derivatives of (66), (67), and (69) with respect to the primal and dual variables imposing then first order conditions. For the derivatives of (66), similarly to (20), we have

$$\frac{\partial}{\delta m} \left[ (h_i(m) - d_i) - x_i \sqrt{(h_i(m) - d_i)^2 + \beta} \right] = (I - XE^{-1}F)J(m) \quad (70)$$

$$\frac{\partial}{\delta x} \left[ (h_i(m) - d_i) - x_i \sqrt{(h_i(m) - d_i)^2 + \beta} \right] = -E \quad (71)$$

$$\frac{\partial}{\delta y} \left[ (h_i(m) - d_i) - x_i \sqrt{(h_i(m) - d_i)^2 + \beta} \right] = 0 \quad (72)$$

for (67), similarly to (50) and (51), we have

$$\frac{\partial}{\delta m} \left[ L_j(m - m^*) - y_j \sqrt{(L_j(m - m^*))^2 + \beta} \right] = (I - YS^{-1}G)L \quad (73)$$

$$\frac{\partial}{\delta x} \left[ L_j(m - m^*) - y_j \sqrt{(L_j(m - m^*))^2 + \beta} \right] = 0 \quad (74)$$

$$\frac{\partial}{\delta y} \left[ L_j(m - m^*) - y_j \sqrt{(L_j(m - m^*))^2 + \beta} \right] = -S \quad (75)$$

for (68), we have

$$\frac{\partial}{\delta m} [J^T(m)Wx + \alpha L^T y] = 0 \quad (76)$$

$$\frac{\partial}{\delta x} [J^T(m)Wx + \alpha L^T y] = J^T(m)W \quad (77)$$

$$\frac{\partial}{\delta y} [J^T(m)Wx + \alpha L^T y] = \alpha L^T \quad (78)$$

where the Jacobian  $J(m)$  has been considered constant with  $m$  as this is one step of the Gauss Newton method and the objective function is approximated with its first order

Taylor series. A system of equations can be formed now, computing the updates for the primal and dual updates as

$$\begin{bmatrix} 0 & J^T W & \alpha L^T \\ (I - XE^{-1}F)J & -E & 0 \\ (I - YS^{-1}G)L & 0 & -S \end{bmatrix} \begin{bmatrix} \delta m \\ \delta x \\ \delta y \end{bmatrix} = - \begin{bmatrix} J^T W x + \alpha L^T y \\ f - E x \\ g - S y \end{bmatrix} \quad (79)$$

As for the previous cases, updates for the primal variables and for the two sets of the dual variables can be computed separately, with the following equations deriving from the system above

$$\delta m = -[J^T W E^{-1}(I - XE^{-1}F)J + \alpha L^T S^{-1}(I - YS^{-1}G)L]^{-1}[J^T W E^{-1}f - \alpha L^T S^{-1}g] \quad (80a)$$

$$\delta x = E^{-1}(f - E x) + E^{-1}(I - XE^{-1}F) J \delta m \quad (80b)$$

$$\delta y = S^{-1}(g - S y) + S^{-1}(I - YS^{-1}G) L \delta m \quad (80c)$$

Updates on the primal can be computed with a traditional line search procedure, while updates on the duals can be computed with a step length procedure.

### 3. Example Problem: Electrical Impedance Tomography

We use Electrical Impedance Tomography (EIT) as an example inverse problem to which we apply the PD-IPM framework derived in the previous sections. EIT is a imaging technique which allows to reconstruct the spatial distribution of electrical conductivity within a volume. A number of electrodes are applied to the surface of the body under investigation, certain electrodes are used for injecting and sinking currents, and the non-injecting electrodes are used for sensing potential differences that result from the current excitation. By using several injection / sense patterns it is possible to collect a full, linearly independent, set of data. The inversion of such data, through an inverse problem formulation, allows estimating the spatial distribution of conductivity within the volume encompassed by the electrodes. Image reconstruction is typically formulated as in (4) with  $n_d = 2$  and  $n_m = 2$ , where now the model parameters  $m$  represent the spatial distribution of conductivity, on a discretized grid, the data vector  $d$  represents the recorded potentials resulting from multiple excitation/sensing electrode pairs, and  $L$ , the regularization matrix, is usually a discrete representation of of a first or second order differential operator, with the effect of imposing a certain smoothness on the solution.

In EIT currents injected into the imaged body are low frequency alternating currents, with frequencies in the range 10Hz to 10MHz (Holder 2004). The forward problem is modeled with a low-frequency approximation, where the electric field is conservative and conduction currents dominant with respect to their displacement counterparts, which leads to the Laplace equation (Holder 2004)

$$\nabla \cdot \sigma \nabla u = 0 \quad \text{on } \Omega \quad (81)$$

where  $\sigma$  is the continuous distribution of conductivity of the body to be imaged,  $u$  is the electric potential, and  $\Omega$  represents the body to be imaged. Electrodes are modeled with boundary conditions that are referred to as the Complete Electrode Model (Somersalo et al. 1992) in the EIT literature. The model specifies the following boundary condition for each portion of the boundary,  $\partial\Omega_\ell$ , underneath electrode  $\ell$

$$u + z_c \sigma \frac{\partial u}{\partial n} = V_\ell \quad \text{on } \partial\Omega_\ell \quad \ell = 1 \dots K \quad (82)$$

where  $z_c$  is the contact impedance,  $\ell$  is the electrode number,  $V_\ell$  is the potential developed by electrode  $\ell$ ,  $K$  is the number of electrodes, and  $n$  the normal direction to the boundary of  $\Omega$ . An additional condition is that the flux of the normal component of the current density over the surface of each electrode must equal the applied currents

$$I_\ell = \int_{\partial\Omega_\ell} \sigma \frac{\partial u}{\partial n} \quad \ell = 1 \dots K \quad (83)$$

where  $I_\ell$  is the current applied at electrode  $\ell$ . In the interelectrode gaps, the following boundary condition is applied

$$\frac{\partial u}{\partial n} = 0 \quad \text{on } \partial\Omega \setminus \{\partial\Omega_1 \cup \dots \cup \partial\Omega_K\} \quad (84)$$

as currents enter and leave the body only at the electrode surfaces.

The forward model constituted by (81), (82), (83), and (84) is solved by discretizing the imaging domain  $\Omega$ , often by using the Finite Element Method (Polydorides & Lionheart 2002). In the discretized setting the continuous conductivity  $\sigma$  is represented with a finite number of regions of constant conductivity, the values of which constitute the model parameters  $m$ . The set of collected difference of potentials  $V_m - V_n$ , for various pairs  $(m, n)$  of sensing electrodes, and for various pairs of current injection/sink electrodes  $(k, p)$  constitutes the vector of data  $d$  to be fitted.

#### 4. Implementation

We use EIT as an example problem for implementing the four formulations (4) which derive from the possible choices of the  $n_d = 1, 2$  and  $n_m = 1, 2$  norms. Our goal is to highlight the effect of choosing the L2 or L1 norm on the data and regularization terms through numerical simulations, and to demonstrate that the PD-IPM framework is an effective approach. In this section we discuss the implementation of the algorithms, while the next section discusses numerical experiments.

Formulation (4) with  $n_d = n_m = 2$  does not need particular consideration as it is the classical quadratic approach, which is solved using a Gauss-Newton method, with the following update for the model parameters

$$\delta m = -[J^T W^T W J + \alpha L^T L]^{-1} [J^T W^T W (h(m) - d) - \alpha L^T L m] \quad (85)$$

$$m = m + \lambda_m \delta m \quad (86)$$

where  $\lambda_m$  is a scalar number, determined in this work using a parabolic line search procedure (Nocedal & Wright 1999).

Following (28), (53), and (80) the formulations L1-L2, L2-L1, and L1-L1 were respectively implemented, and pseudo code is provided in Figures 1, 2, and 3, to facilitate the reader in the implementation in EIT, or other problems. A public domain MATLAB-based implementation of these algorithms, for the EIT problem, is also available within the EIDORS project (Adler & Lionheart 2006): <http://www.eidors.org>

As mentioned earlier most algorithms using the L1 norm rely on a smoothing parameter  $\beta$ . Examples of these algorithms are IRLS, methods that approximate the absolute value of a variable  $z$  with  $\sqrt{z^2 + \beta}$ , and the PD-IPM method itself through the centering condition. Methods that are not based on PD-IPM are generally known to converge extremely slowly for small values of  $\beta$  (e.g.  $\beta = 0.01$ ) or to diverge (Chan et al. 1995) (Borsic 2002), and to generally require complex schemes where  $\beta$  is carefully decreased while the algorithm is progressing.

Guaranteeing stability and good convergence properties for these algorithms is hard. Contrary to these methods, the PD-IPM algorithm converges well even for small values of  $\beta$ . In our implementations we use a constant value of  $\beta = 1 \times 10^{-12}$  ( $\beta$  in this context is the centering parameter, which nevertheless is a smoothing parameter - see (20)). Not requiring a continuation scheme for  $\beta$  is a benefit of PD-IPM that greatly simplifies the implementation of MAP estimation algorithms based on the L1 norm, besides the fact that convergence is faster (Chan et al. 1995) (Borsic 2002).

The three algorithms in Figures 1, 2, and 3 start by finding the best uniform value of the parameters  $m$  that fits the data, used as a good starting point, and by setting the initial value for the dual variables to zero. In a loop the Jacobian of the forward model is computed and updated, and updates for the primal and dual variables are computed. A parabolic line search procedure (Nocedal & Wright 1999) is used for applying the update on the primal variables and a step length rule is used for applying the update to the dual variables. The value of the primal dual gap  $G_{PD}$  is computed at each cycle, and it is used in a termination condition. As  $G_{PD}$  tends to zero at convergence, iterations are terminated when a sufficient reduction is achieved. In our numerical experiments we require a relative reduction to  $1 \times 10^{-4}$  times with respect to the initial value. We found experimentally that iterating further does not result in any appreciable change in the images.

One last note regards the computational cost of PD-IPM compared to the Gauss Newton algorithm applied to quadratic formulations. Overall the single cycle of the PD-IPM algorithm requires almost the same cost than a Gauss Newton iteration. We compare below an L2-L2 update equation and an L1-L1 update equation for the primal variables

L2-L2 – Gauss Newton update

$$[J^T W^T W J + \alpha L^T L] \delta m = -[J^T W^T W (h(m) - d) - \alpha L^T L m] \quad (87)$$

L1-L1 – PD-IPM update

$$[J^T W E^{-1} (I - X E^{-1} F) J + \alpha L^T S^{-1} (I - Y S^{-1} G) L] \delta m = -[J^T W E^{-1} f - \alpha L^T S^{-1} g] \quad (88)$$

The left hand side of the two updates differs. In the first case ( $J^T \dots J$ ) is inner multiplied by  $W^T W$ , in the second by  $W E^{-1} (I - X E^{-1} F)$ . The matrices  $W$ ,  $E$ ,  $I$ ,  $X$ , and  $F$  are diagonal, and therefore computing their inverses and products is fast and negligible compared to the other computational costs such, for example, forward solving and Jacobian assembly. Similarly, the product ( $L^T L$ ) of the Gauss Newton update is inner multiplied by  $S^{-1} (I - Y S^{-1} G)$  in the PD-IPM update, which is again a diagonal matrix. The right hand sides in both (87) and (88) are readily computed, considering that  $E$  and  $S$  are diagonal matrices. The same considerations apply to computing the dual updates

$$\delta x = E^{-1} (f - E x) + E^{-1} (I - X E^{-1} F) J \delta m \quad (89)$$

$$\delta y = S^{-1} (g - S y) + S^{-1} (I - Y S^{-1} G) L \delta m \quad (90)$$

the matrices  $E$  and  $S$  are diagonal, so their inverses are immediate to compute.

The single iteration of the PD-IPM algorithm has therefore a cost which is almost identical to that of the Gauss Newton algorithms for quadratic inverse problems. The difference in the two methods resides instead in the number of iterations the two algorithms require. The PD-IPM method involves in the primal and dual updates diagonal weighting matrices that are functions of the current model parameters and of the current dual variables. As a consequence these algorithms require a larger number of iterations to reach convergence. In our numerical experiments reconstructions converged typically in approximately 10 iterations, while the Gauss Newton converges in 3 to 5 iterations. The computational cost of the PD-IPM algorithm resides therefore in a larger number of iterations.

## 5. Numerical Experiments

In this section we report numerical experiments comparing performance and characteristics of the four inverse formulations (4) resulting from the choices  $n_d = 1, 2$  and  $n_m = 1, 2$ , applied to the EIT inverse problem. All the simulations are 2D, as this allows an easier visualization of the results and faster simulation, but the formwork and formulations we provide are equally applicable to 3D reconstruction.

### 5.1. Meshes, Generation of Synthetic Data, Noise, and Outliers

A test resistivity profile was generated with a background value of  $100 \Omega \cdot \text{m}$ , with two round inclusions with a resistivity of  $50 \Omega \cdot \text{m}$ , as illustrated in Figure 4(b). A FEM mesh with 16 electrodes was generated (see Figure 4(a)), having 30,937 nodes, and synthetic voltages were computed using the opposite EIT protocol, where current is injected between electrodes that are opposite to each other, and voltages are sensed on pairs of

```

                                PD-IPM Algorithm - L1-L2 Norm

set  $\beta=1e-12$ ;
set terminate=0;

find homogeneous  $m : m = \operatorname{argmin} \|W(h(m) - d)\|$ ;
initialise dual variables  $x$  to zero;

set k=0;    (iteration number)

while not terminate
     $f = (h(m) - d)$ ;    (forward solve)
     $F = \operatorname{diag}(f_i)$ ;
     $X = \operatorname{diag}(x_i)$ ;
     $\eta_i = \sqrt{f_i^2 + \beta}$ ;
     $E = \operatorname{diag}(\eta_i)$ ;
     $G_{PD} = \sum_i |W_i f_i| - x^T W f$ ;    (compute the PD gap value)
     $J = J(m)$ ;    (compute/update Jacobian)
     $\delta m = -[J^T W E^{-1} (I - X E^{-1} F) J + 2\alpha L^T L]^{-1} [J^T W E^{-1} f - 2\alpha L^T L (m - m_0)]$ ;
     $\delta x = E^{-1} (f - E x) + E^{-1} (I - X E^{-1} F) J \delta m$ ;
     $\lambda_m = \operatorname{argmin} \sum_i |W_i (h(m + \lambda_m \delta m) - d)| + \alpha \|L(m - m_0)\|^2$ ;    (line search)
     $\lambda_x = \max \{ \lambda_x : \|x_i + \lambda_x \delta x_i\| \leq 1, i = 1, \dots, n \}$ ;    (step length rule)
     $m = m + \lambda_m \delta m$ ;
     $x = x + \min(1, \lambda_x) \delta x$ ;
    if sufficient reduction of PDGAP, set terminate=1;
    if k>kmax, set terminate=1;
    k=k+1;
end while

```

**Figure 1.** Pseudo code for the implementation of the image reconstruction algorithm using an L1-L2 norm formulation as in (11) and (28)

neighboring electrodes, resulting in 96 independent data points.

Synthetic additive Gaussian noise was generated to simulate systematic and random errors occurring in data acquisition processes. A noise vector,  $\nu$ , was generated to have the same dimension of the data by extracting samples from a Gaussian distribution with zero mean and standard deviation of 1, and then normalized to have the same standard deviation of the data as

$$\nu = \nu \frac{\operatorname{std}(d)}{\operatorname{std}(\nu)} \quad (91)$$

where  $\operatorname{std}(\cdot)$  is the standard deviation of a vector. We specify noise level in percentages,

```

                                PD-IPM Algorithm - L2-L1 Norm

set  $\beta=1e-12$ ;
set terminate=0;

find homogeneous  $m : m = \operatorname{argmin} \|W(h(m) - d)\|$ ;
initialise dual variables  $x$  to zero;

set k=0;    (iteration number)

while not terminate
     $g = Lm$ ;
     $G = \operatorname{diag}(g_i)$ ;
     $Y = \operatorname{diag}(y_i)$ ;
     $s_i = \sqrt{(L_i m)^2 + \beta}$ ;
     $S = \operatorname{diag}(s_i)$ ;
     $G_{PD} = \alpha \sum_i |L_i(m - m_0)| - \alpha y^T L(m - m_0)$  (compute the PD gap value)
     $J = J(m)$ ;    (compute/update Jacobian)
     $\delta m = -[2J^T W^T W J + \alpha L^T S^{-1}(I - Y S^{-1} G)L]^{-1}[2J^T W^T W(h(m) - d) - \alpha L^T S^{-1}g]$ ;
     $\delta y = S^{-1}(g - S y) + S^{-1}(I - Y S^{-1} G)L\delta m$ ;
     $\lambda_m = \operatorname{argmin} \|W(h(m) - d)\|^2 + \alpha \sum_i |L_i(m - m_0)|$ ; (line search)
     $\lambda_y = \max \{\lambda_y : \|y_i + \lambda_y \delta y_i\| \leq 1, \forall i\}$ ; (step length rule)
     $m = m + \lambda_m \delta m$ ;
     $y = y + \min(1, \lambda_y) \delta y$ ;
    if sufficient reduction of PDGAP, set terminate=1;
    if k>kmax, set terminate=1;
    k=k+1;
end while

```

**Figure 2.** Pseudo code for the implementation of the image reconstruction algorithm using an L2-L1 norm formulation as in (33) and (53)

for example we indicate the noisy data produced as  $d+0.01\nu$  to have a 1% level of noise.

Additionally to the Gaussian noise, a set of 4 synthetic outliers was generated, to be used for testing robustness of inverse formulations. In order to generate the outliers, four random data points were selected out of the total 96, and the arbitrary value of 10 was added or subtracted from the original data. Figure 5 shows plots for the original noiseless data, in blue color, and for the data corrupted with outliers, in red color. Data points are indicated with circles.

The simulation of outliers is rather important to represent situations that occur in the actual application of inverse problems techniques, and that often result in significant

```

                                PD-IPM Algorithm - L1-L1 Norm

set  $\beta=1e-12$ ;
set terminate=0;

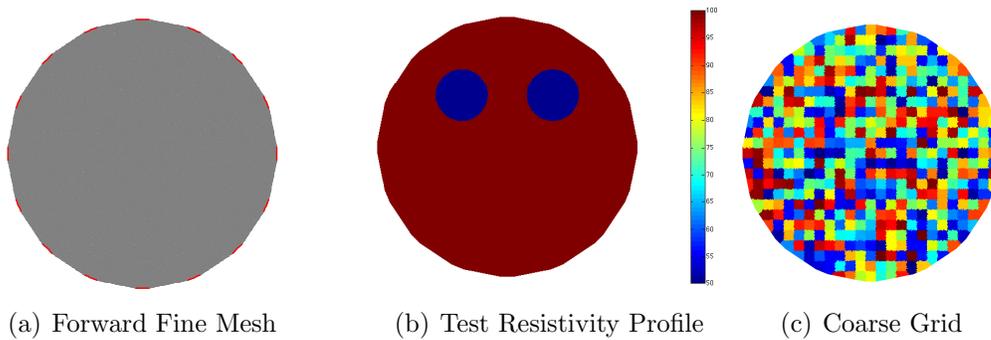
find homogeneous  $m : m = \operatorname{argmin} \|W(h(m) - d)\|$ ;
initialise dual variables  $x$  to zero;

set k=0;    (iteration number)

while not terminate
     $f = (h(m) - d)$ ;    (forward solve)
     $F = \operatorname{diag}(f_i)$ ;
     $X = \operatorname{diag}(x_i)$ ;
     $\eta_i = \sqrt{f_i^2 + \beta}$ ;
     $E = \operatorname{diag}(\eta_i)$ ;
     $g = Lm$ ;
     $G = \operatorname{diag}(g_i)$ ;
     $Y = \operatorname{diag}(y_i)$ ;
     $s_i = \sqrt{(L_i m)^2 + \beta}$ ;
     $S = \operatorname{diag}(s_i)$ ;
     $G_{PD} = \sum_i W_i |h(m) - d| + \alpha \sum_i |L_i(m - m_0)| - x^T W(h(m) - d) - \alpha y^T L(m - m_0)$ 
     $J = J(m)$ ;    (compute/update Jacobian)
     $\delta m = -[J^T W E^{-1}(I - X E^{-1} F) J + \alpha L^T S^{-1}(I - Y S^{-1} G) L]^{-1} [J^T W E^{-1} f - \alpha L^T S^{-1} g]$ ;
     $\delta x = E^{-1}(f - E x) + E^{-1}(I - X E^{-1} F) J \delta m$ ;
     $\delta y = S^{-1}(g - S y) + S^{-1}(I - Y S^{-1} G) L \delta m$ ;
     $\lambda_m = \operatorname{argmin} \sum_i W_i |h(m) - d| + \alpha \sum_i |L_i(m - m_0)|$ ;    (line search)
     $\lambda_x = \max \{ \lambda_x : \|x_i + \lambda_x \delta x_i\| \leq 1, \forall i \}$ ;    (step length rule)
     $\lambda_y = \max \{ \lambda_y : \|y_i + \lambda_y \delta y_i\| \leq 1, \forall i \}$ ;    (step length rule)
     $m = m + \lambda_m \delta m$ ;
     $x = x + \min(1, \lambda_x) \delta x$ ;
     $y = y + \min(1, \lambda_y) \delta y$ ;
    if sufficient reduction of PDGAP, set terminate=1;
    if k>kmax, set terminate=1;
    k=k+1;
end while

```

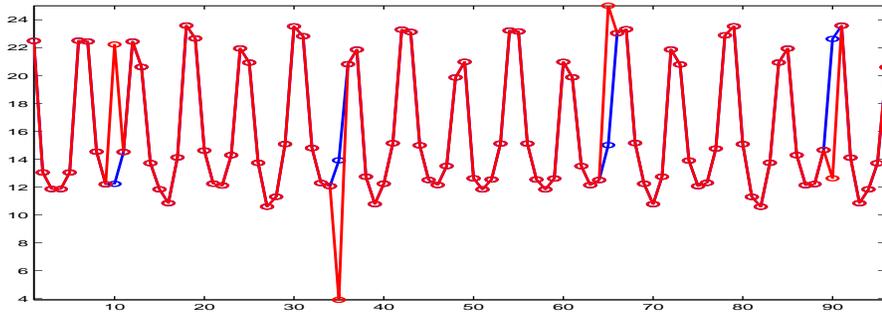
**Figure 3.** Pseudo code for the implementation of the image reconstruction algorithm using an L1-L1 norm formulation as in (54) and (80)



**Figure 4.** Left, a FEM mesh presenting 30,937 nodes, and 16 electrodes (indicated in red color) is used for computing synthetic data. Center, two round inclusions with a resistivity value of  $50 \Omega \cdot \text{m}$  and a background resistivity of  $100 \Omega \cdot \text{m}$  are simulated and used for generating synthetic data. Right, by grouping elements on a fine mesh, and constraining them to have the same conductivity value, it is possible to generate a coarse grid of 489 coarse elements to be used for estimating the resistivity. Coarse elements are shown in random colors.

artifacts in the reconstructed data. As an example, EIT finds applications in medical and geophysical applications. In both these fields it is not uncommon for a few electrodes to present contact impedances that are much higher than anticipated, resulting in data points that are outliers. In biomedical applications of EIT this can occur when one electrode is about to detach from the skin, presenting a poorer electrical contact with the body. In geophysical applications of EIT electrodes are applied to the surface in order to image the sub-surface. Locations where the soil is particularly dry might result in a particularly bad electrical contact between electrodes and the soil. Similar situations occur in other imaging modalities like electromagnetic and diffuse optical imaging, where a few sensors might not couple with the imaged medium as expected, resulting in bad data generated by one sensing channel. Being many of these problems ill-posed, an outlier of a moderate amplitude, for example of 20% relative amplitude, might be hard to spot by inspection of the data, but can disrupt the reconstruction. As we will show through numerical experiments, the availability of algorithms that are largely insensitive to outliers is therefore an important tool to deal with these sources of uncertainty, which are common in many practical applications, and many times hard to get rid of.

In the image reconstruction experiments, in order to avoid committing what is referred to as an *inverse crime* (Wirgin 2004), we use a mesh for image reconstruction that is different from the one used for computing the synthetic measurements. This second mesh was generated to have 20,663 nodes, and is used for forward solving in the reconstruction process. The number of inverse parameters is reduced to 489 by constraining several mesh elements to have the same resistivity value, forming thus “coarse pixels”. This arrangement is described in (Borsic, Halter, Wan, Hartov & Paulsen 2010). Figure 4(c) shows a rendering of the coarse grid resulting from grouping



**Figure 5.** The above graph shows the noiseless synthetic data, plotted in blue color, to which four outliers were added. The resulting noisy dataset is plotted in red color. The vertical axis represents measured potential differences at electrode pairs, while the horizontal line represents the data point index. Outliers were generated by adding or subtracting an arbitrary value of 10 to 4 random data points. This results in outliers with a relative amplitude ranging from 40% to 80%, depending on the data point.

FEM elements on the fine grid; coarse pixels have been randomly colored in order to highlight them.

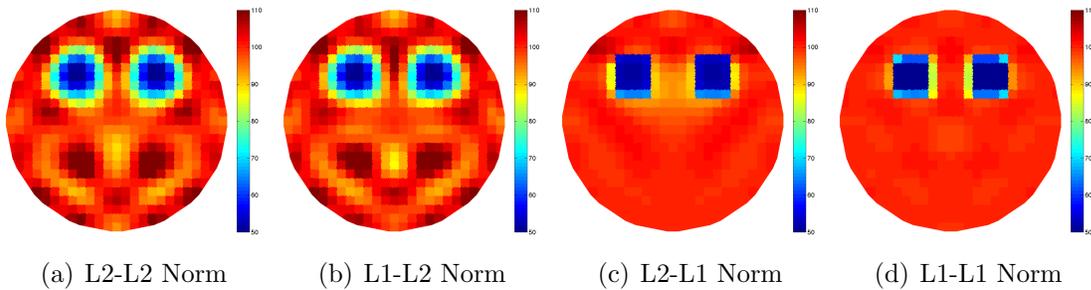
### 5.2. Choice of the Tikhonov Factor

Though not the focus of the present work, the choice of the Tikhonov factor affects greatly image reconstruction results. Our purpose is to show the effect of using different norms on the data and regularization terms of an inverse problem, and to demonstrate a practical framework for implementing different formulations, without focusing on a specific method for choosing the Tikhonov factor. Several methods exist for estimating the Tikhonov factor, including the L-Curve criterion, Generalized Cross Validation (Hansen 1998), and others (Graham & Adler 2006). While these methods have been developed traditionally for the quadratic formulations (L2-L2), some of these methods have been considered for application to inverse formulations involving the L1 norm (Clason et al. 2010) (Ito et al. 2011).

Our aim in the present work is to generate several test cases, involving noise and outliers, and to compare, test case by test case, the “best result” that each of the four formulations resulting from the choices  $n_d = 1, 2$  and  $n_m = 1, 2$  can produce. We have adopted a simple method method for choosing a Tikhonov factor, as follows

- (i) Synthetic data is produced from the resistivity profile  $\rho_{test}$  of Figure 4(b).
- (ii) Noise and/or outliers are added to the synthetic data, to generate different test cases.
- (iii) For each inverse formulation and for each test case, the Tikhonov factor is optimized as

$$\alpha_{optimal} = \underset{\alpha}{\operatorname{argmin}} \|\rho(\alpha)_{recon} - \rho_{test}\| \quad (92)$$



**Figure 6.** Reconstructions of synthetic data with no noise added for the four possible combinations of L1 and L2 norms on the data and regularization terms. Reconstructions using the L1 norm (Figures 6(c) and 6(d)) on the regularization term show sharper reconstructed profiles.

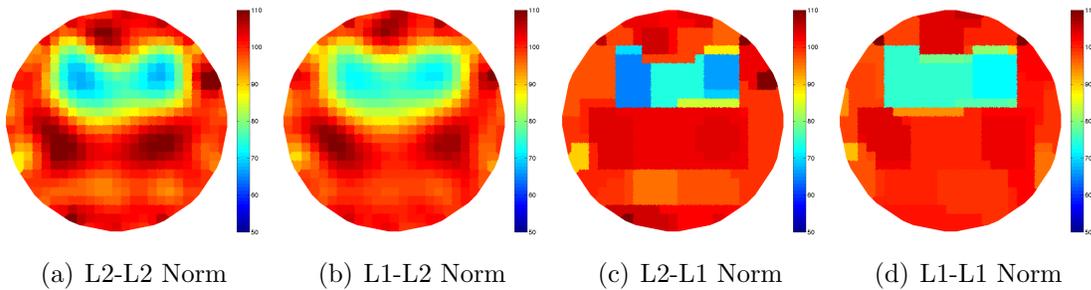
where  $\rho(\alpha)_{recon}$  is the reconstructed resistivity profile as a function of  $\alpha$ , the Tikhonov factor. In other words we optimize the Tikhonov factor in such a way that the reconstructed image best matches the original image. This arrangement allows comparing the “best results” that each method can produce. This is important, for example, for comparing algorithms in the presence of outliers, where the formulations using the L1 norm on the data term are able to correctly reconstruct images, but the L2-based formulations are unable to produce meaningful results, even if the Tikhonov factor is optimized against the known image. This approach of choosing the Tikhonov factor is of course applicable only where the original resistivity profile is known. We believe it might be appropriate in showing performances of different algorithms with an optimal choice of the Tikhonov factor and to abstract this discussion as possible from the choice of methods that estimate the Tikhonov factor.

### 5.3. Reconstructions

Several test cases have been prepared to compare reconstructed images resulting from different inverse formulations using the L1 and L2 norms. In all simulations we use a regularization matrix  $L$  which is a first order differential operator. This regularization matrix, when used with the L1-norm, corresponds to the Total Variation regularization, and it is well known to result in sharp reconstructed profiles.

In a first test case no noise was added to the synthetic data. Reconstructions for the L2-L2, L1-L2, L2-L1, and L1-L1 formulations are shown respectively in Figures 6(a), 6(b), 6(c), and 6(d). In this test case data is corrupted only by the numerical imprecisions in solving the forward model with finite computing precision. The four reconstructed profiles are plotted using the same color scale, ranging from 50  $\Omega\text{hm}$  (blue color) to 110  $\Omega\text{hm}$  (red color). This scale extends beyond the range of the test profile (50 to 100  $\Omega\text{hm}$ ) to accommodate for the fact that reconstructed resistivity “shoots” over this range at particular image locations.

The first two images on the left (L2-L2 and L1-L2 reconstructions) use the L2-norm on the regularization term, and thus result in reconstructions that correctly identify the



**Figure 7.** Reconstructions of synthetic data with 2% additive Gaussian noise for the four possible combinations of L1 and L2 norms on the data and regularization terms. The effect of using the L1 and L2 norms is similar to results of Figure 6 where no noise was used: the use of the L1 norm on the regularization term results in sharper reconstructions (Figures 7(c) and 7(d)), the use of the L1 or L2 norms on the data term produces results that are equivalent.

location and size of the contrasts and that are relatively smooth. As there is no noise on the data, the effect of using the L2-norm or the L1-norm on the data term is negligible and the first two reconstructed profiles are almost identical. The last two images on the right (L2-L1 and L1-L1 reconstructions) use the L1-norm on the regularization term, and demonstrate how this results in much sharper reconstructed profiles. The effect of using the L1-norm on the regularization term is therefore quite profound in terms of reconstructed profiles, and it is useful when the original profile is known to present step, or very fast, spatial changes.

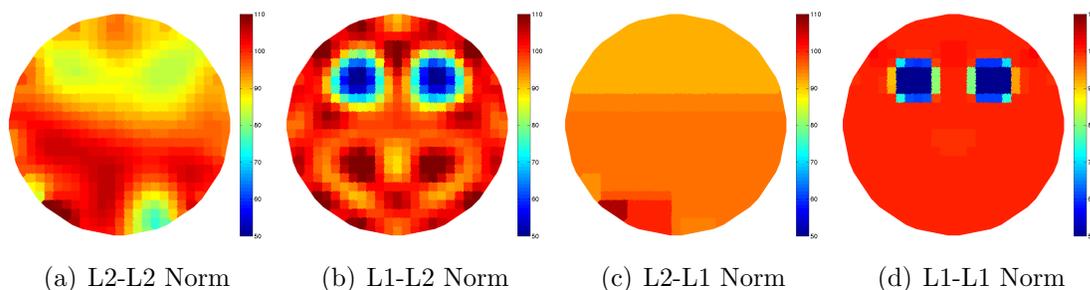
These regularization characteristics of the L1-norm are preserved in the presence of noise. Figure 7 illustrates results from a second test case, where 2% Gaussian noise was added to the synthetic data. Reconstructions for the L2-L2, L1-L2, L2-L1, and L1-L1 formulations are shown respectively in figures 7(a), 7(b), 7(c), and 7(d). This test case represents a typical practical situation, where a certain amount of noise is present on the data. The L2-L2 and L1-L2 reconstructions, shown in the left two images, being both based on L2-norm regularization are similar. Both reconstructions correctly identify and locate the inclusions and show some artifacts due to the presence of noise. Both reconstructions are smooth, as quadratic regularization is used. The L2-L2 reconstruction, corresponding to a Least Squares fitting of the data performs slightly better than the L1-L2 reconstruction, as in the presence of purely Gaussian noise Least Squares fitting is optimal (Tarantola 2005). The two images on the right show L2-L1 and L1-L1 reconstructions. As both these cases are based on L1-norm regularization, reconstructions are sharper, and the L2-L1 reconstruction performs slightly better, as, again, the use of the L2-norm on the data term is optimal in this case.

The previous two test cases (Figure 6 and 7) are useful to demonstrate the effect of using different norms on the regularization term (smooth/sharp regularization). In the next two test cases outliers are used to demonstrate the effect of the L2/L1-norm on the data term when outliers are present in the data.

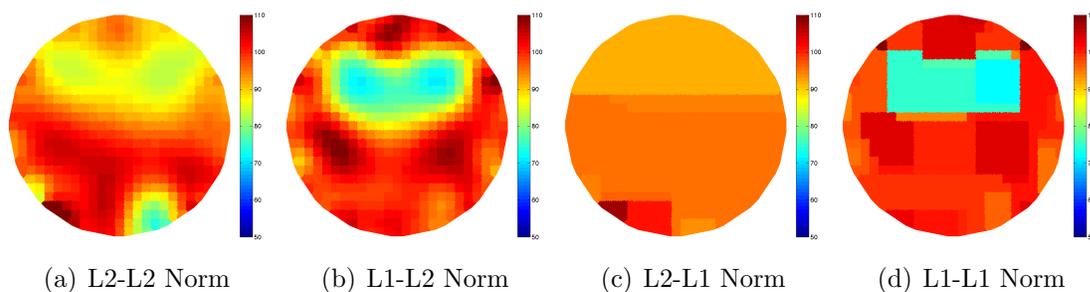
In a first test case the 4 outliers generated as discussed in Section 5.1 were added to the noiseless synthetic data, with no addition of Gaussian noise. Reconstruction results for this test case are shown in Figure 8. The second and fourth images from the left, Figures 8(b) and 8(d), correspond to the L1-L2 and L1-L1 formulations, they both are based on Least Absolute Fitting of the data. These reconstructions are insensitive to the presence of the outliers. Figure 8(b) is almost identical to Figure 6(b), which is the corresponding L1-L2 case where no noise was used. Similarly, Figure 8(d) is almost identical to Figure 6(d), which is the corresponding L1-L1 case where no noise was used. For these two reconstructions the presence of the outliers has no visible effect compared to the no-noise case. The difference between Figure 8(b) and 8(d) is that the first uses use of the L2-norm regularization, and the second L1-norm regularization, resulting respectively in a smooth or in a sharp reconstructed profile.

The first and the second images from the left, Figure 8(a) and 8(c), show results for the L2-L2 and L2-L1 formulations, correspond instead to a Least Squares fitting of the data. The Least Squares method presents a high sensitivity to the outliers. In Figure 8(a) the optimal choice of the Tikhonov factor, resulting in the best resemblance to the original test image as discussed in Section 5.2, produces an image that is very little informative about the location of the original contrasts. Given the sensitivity to outliers of L2 formulations, the optimal Tikhonov factor results very strong, as otherwise the image would be dominated by artifacts. Similarly the reconstruction based on the L2-L1 formulation, shown in Figure 8(c), results in a Tikhonov factor so strong that the reconstructed image is almost flat and completely uninformative. Any other choice for the Tikhonov factor would result in a worse discrepancy with respect to the original image. This test case shows therefore how the presence of outliers can severely affect formulations based on L2-norm fitting of the data.

In a last test case both outliers and Gaussian noise were added to the original synthetic data, to demonstrate performance in an application relevant scenario, where both types of disturbances are likely to be present. Precisely the 4 outliers of the previous case were used in conjunction with 2% Gaussian noise. As in the previous test case reconstructions based on a Least Squares formulation suffer significantly from the presence of outliers, while reconstructions based on Least Absolute Fitting cope well with them. Figures 9(b) and 9(d), corresponding to the L1-L2 and L1-L1 formulations, detect well the location of the contrasts. In Figure 9(b) the reconstructed profile is smooth and very similar to the reconstructed profile in the sole presence of Gaussian noise shown in Figure 7(b). The L1-L1 reconstruction, Figure 9(d), presents the sharp characteristic of the L1-norm regularization, and the reconstructed profile is similar to the profile in Figure 7(d), where only the gaussian noise was present. The use of the L1-norm on the data provides therefore robustness to the presence of outliers. In further numerical experiments we have found that the algorithms based on a Absolute Values fitting formulation continued to provide a similar quality in the reconstructed images with up to 20 outliers.



**Figure 8.** Reconstructions of synthetic data with outliers for the four possible combinations of L1 and L2 norms on the data and regularization terms. Reconstructions using the L2 norm on the data term (Figures 8(a) and 8(c)) are very sensitive to outliers, and result in the selection of large Tikhonov factors. Given the procedure for choosing the Tikhonov factor, reconstructed images are the closest possible to the original profile - any other choice of regularization strength would result in worse reconstructions. For the L2-L1 case the optimal Tikhonov factor is so large that the reconstructed image is almost flat. Reconstructions using the L1 norm on the data term (Figures 8(b) and 8(d)) are insensitive to the presence of the outlier, and result in the same image quality as in shown in Figure 6 where no noise was present.



**Figure 9.** Reconstructions of synthetic data with additive outliers plus additive 2% noise for the four possible combinations of L1 and L2 norms on the data and regularization terms. As in Figure 8, reconstructions using the L2 norm on the data term (Figures 9(a) and 9(c)) are very sensitive to outliers, and result in the selection of large Tikhonov factors. Given the procedure for choosing the Tikhonov factor, reconstructed images are the closest possible to the original profile - any other choice of regularization strength would result in worse reconstructions. For the L2-L1 case the optimal Tikhonov factor is so large that the reconstructed image is almost flat. Reconstructions using the L1 norm on the data term are insensitive to the presence of the outlier (Figures 9(b) and 9(d)), and result in the same image quality as in Figure 7, where only 2% noise was present (no outlier).

From the numerical experiments presented so far it is possible to appreciate how L1 regularization facilitates the reconstruction of sharp spatial changes in the images. At the same time L1 regularization might over-enhance certain transitions that are smooth in the original image. For example, the round inclusions used in the tests are reconstructed as two square objects when L1 regularization is used. In order to further demonstrate when certain choices on the data / regularization norms might or might

not be appropriate, we have conducted further experiments with a test resistivity profile which is smooth. These experiments are presented in the next subsection.

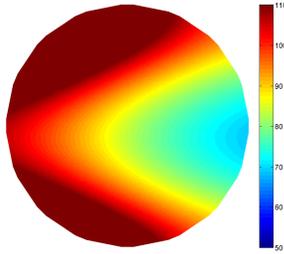
#### 5.4. Reconstructions With a Smooth Test Profile

The use of L1 regularization corresponds to the prior assumption that the profile to be reconstructed presents sharp transitions. When this is not the case, use of L1 regularization might not be optimal, and reconstructions will typically present a blocky appearance, even if the original profile is smooth. In order to demonstrate this behavior we have generated a test profile presenting a slow spatial variation, shaped like a cone, as illustrated in Figure 10. Synthetic data from this profile was generated with the same procedures used for the previous test cases, and the same Gaussian noise and outliers were used where needed.

A general consideration applies to the next reconstructions we present, which are produced from the smooth test profile. As this profile presents lower spatial frequencies compared to the test profile in Figure 4(b), it is less challenging to reconstruct, as few lower singular vectors of the Jacobian matrix can span relatively well the shape of the resistivity variation. The effect of the outliers across the simulations is therefore somewhat less pronounced: optimal values of the Tikhonov factor result in a sufficient strength of the regularization that artifacts are attenuated, at the same time the lower spatial frequencies allowed by this regularization settings can span relatively well the image.

Figure 11 shows reconstructions with no noise or outliers for the four possible combinations of the L1 and L2 norms. As no noise or outliers are present, the optimal Tikhonov factor is rather small for all the four formulations of the inverse problem, resulting in a lesser effect of the regularization, and in similar reconstructions. In Figure 11(d) it is possible to appreciate a somewhat a blocky effect deriving from the use of L1 regularization.

Differences between formulations become apparent when errors are added to the data, as in Figure 12, where the previously generated 2% Gaussian noise and outliers are used to corrupt the synthetic data. Figure 12(a) presents an L2-L2 reconstruction, where a strong value of the Tikhonov factor suppresses to some degree the effects of the outliers, which remain visible particularly at the bottom of the figure. The reconstructed image captures an area of lower resistivity on the right part of the figure, but the profile overall is disrupted by the presence of artifacts caused by outliers. Figure 12(b) is relative to an L1-L2 formulation. This formulation is optimal for this particular test case and results in the most faithful reconstruction: the L1 norm on the data renders the reconstruction robust to outliers, while the L2 regularization assumes a smooth resistivity prior, which matches the test case. Being this formulation less sensitive to outliers, the optimal Tikhonov factor is relatively small, allowing more singular vectors to contribute effectively to the image compared to the L2-L2 case, and results in a more faithful reconstruction which captures the conic characteristic of test profile.



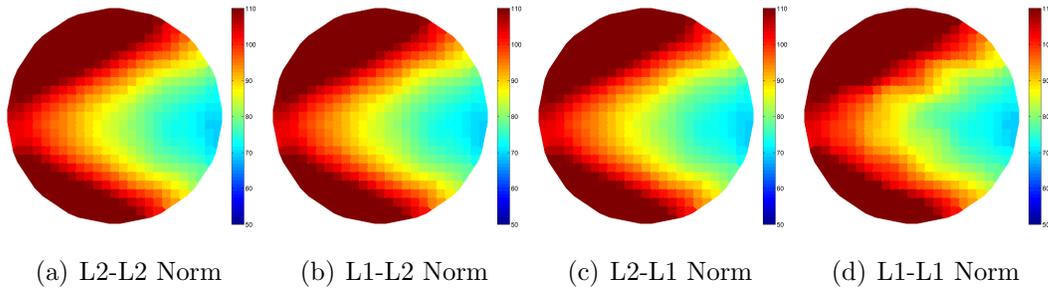
**Figure 10.** In order to study performance of different formulations involving the L2 and L1 norms in the presence of slowly varying resistivity profiles, a new test profile, illustrated in the above figure, was generated. The profile presents a conical resistivity change which spans smoothly the whole domain.

Figures 12(c) and 12(d) show results for formulations implementing L1 regularization, respectively the L2-L1 and L1-L1 formulations. As the test profile is smooth, this choice on the regularization norm, which assumes sharp transitions, is not optimal. Reconstructions are blocky even if the test profile is not. The main difference between Figure 12(c) and Figure 12(d) resides in the fact that the L1-L1 reconstruction is robust to outliers, and the resulting optimal Tikhonov factor is smaller than the one for L2-L1. This allows capturing better the shape of the test profile. The L2-L1 reconstruction, in Figure 12(c) is instead sensitive to outliers, and the optimal Tikhonov factor is so high that the test profile is reconstructed incorrectly.

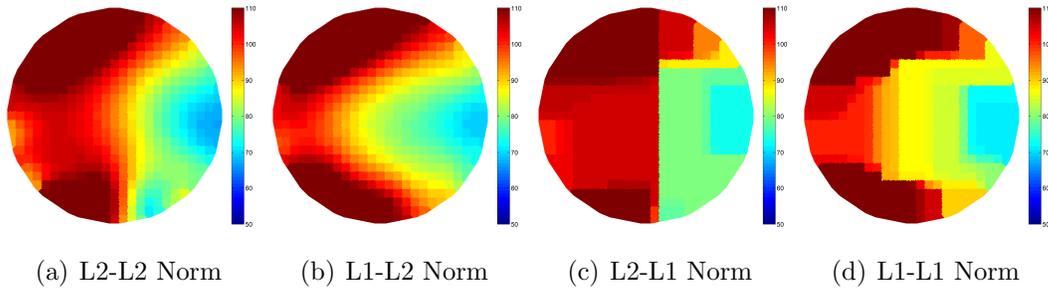
The tendency of reconstructing blocky images in the presence of slow spatial variations shown by L1 regularization, which appears in this second test case, is known in literature, and referred to as *staircase effect*. Approaches to avoid this undesired behavior are usually based on building regularization functionals that comprise an L1 term and an L2 term, and where the relative weight of the two terms is adaptively adjusted in such a way that L1 regularization is dominant in those regions of the image where fast spatial changes occur. One of the first of these approaches is the Higher Order Total Variation proposed by T. Chan, (Chan et al. 2000), several other works exist where L1 and L2 terms are mixed in an adaptive fashion. In the present manuscript we did not consider the possibility of building regularization terms that mix L1 and L2 norms, but this is possible with the PD-IPM framework, and it will be considered as future work.

### 5.5. Objective Function Plots

A last figure is used for illustrating the non-smooth characteristics of the optimization problems involving the L1 norm. Figure 13 illustrates contour plots for the objective functions resulting from the four possible inverse formulations. In each figure two resistivity image elements were swept in the interval of values  $[-0.005 \ +0.005]$  around the optimal value found by the reconstruction algorithms. Depending on whether one pixel value becomes greater than the other, resulting in a sign change in  $|L_j(m - m_0)|$ ,



**Figure 11.** In a first test no noise was added to the synthetic test data, and reconstructions were performed for the four possible combinations of L1 and L2 norms on the data and regularization terms. As in this case the optimal Tikhonov factor is small in value, the effect of regularization is limited, and all the reconstructed images shows similar characteristics: they match well the original profile. In Figure 11(d), on the top left part, it is possible to faintly appreciate the blocky behavior of L1 regularization.

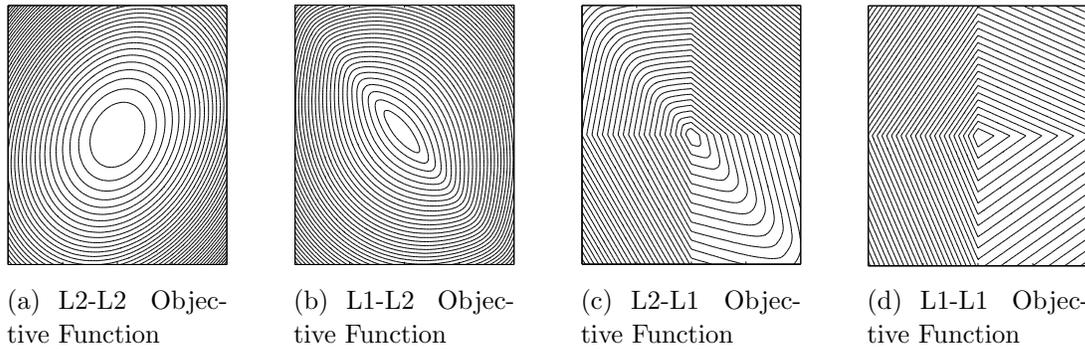


**Figure 12.** Reconstructions with 2% additive Gaussian noise and outliers for the four possible combinations of L1 and L2 norms on the data and regularization terms. The use of the L2 norm on the data term in Figure 12(a) results in sensitivity to outliers. The L1-L2 reconstruction in Figure 12(b) is instead optimal, as the use of the L1 norm on the data term reduces sensitivity to outliers, and the L2 model prior matches well the test profile. Reconstructions that use L1 norm regularization, Figures 12(c) and 12(c), suffer from the characteristic blockiness of this formulation, and the L2-L1 reconstruction of 12(c) suffers additionally from sensitivity to outliers, resulting in an incorrect reconstruction.

or, on whether the resistivity change results in a sign change in  $|h_i(m) - d_i|$ , ridges of non-differentiability are crossed. Ridges are evident in Figures 13(c), and 13(d), while Figure 13(b) is evidently non-quadratic.

## 6. Conclusions

The possibility of formulating an inverse problem using the L1-norm or the L2-norm on the data and regularization terms is attractive, as this allows better robustness to outliers and sharper profiles compared to reconstructions of quadratic algorithms. Unfortunately, the use of the L1 norm on either the data or the regularization terms results in an objective function that is not differentiable everywhere. While quadratic formulations of inverse problems can be solved with Gauss Newton methods and do



**Figure 13.** The above graphs represent contour plots for the objective functions of the four MAP formulations. The graphs were produced by perturbing the values of two image pixels around the optimal point, therefore producing a surface in 3D space. Kinks occur in the objective functions when either one pixel value becomes greater than the other, or when the difference between any predicted data point and measured data point switches sign.

not pose issues from the optimization point of view, formulations that use the L1-norm require specific implementations. We derive a framework, based on the Primal-Dual Interior Point Method, that allows using the L1/L2 norm on the data term and the L1/L2 norm on the regularization term. The framework is efficient at dealing with the resulting non-differentiability, and it is of general application to inverse problems. We use EIT, an ill-posed inverse problem, as a test case and provide detailed pseudo-code for facilitating re-implementation of the algorithms (a public domain implementation for EIT has been made available also through the toolbox EIDORS (Adler & Lionheart 2006) at <http://eidors.org>). We show through numerical test cases the effect of using different norms on the two terms of an inverse problem. Reconstructions using the L1-norm on the regularization term, with an appropriate choice of the regularization matrix, correspond to Total Variation regularization, and show the sharp transitions in the reconstructed profiles that are typical of this technique. Reconstructions using the L1-norm on the data, corresponding to a Least Absolute Squares fitting, result in improved robustness to the presence of outliers in the data, producing meaningful reconstructions where the quadratic approach fails or shows great sensitivity to the data errors. In terms of practical implementation, the computational time per single iteration of PD-IPM is similar to that of the Gauss Newton algorithm, but more steps are usually required for convergence, approximately 10 in our experiments.

The PD-IPM framework we present is therefore a viable solution to using the L1 norm in inverse problems, allowing to benefit from increased robustness to outliers and from shaper reconstructions, where necessary. We believe these results can be useful in numerous inverse problems. The proposed implementations are shown to work on test cases with syntectic noise and to cope well with data outliers. Work is currently under way to test these implementations on experimental data.

## 7. Acknowledgments

This work was partially supported from the NIH grants, P01-CA80139 awarded by the National Cancer Institute, and RC1-EB011000 awarded by the National Institute of Biomedical Imaging and Bioengineering.

## 8. References

- Acar R & Vogel C R 1994 *Inverse Problems* **10**, 1217–1229.
- Adler A & Lionheart W 2006 *Physiological measurement* **27**, S25.
- Andersen K D, Christiansen E, Conn A & Overton M L 2000 *SIAM J. on Scientific Computing* **22**, 243–262.
- Bertsekas D 1982 *Computer Science and Applied Mathematics, Boston: Academic Press, 1982* **1**.
- Bonnans J 2006 *Numerical optimization: theoretical and practical aspects* Springer-Verlag New York Inc.
- Borsic A 2002 Regularisation Methods for Imaging from Electrical Measurements PhD thesis Oxford Brookes University, School of Engineering.
- Borsic A, Graham B, Adler A & Lionheart W 2010 *IEEE Transaction on Medical Imaging* **29**, 44–54.
- Borsic A, Halter R, Wan Y, Hartov A & Paulsen K 2010 *Physiological measurement* **31**, S1.
- Candès E, Romberg J & Tao T 2006 *Information Theory, IEEE Transactions on* **52**(2), 489–509.
- Chan T F, Zhou H M & Chan R H 1995 *UCLA CAM Report 95-18*.
- Chan T, Golub G & Mulet P 1996 *ICAOS'96* pp. 241–252.
- Chan T, Marquina A & Mulet P 2000 *SIAM J. on Scientific Computing* **22**(2), 503–516.
- Chan T & Mulet P 1996 *UCLA CAM Report 96-38*.
- Clason C, Jin B & Kunisch K 2010 *SIAM Journal on Imaging Sciences* **3**, 199.
- Daubechies I, DeVore R, Fornasier M & Güntürk C 2010 *Communications on Pure and Applied Mathematics* **63**(1), 1–38.
- De Zaeytijd J, Franchois A, Eyraud C & Geffrin J 2007 *Antennas and Propagation, IEEE Transactions on* **55**(11), 3279–3292.
- Dobson D C & Vogel C R 1997 *SIAM J. on Numerical Analysis* **43**, 1779–1791.
- Graham B & Adler A 2006 *Physiological measurement* **27**, S65.
- Hansen P 1998 *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion* number 4 Society for Industrial Mathematics.
- Holder D 2004 *Electrical Impedance Tomography: Methods, History and Applications* Institute of Physics Publishing.
- Ito K, Jin B & Takeuchi T 2011 *SIAM Journal on Scientific Computing* **33**, 1415.
- Kaipio J P, Kolehmainen V, Somersalo E & Vauhkonen M 2000 *Inverse Problems* **16**, 1487–1522.
- Koh K, Kim S & Boyd S 2007 *Journal of Machine learning research* **8**(8), 1519–1555.
- Loris I 2009 *Inverse Problems* **25**, 035008.
- Nocedal J & Wright S J 1999 *Numerical Optimization* Springer.
- Osborne M 1985 *Finite algorithms in optimization and data analysis* John Wiley & Sons, Inc.
- Polydorides N & Lionheart W R B 2002 *Meas. Sci. Technol.* **13**, 1871–1883.
- Rockefeller R T 1970 *Convex Analysis* Princeton University Press Princeton University.
- Rudin L I, Osher S & Fatemi E 1992 *Physica D* **60**, 259–268.
- Schweiger M, Arridge S & Nissilä I 2005 *Physics in medicine and biology* **50**, 2365.
- Somersalo E, Cheney M & Isaacson D 1992 *SIAM J Appl Math* **52**, 1023–1040.
- Tarantola A 2005 *Inverse problem theory and methods for model parameter estimation* Society for Industrial Mathematics.
- Tikhonov A N & Arsenin V Y 1977 *Solutions of Ill-Posed Problems* Winston & Sons Washington D.C.
- Vogel C R & Oman M E 1996 *SIAM J. Sci. Computing* **17**, 227–238.

Weiszfeld E 1937 *Sur le point pour lequel la somme des distances de  $n$  points donnés est minimum*  
Print. Tôhoku Imp. Univ.

Wirgin A 2004 *Arxiv preprint math-ph/0401050* .