

4-1-2010

Least Squares Problems with Inequality Constraints as Quadratic Constraints

Jodi Mead

Boise State University

Rosemary A. Renaut

Arizona State University at the Tempe Campus



This is an author-produced, peer-reviewed version of this article. © 2009, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). The final, definitive version of this document can be found online at *Linear Algebra and its Applications*, doi: 10.1016/j.laa.2009.04.017

Least squares problems with inequality constraints as quadratic constraints

Jodi L. Mead* Rosemary A Renaut †

April 7, 2009

Abstract

Linear least squares problems with box constraints are commonly solved to find model parameters within bounds based on physical considerations. Common algorithms include Bounded Variable Least Squares (BVLS) and the Matlab function *lsqlin*. Here, the goal is to find solutions to ill-posed inverse problems that lie within box constraints. To do this, we formulate the box constraints as quadratic constraints, and solve the corresponding unconstrained regularized least squares problem. Using box constraints as quadratic constraints is an efficient approach because the optimization problem has a closed form solution.

The effectiveness of the proposed algorithm is investigated through solving three benchmark problems and one from a hydrological application. Results are compared with solutions found by *lsqlin*, and the quadratically constrained formulation is solved using the L-curve, maximum a posteriori estimation (MAP), and the χ^2 regularization method. The χ^2 regularization method with quadratic constraints is the most effective method for solving least squares problems with box constraints.

Linear least squares, Box constraints, Regularization

AMS Classification: 65F22, 93E24, 62F30

1 Introduction

The linear least squares problems discussed here are often used to incorporate observations into mathematical models. For example, least squares formulations are often used to solve inverse problems in imaging and data assimilation from medical and geophysical applications. In many of these applications the variables in the mathematical models are known to lie within prescribed intervals. This leads to a bound constrained least squares problem:

$$\min \|\mathbf{Ax} - \mathbf{b}\|_2^2 \quad \boldsymbol{\alpha} \leq \mathbf{x} \leq \boldsymbol{\beta}, \quad (1)$$

where $\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta} \in R^n$, $\mathbf{A} \in R^{m \times n}$, and $\mathbf{b} \in R^m$. If the matrix \mathbf{A} has full column rank, then this problem has a unique solution for any vector \mathbf{b} [4]. Here we focus on the more general condition in which \mathbf{A} need not have full column rank.

Successful approaches to solving bound-constrained optimization problems for general linear or nonlinear objective functions can be found in [6], [13], [8], [14] and the Matlab® function *fmincon*. Approaches which are specific

*Corresponding Author. Supported by NSF grant EPS 0447689, Boise State University, Department of Mathematics, Boise, ID 83725-1555, USA. Tel: 208426-2432, Fax: 208-426-1354. Email: jmead@boisestate.edu

†Supported by NSF grants DMS 0513214 and DMS 0421846. Arizona State University, Department of Mathematics and Statistics, Tempe, AZ 85287-1804, USA. Tel: 480-965-3795, Fax: 480-965-4160. Email: renaut@asu.edu

to least squares problem are described in [3], [9] and [15] and the Matlab function *lsqlin*. In this work, we implement a novel approach to solving the bound constrained least squares problem by writing the constraints in quadratic form, and solving the corresponding unconstrained least squares problem.

Most methods for solutions of bound-constrained least squares problems of the form (1) can be categorized as active-set or interior point methods. In active-set methods, a sequence of equality constrained problems are solved with efficient solution methods. The equality constrained problem involves those variables x_i which belong to the active set, i.e. those which are known to satisfy the equality constraint [17]. It is difficult to know the active set *a priori* but algorithms for it include Bounded Variable Least Squares (BVLS) given in [20]. These methods can be expensive for large-scale problems, and a popular alternative to them are interior point methods.

Interior point methods use variants of Newton's method to solve the KKT equality conditions for (1). In addition, the search directions are chosen so the inequalities in the KKT conditions are satisfied at each iteration. These methods can have slow convergence, but if high-accuracy solutions are not necessary, they are a good choice for large scale applications [17]. In this work we write the inequality constraints as quadratic constraints and solve the optimization problem with a penalty-type method that is commonly used for equality constrained problems. This formulation is advantageous because the unconstrained quadratic optimization problem corresponding to the constrained one has a known unique solution.

When \mathbf{A} is not full rank, regularized solutions are necessary for both the constrained and unconstrained problem. A popular approach is Tikhonov regularization [21]

$$\min \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda^2 \|\mathbf{L}(\mathbf{x} - \mathbf{x}_0)\|_2^2, \quad (2)$$

where \mathbf{x}_0 is an initial parameter estimate and L is typically chosen to yield approximations to the l th order derivative, $l = 0, 1, 2$. There are different methods for choosing the regularization parameter λ ; the most popular of which include L-curve, Generalized Cross-Validation (GCV) and the Discrepancy principle [5]. In this work, we will use a χ^2 method introduced in [11] and further developed in [12]. The efficient implementation of this χ^2 approach for choosing λ compliments the solution of bound-constrained least squares problem with quadratic constraints.

The rest of the paper is organized as follows. In Section 2 we re-formulate the bound-constrained least squares problem as an unconstrained quadratic optimization problem by writing the box constraints as quadratic constraints. In Section 3 we give numerical results from benchmark problems [5] and from a hydrological application, and in Section 4 we give conclusions.

2 Bound-Constrained Least Squares

2.1 Quadratic Constraints

Here we introduce an approach whereby the bound constrained problem is written with n quadratic inequality constraints, i.e. (1) becomes

$$\min \|\mathbf{Ax} - \mathbf{b}\|_2^2 \quad (3)$$

$$\text{subject to } (x_i - \bar{x}_i)^2 \leq \sigma_i^2 \quad i = 1, \dots, n \quad (4)$$

where $\bar{\mathbf{x}} = [\bar{x}_i; i = 1, \dots, n]^T$ is the midpoint of the interval $[\alpha, \beta]$, i.e. $\bar{\mathbf{x}} = (\beta + \alpha)/2$ and $\sigma = (\beta - \alpha)/2$. The necessary and sufficient KKT conditions for a feasible point \mathbf{x}_* to be a solution of (3) are:

$$(\mathbf{A}^T \mathbf{A} + \boldsymbol{\lambda}_*) \mathbf{x}_* = \boldsymbol{\lambda}_* \bar{\mathbf{x}} + \mathbf{A}^T \mathbf{b} \quad (5)$$

$$(\lambda_*)_i \geq 0 \quad i = 1, \dots, n \quad (6)$$

$$(\lambda_*)_i [\sigma_i^2 - (x_i - \bar{x}_i)^2] = 0 \quad i = 1, \dots, n \quad (7)$$

$$\sigma_i^2 - (x_i - \bar{x}_i)^2 \geq 0 \quad i = 1, \dots, n \quad (8)$$

where $\boldsymbol{\lambda}_* = \text{diag}((\lambda_*)_i)$.

Reformulating the box constraints $\alpha \leq \mathbf{x} \leq \beta$ as quadratic constraints $(x_i - \bar{x}_i)^2 \leq \sigma_i^2, i = 1, \dots, n$ effectively circumscribes an ellipsoid constraint around the original box constraint. In [18] box constraints were reformulated in exactly the same manner, however the optimization problem was not solved with the penalty or weighted approach as is done here, and described in the Section 2.2. Rather, in [18] parameters were found which ensure there is a convex combination of the objective function and the constraints. This ensures the ellipsoid defined by the objective function intersects that defined by the inequality constraints.

Tikhonov regularization can be viewed as a quadratically constrained least squared problem when the constraint (4) replaced with $\|\mathbf{L}(\mathbf{x} - \mathbf{x}_0)\|_2^2 \leq \delta$. The advantage of viewing regularization as a constraint is that the constrained formulation can give the problem physical meaning. In [19] they give the example in image restoration where δ represents the energy of the target image. For a more general set of problems, the authors in [19] successfully find the regularization parameter λ by solving the quadratically constrained least squares problem.

2.2 Penalty or Weighted Approach

We apply the penalty or weighted approach to the quadratic, inequality constrained problem (3)-(4). In this case, a penalty term $\mathbf{x} - \bar{\mathbf{x}}$ is added to the objective function, and multiplied by a matrix that contains the bounds of the inequality constraints, σ_i . This matrix will come from the first KKT condition (5), which is the solution of the least squares problem:

$$\min \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \|\boldsymbol{\lambda}_*^{1/2}(\mathbf{x} - \bar{\mathbf{x}})\|_2^2. \quad (9)$$

We view the inequality constraints as a penalty term by replacing $\boldsymbol{\lambda}_*$ by $\mathbf{C} = \text{diag}((\sigma_i)^2)$. Since the quadratic constraints circumscribe the box constraints, a sequence of problems for decreasing ϵ are solved which effectively decreases the radius of the ellipsoid until the constraints are satisfied, i.e. solve

$$\min \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \|\mathbf{C}_\epsilon^{-1/2}(\mathbf{x} - \bar{\mathbf{x}})\|_2^2, \quad (10)$$

where $\mathbf{C}_\epsilon = \epsilon \mathbf{C}$. Starting with $\epsilon = 1$, the penalty parameter ϵ decreases until the solution of the inequality constrained problem (3) is identified. Since $\epsilon \rightarrow 0$ solves the equality constrained problem $\mathbf{x} = \bar{\mathbf{x}}$, these iterations are guaranteed to converge when \mathbf{A} is full rank.

Algorithm 1 Solve least squares problem with box constraints as quadratic constraints.

Initialization: $\bar{\mathbf{x}} = (\boldsymbol{\beta} + \boldsymbol{\alpha})/2$, $\mathbf{C}_\epsilon^{-1} = \text{diag}((2/(\beta_i - \alpha_i))^2)$, $\epsilon = 1$ 4
 $\mathcal{Z} = \{j : j = 1, \dots, n\}$, $\mathcal{P} = \text{NULL}$
 count=0
Do (until all constraints are satisfied)
 Solve $(\mathbf{A}^T \mathbf{A} + \mathbf{C}_\epsilon^{-1})\mathbf{y} = \mathbf{A}^T(\mathbf{b} - \mathbf{A}\bar{\mathbf{x}})$ for \mathbf{y}
 $\tilde{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{y}$
 if $\alpha_j \leq \tilde{x}_j \leq \beta_j$ $j \in \mathcal{P}$, else $j \in \mathcal{Z}$
 if $\mathcal{Z} = \text{NULL}$, end
 $\epsilon = 1/(1 + \text{count}/10)\epsilon$
 if $j \in \mathcal{Z}$, $(\mathbf{C}_\epsilon^{-1})_{jj} = ((\epsilon \mathbf{C}_\epsilon)^{-1})_{jj}$
End

This algorithm performs poorly because it over-smoothes the solution \mathbf{x} . In particular, if the interval is small, then σ_i is small, and the solution is heavily weighted towards the mean, $\bar{\mathbf{x}}$. This approach to inequality constraints is not recommended unless prior information about the parameters, or a regularization term is included in the optimization as described in Section 2.3.

2.3 Regularization and quadratic constraints

Algorithm 1 is not useful for well-conditioned or full rank matrices \mathbf{A} because it over-smoothes the solution. In addition, for rank deficient or ill-conditioned \mathbf{A} , we may not be able to calculate the least squares solution $\tilde{\mathbf{x}}$ to (10)

$$\tilde{\mathbf{x}} = \bar{\mathbf{x}} + (\mathbf{A}^T \mathbf{A} + \mathbf{C}_\epsilon^{-1})^{-1} \mathbf{A}^T (\mathbf{b} - \mathbf{A}\bar{\mathbf{x}}),$$

when ϵ is near 1 because $(\mathbf{A}^T \mathbf{A} + \mathbf{C}_\epsilon^{-1})$ may not be invertible. Regularization methods can be used to address these issues. The approach to inequality constraints proposed here should be used after a regularized solution is found which does not satisfy the box constraints.

As mentioned in the Introduction, a typical way to regularize a problem is with Tikhonov regularization (2), but any regularization method can be used to implement box constraints as quadratic constraints. Methods such as the *discrepancy principle* [16], L-curve [5], χ^2 regularization [11] and maximum a posteriori estimation (MAP) [1] often weight the least squares problem with the inverse covariance matrix for the errors in the data, \mathbf{C}_b . In addition the χ^2 method and MAP estimation weight the regularization term with the inverse covariance matrix on the mean zero initial parameter estimate, \mathbf{C}_x , i.e. from (2) $\lambda \mathbf{L} = \mathbf{C}_x^{-1/2}$, in which case we solve

$$\min_{\mathbf{x}} \mathcal{J}(\mathbf{x})$$

where

$$\mathcal{J}(\mathbf{x}) = \|\mathbf{C}_b^{-1/2}(\mathbf{A}\mathbf{x} - \mathbf{b})\|_2^2 + \|\mathbf{C}_x^{-1/2}(\mathbf{x} - \mathbf{x}_0)\|_2^2.$$

Applying quadratic constraints to the regularized functional amounts to solving the following problem:

$$\min_{\mathbf{x}} \mathcal{J}_\epsilon(\mathbf{x}) \tag{11}$$

where

$$\mathcal{J}_\epsilon(\mathbf{x}) = \|\mathbf{C}_b^{-1/2}(\mathbf{A}\mathbf{x} - \mathbf{b})\|_2^2 + \|\mathbf{C}_x^{-1/2}(\mathbf{x} - \mathbf{x}_0)\|_2^2 + \|\mathbf{C}_\epsilon^{-1/2}(\mathbf{x} - \bar{\mathbf{x}})\|_2^2.$$

This formulation has three terms in the objective function and seeks a solution which lies in an intersection of three ellipsoids. It is possible, but not necessary, to write the regularization term and the inequality constrained term as a single term. The solution, or minimum value of $\mathcal{J}_\epsilon(\mathbf{x})$ occurs at

$$\tilde{\mathbf{x}}_\epsilon = \mathbf{x}_0 + (\mathbf{A}^T \mathbf{C}_b^{-1} \mathbf{A} + \mathbf{C}_x^{-1} + \mathbf{C}_\epsilon^{-1})^{-1} (\mathbf{A}^T \mathbf{C}_b^{-1} \mathbf{r} + \mathbf{C}_\epsilon^{-1} \Delta \mathbf{x}), \quad (12)$$

where $\Delta \mathbf{x} = \bar{\mathbf{x}} - \mathbf{x}_0$ and $\mathbf{r} = \mathbf{b} - \mathbf{A} \mathbf{x}_0$.

In order for the solution (12) to exist, $\mathbf{D} \mathbf{x} = \mathbf{f}$ must have a solution where

$$\mathbf{D} = \begin{bmatrix} \mathbf{C}_x^{-1/2} \\ \mathbf{C}_\epsilon^{-1/2} \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} \mathbf{C}_x^{-1/2} \mathbf{x}_0 \\ \mathbf{C}_\epsilon^{-1/2} \bar{\mathbf{x}} \end{bmatrix},$$

i.e. \mathbf{f} must be in the range of \mathbf{D} . If there is no such solution the two ellipsoids defined by the last two terms in $\mathcal{J}_\epsilon(\mathbf{x})$ do not intersect and we cannot find a solution that lies within the constraints for the given \mathbf{C}_x .

Algorithm 2 given below takes as inputs \mathbf{C}_x and \mathbf{C}_b which result in regularized solutions that may or may not lie within the box constraints. The output from the algorithm is $\tilde{\mathbf{x}}_\epsilon$ defined by (12), that satisfies the box constraints.

Algorithm 2 Solve regularized least squares problem with box constraints as quadratic constraints.

Initialization: $\bar{\mathbf{x}} = (\boldsymbol{\beta} + \boldsymbol{\alpha})/2, \mathbf{C}_\epsilon^{-1} = \text{diag}((2/(\beta_i - \alpha_i))^2), \epsilon = 1$
 $\mathcal{Z} = \{j : j = 1, \dots, n\}, \mathcal{P} = \text{NULL}$
 count=0
Do (until all constraints are satisfied)
 Solve $(\mathbf{A}^T \mathbf{C}_b^{-1} \mathbf{A} + \mathbf{C}_\epsilon^{-1} + \mathbf{C}_x^{-1}) \mathbf{y}_\epsilon = (\mathbf{A}^T \mathbf{C}_b^{-1} \mathbf{r} + \mathbf{C}_\epsilon^{-1} \Delta \mathbf{x})$ for \mathbf{y}_ϵ
 $\tilde{\mathbf{x}}_\epsilon = \mathbf{x}_0 + \mathbf{y}_\epsilon$
 if $\alpha_j \leq \tilde{x}_j \leq \beta_j, j \in \mathcal{P}$, else $j \in \mathcal{Z}$
 if $\mathcal{Z} = \text{NULL}$, end
 $\epsilon = 1/(1 + \text{count}/10)\epsilon$
 if $j \in \mathcal{Z}, (\mathbf{C}_\epsilon^{-1})_{jj} = ((\epsilon \mathbf{C}_\epsilon)^{-1})_{jj}$
 count = count +1
End

The iterations in Algorithm 2 reduce the penalty parameter until the constraints are satisfied. Illustrative results of the performance of this algorithm for ill-posed problems, as compared to other standard methods, are given in Section 3.

2.4 Regularization methods

Algorithm 2 requires the weight on the initial parameter misfit \mathbf{C}_x as an input. In this section we describe three different methods for the calculation of it: the L-curve, χ^2 regularization and maximum a posteriori estimation (MAP).

The L-curve approach finds the parameter λ in (2), for specified \mathbf{L} . This is done by solving (2) multiple times with various λ to get multiple solutions \mathbf{x}_λ . Once these solutions are obtained, a log-log plot of $\|\mathbf{L}(\mathbf{x}_\lambda - \mathbf{x}_0)\|_2^2$ versus $\|\mathbf{b} - \mathbf{A} \mathbf{x}_\lambda\|_2^2$ will typically be in the shape of an L, and the optimal value of λ is the one at the corner. The parameter values resulting from this choice of λ are optimal in the sense that the error in the weighted parameter misfit and data misfit are balanced. For the purposes of Algorithm 2, the L-curve method finds λ , for specified \mathbf{L} with $\mathbf{C}_x = \lambda^{-2}(\mathbf{L}^T \mathbf{L})^{-1}$, and has the potential to include random noise in the data when \mathbf{C}_b is specified.

MAP estimation and the χ^2 regularization method are two which not only assume that the data contain noise, but so do the initial parameter estimates. The MAP estimate assumes the data \mathbf{b} are random, independent and identically

distributed, and follow a normal distribution with probability density function

6

$$\rho(\mathbf{b}) = \text{const} \times \exp \left\{ -\frac{1}{2}(\mathbf{b} - \mathbf{Ax})^T \mathbf{C}_b^{-1}(\mathbf{b} - \mathbf{Ax}) \right\}, \quad (13)$$

with \mathbf{Ax}_0 the expected value of \mathbf{b} and \mathbf{C}_b the corresponding covariance matrix. In addition, it is assumed that the parameter values \mathbf{x} are also random following a normal distribution with probability density function

$$\rho(\mathbf{x}) = \text{const} \times \exp \left\{ -\frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \mathbf{C}_x^{-1}(\mathbf{x} - \mathbf{x}_0) \right\}, \quad (14)$$

with \mathbf{x}_0 the expected value of \mathbf{x} and \mathbf{C}_x the corresponding covariance matrix.

In order to maximize the probability that the data were in fact observed we find \mathbf{x} where the probability density is maximum. The maximum a posteriori estimate of the parameters occurs when the joint probability density function is maximum [1], i.e. optimal parameter values are found by solving

$$\min_{\mathbf{x}} \{ (\mathbf{b} - \mathbf{Ax})^T \mathbf{C}_b^{-1}(\mathbf{b} - \mathbf{Ax}) + (\mathbf{x} - \mathbf{x}_0)^T \mathbf{C}_x^{-1}(\mathbf{x} - \mathbf{x}_0) \}. \quad (15)$$

This optimal parameter estimate is found under the assumption that the data and parameters follow a normal distribution and are independent and identically distributed. The χ^2 regularization method is based on this idea, but the assumptions are relaxed, see [11] [12]. Since these assumptions are typically not true, we do not expect the MAP or χ^2 estimate to give us the exact parameter values.

The estimation procedure behind the χ^2 regularization method is equivalent to that for MAP estimation. However the χ^2 regularization method is an approach for finding \mathbf{C}_x (\mathbf{C}_b) given \mathbf{C}_b (\mathbf{C}_x), while MAP estimation simply takes them as inputs. This χ^2 method is based on the fact that the minimum value of the functional $\mathcal{J}(\tilde{\mathbf{x}})$ is a random variable which follows a χ^2 distribution with m degrees of freedom [2, 11]. In particular, given values for \mathbf{C}_b and \mathbf{C}_x , the difference $|J(\tilde{\mathbf{x}}) - m|$ is an estimate of confidence that \mathbf{C}_b and \mathbf{C}_x are accurate weighting matrices. Mead [11] noted these observations, and suggested a matrix \mathbf{C}_x can be found by requiring that $\mathcal{J}(\tilde{\mathbf{x}})$, to within a specified $(1 - \alpha)$ confidence interval, is a χ^2 random variable with m degrees of freedom, namely such that

$$m - \sqrt{2m}z_{\alpha/2} < \mathbf{r}^T (\mathbf{AC}_x \mathbf{A}^T + \mathbf{C}_b)^{-1} \mathbf{r} < m + \sqrt{2m}z_{\alpha/2}, \quad (16)$$

where $\mathbf{r} = \mathbf{b} - \mathbf{Ax}_0$ and $z_{\alpha/2}$ is the relevant z -value for a χ^2 -distribution with m degrees of freedom. In [12] it was shown that for accurate \mathbf{C}_b , this χ^2 approach is more efficient and gives better results than the discrepancy principle, the L-curve and generalized cross validation (GCV) [5].

In the numerical results in Section 3, MAP is implemented only for the benchmark problems where the true, or mean, parameter values are known. In the benchmark problems \mathbf{x}_0 is randomly generated with error covariance \mathbf{C}_x , just as the data are generated with error covariance \mathbf{C}_b . The MAP estimate uses the exact value for \mathbf{C}_x , while the χ^2 method finds $\mathbf{C}_x = \sigma_x^2 \mathbf{I}$ by solving (16), thus the MAP estimate is the ‘‘exact’’ solution for the χ^2 regularized estimate but cannot be used in practice when \mathbf{C}_x is unknown.

Note that the L-curve is similar to the MAP and χ^2 estimates when $\mathbf{C}_x = \lambda^{-2}(\mathbf{L}^T \mathbf{L})$. The advantage of the MAP and χ^2 estimates is when \mathbf{C}_x is not a constant matrix and hence the weights on the parameter misfits vary. Moreover,

when \mathbf{C}_x has off diagonal elements, correlation in initial parameter estimate errors can be modeled. The disadvantage⁷ of MAP is that *a priori* information is needed. On the other hand the χ^2 regularization method is an approach for finding elements of \mathbf{C}_x , thus matrices rather than parameters may be used for regularization, but not as much *a priori* information is needed as with MAP. However, in the Section 3 the χ^2 methods uses $\mathbf{C}_x = \sigma_x^2 \mathbf{I}$. Future work involves developing efficient algorithms for more dense \mathbf{C}_x .

In Section 3 we give numerical results where the box constrained least squares problem (1) is solved by (11), i.e. by implementing the box constraints as quadratic constraints using *Algorithm 2*.

3 Numerical Results

3.1 Benchmark Problems

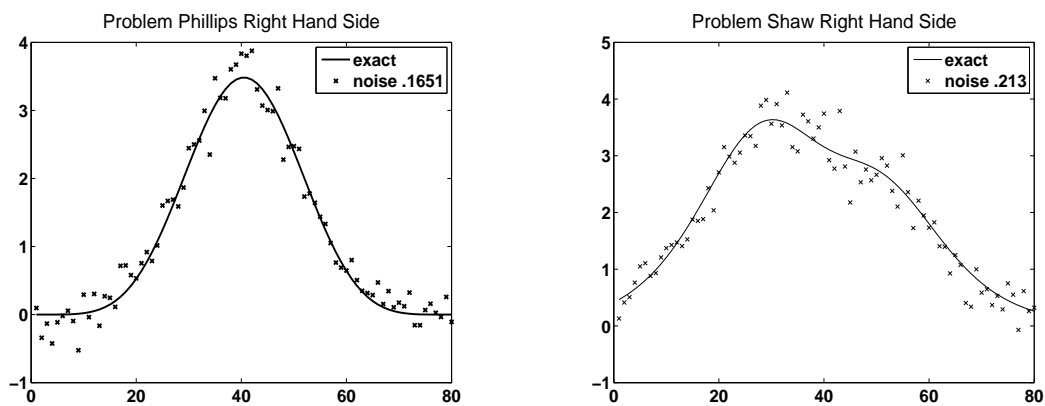
We present a series of representative results from *Algorithm 2* using benchmark cases from [5]. *Algorithm 2* was implemented with the χ^2 regularization method, the L-curve and maximum a posteriori estimation (MAP), and compared with results from the Matlab constrained least squares function *lsqlin*. In particular, system matrices \mathbf{A} , right hand side data \mathbf{b} and solutions \mathbf{x} are obtained from the following test problems: *phillips*, *shaw*, and *wing*. These benchmark problems do not have physical constraints, so we set them arbitrarily as follows: *phillips* ($0.2 < x < 0.6$), *shaw* ($0.5 < x < 1.5$), and *wing* ($0 < x < 0.1$). In all cases, the parameter estimate from *Algorithm 2* is essentially found by (12).

In all cases we generate a random matrix Θ of size $m \times 500$, with columns Θ^c , $c = 1 : 500$, using the Matlab function *randn*. Then setting $\mathbf{b}^c = \mathbf{b} + \text{level} \|\mathbf{b}\|_2 \Theta^c / \|\Theta^c\|_2$, for $c = 1 : 500$, generates 500 copies of the right hand vector \mathbf{b} with normally distributed noise, dependent on the chosen level. Results are presented for level = .1. An example of the error distribution for all cases with $n = 80$ is illustrated in Figure 1. Because the noise depends on the right hand side \mathbf{b} the actual error, as measured by the mean of $\|\mathbf{b} - \mathbf{b}^c\|_\infty / \|\mathbf{b}\|_\infty$ over all c , varies between 0.1651 and 0.2505, and is given for each test problem in Figure 1.

The covariance \mathbf{C}_b between the measured components is calculated directly for the entire data set \mathbf{B} with rows $(\mathbf{b}^c)^T$. Because of the design, \mathbf{C}_b is close to diagonal, $\mathbf{C}_b \approx \text{diag}(\sigma_{b_i}^2)$ and the noise is colored. In all experiments, regardless of parameter selection method, the same covariance matrix \mathbf{C}_b is used. The MAP estimate requires an additional input of \mathbf{C}_x , which is computed in a manner similar to \mathbf{C}_b . The χ^2 method finds \mathbf{C}_x by solving (16) for $\mathbf{C}_x = \sigma_x^2 \mathbf{I}$, while the parameter λ found by the L-curve is used to form $\mathbf{C}_x = \lambda^{-2} \mathbf{I}$. Finally, the matrix \mathbf{C}_ϵ implements the box constraints as quadratic constraints and is the same for all three regularizations methods, with

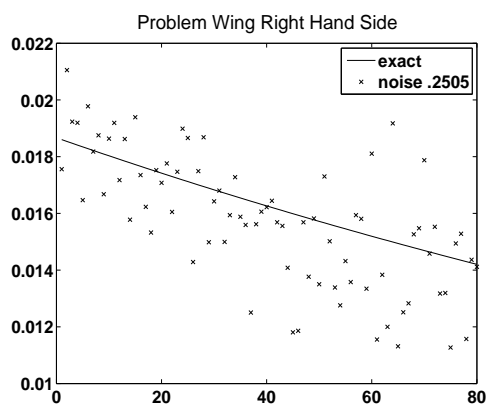
$$\begin{aligned} \mathbf{C}_\epsilon &= \epsilon \mathbf{C} \\ &= \text{diag}(\sigma_i^2), \quad \sigma_i = (\beta_i - \alpha_i)/2 \quad x_i \in [\alpha_i, \beta_i]. \end{aligned}$$

The *a priori* reference solution \mathbf{x}_0 is generated using the exact known solution and noise added with level = .1 in the same way as for modifying \mathbf{b} . The same reference solution \mathbf{x}_0 is used for all right hand side vectors \mathbf{b}^c , see Figure 2.



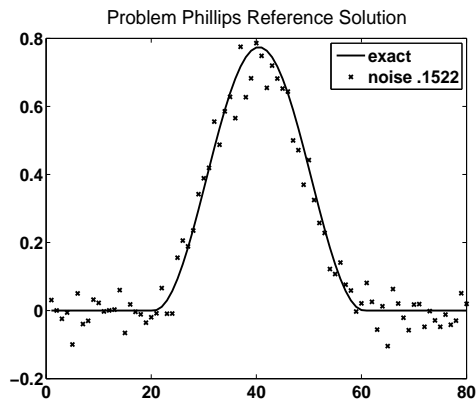
(a)

(b)

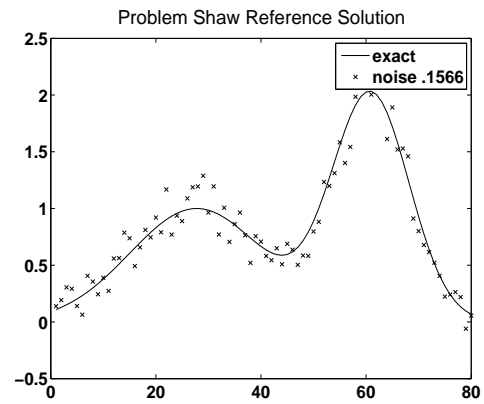


(c)

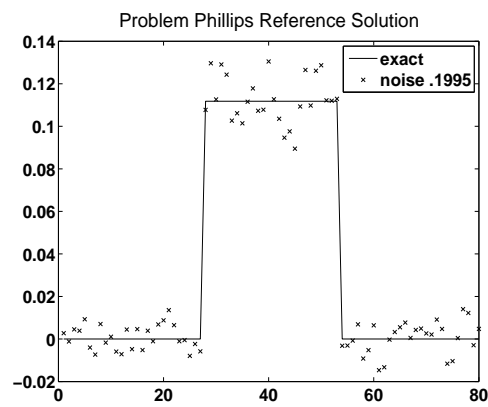
Figure 1: Illustration of the noise in the right hand side for problem (a) *phillips* , (b) *shaw* (c) *wing*.



(a)



(b)



(c)

Figure 2: Illustration of the reference solution x_0 for problem (a) *phillips*, (b) *shaw* (c) *wing*.

The unconstrained and constrained solutions to the *phillips* test problem are given in Figure 3. For the unconstrained solution, the L-curve gives the worst solution, while the MAP and χ^2 estimates are similar. The MAP estimate is an exact version of the χ^2 regularization method because the exact covariance matrix C_x is given. The χ^2 regularization method finds C_x , using the properties of a χ^2 distribution, thus it requires less a priori knowledge.

The methods used in the unconstrained case were implemented with quadratic constraints in Figure 3(b). For comparison, the Matlab function *lsqlin* was used to implement the box constraints in the linear least squares problem. We see here that the *lsqlin* solution stays within the correct constraints, but does not retain the shape of the curve. This is true for all test problems, also see Figures 5(b) and 7(b). Figures 4, 6 and 8 show that, for any regularization method, the significant advantage of implementing the box constraints as quadratic constraints and solving (11) is that we retain the shape of the curve.

The constrained and unconstrained solutions to the *phillips* test problem for each method are given in Figure 4. The quadratic constraints correctly enforce the box constraints in all cases, regardless of the accuracy of the unconstrained solutions. In fact, the poor results from the L-curve are improved with the constraints. However, this is not necessarily true for the *shaw* test problem in Figure 5. Again the L-curve gives the poorest results in the unconstrained case, while in the constrained case it does not retain the correct shape of the curve. The constrained L-curve is still preferable over the results from *lsqlin*, as shown in Figure 5(b).

For all three test problems, in the constrained and unconstrained cases, Figures 4(b)(c), 6(b)(c) and 8(b)(c) each show that the χ^2 estimate gives results as good as the MAP estimate. The χ^2 estimate does not require any a priori information about the parameters, thus it is a significant improvement over the MAP estimate.

The *wing* test problem in Figures 7-8 has a discontinuous solution. Least squares solutions typically do poorly in these instances because they smooth the solution. The L-curve does perform poorly, and is not improved upon by implementing the constraints. Matlab's *lsqlin* is also not able to capture the the discontinuous solution. However, both the MAP and χ^2 estimates were able to capture the discontinuity in the constrained and unconstrained cases.

3.2 Estimating data error: Example from Hydrology

In addition to the benchmark results, we present the results for a real model from hydrology. The goal is to obtain four parameters $\mathbf{x}_0 = [\theta_r, \theta_s, \alpha, n]$ in an empirical equation developed by van Genuchten [22] which describes soil moisture as a function of hydraulic pressure head. A complete description of this application is given in [12]. Hundreds of soil moisture content and pressure head measurements are made at multiple soil pits in the Dry Creek catchment near Boise, Idaho [10], and these are used to obtain \mathbf{b} . We rely on the laboratory measurements for good first estimates of the parameters \mathbf{x}_0 , and their standard deviations σ_{x_i} . It takes 2-3 weeks to obtain one set of laboratory measurements, but this procedure is done multiple times from which we obtain standard deviation estimates and form $C_{\mathbf{x}} = \text{diag}(\sigma_{\theta_r}^2, \sigma_{\theta_s}^2, \sigma_{\alpha}^2, \sigma_n^2)$. These standard deviations account for measurement technique or error. However, measurements on this core may not accurately reflect soils in entire watershed region. We will show results from two soil pits: NU10_15 and SU5_15. They represent pits upstream from a weir 10 and 5 meters, respectively, both 15 meters from the surface.

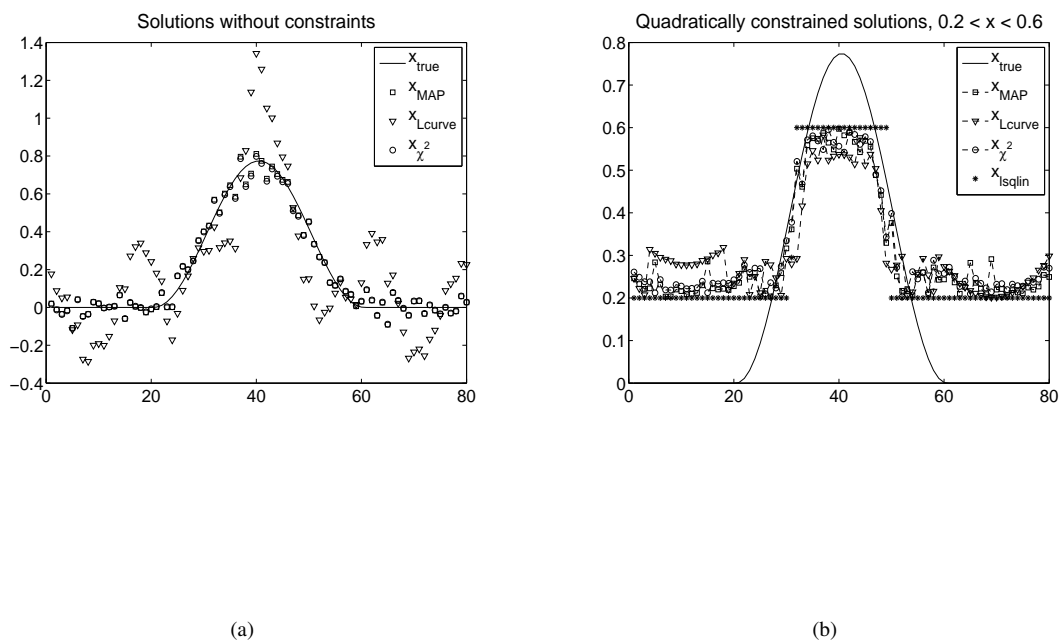
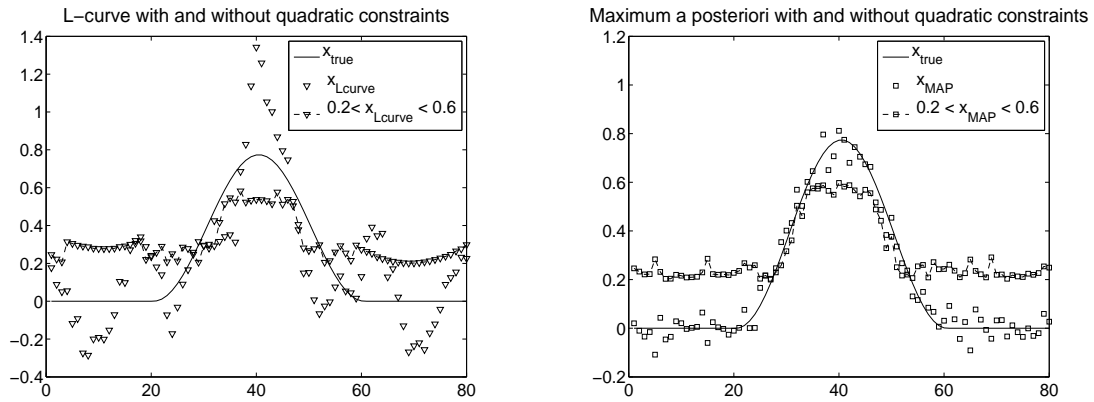
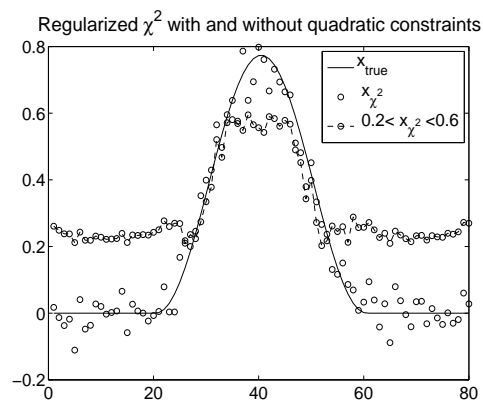


Figure 3: Phillips (a) unconstrained and (b) constrained solutions



(a)

(b)



(c)

Figure 4: Phillips test problem of (a) L-curve , (b) maximum a posteriori estimation (c) regularized χ^2 method.

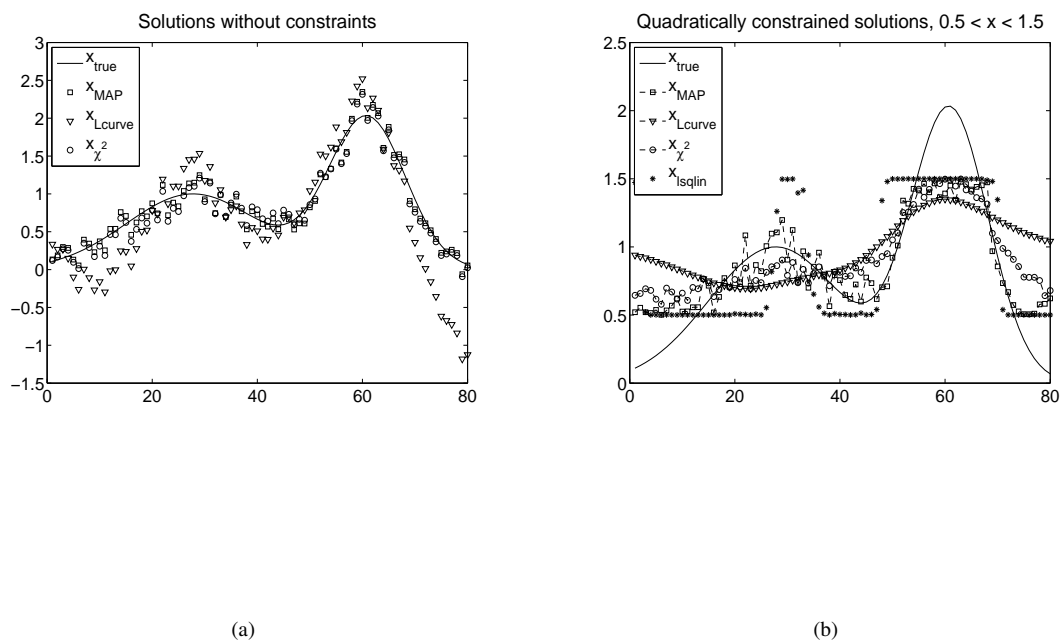
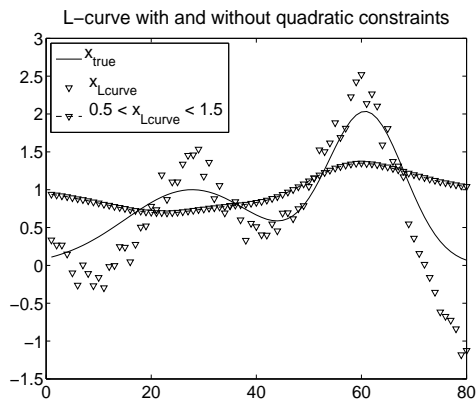
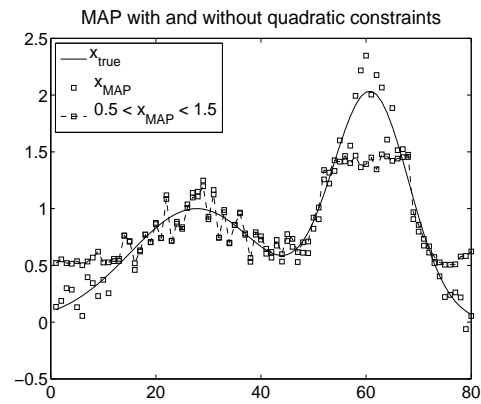


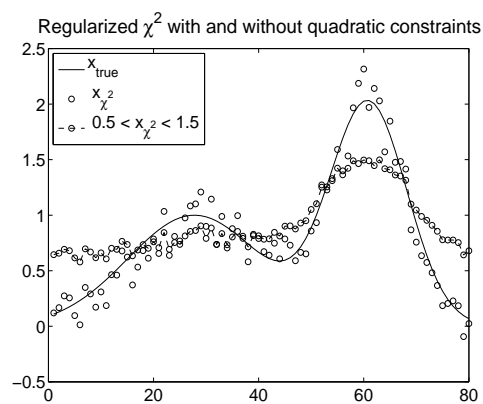
Figure 5: Shaw (a) unconstrained and (b) constrained solutions



(a)



(b)



(c)

Figure 6: Shaw test problem of (a) L-curve , (b) maximum a posteriori estimation (c) regularized χ^2 method.

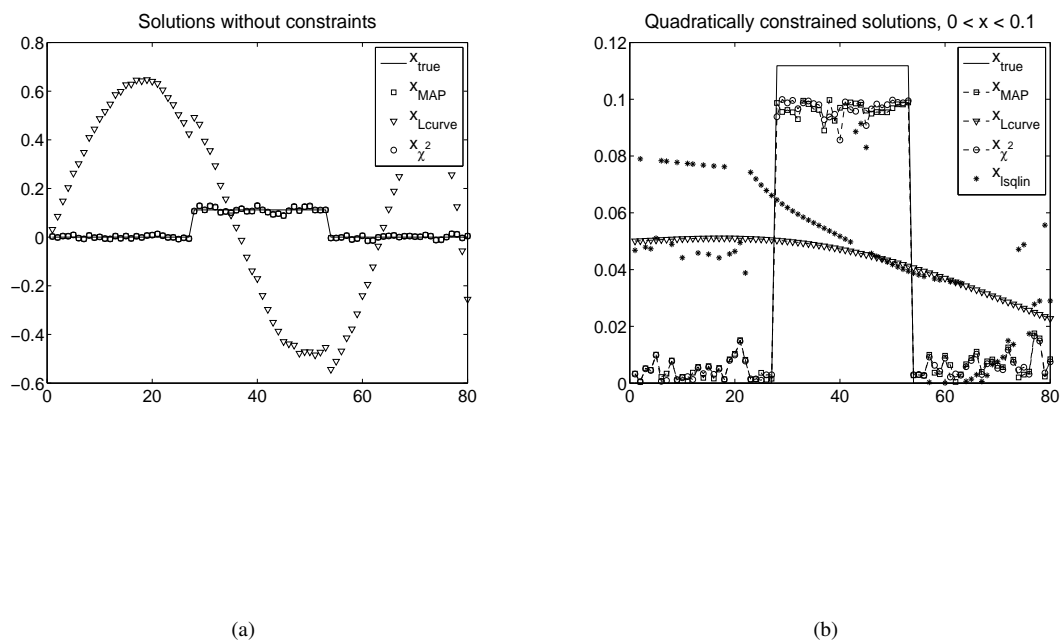


Figure 7: Wing (a) unconstrained and (b) constrained solutions

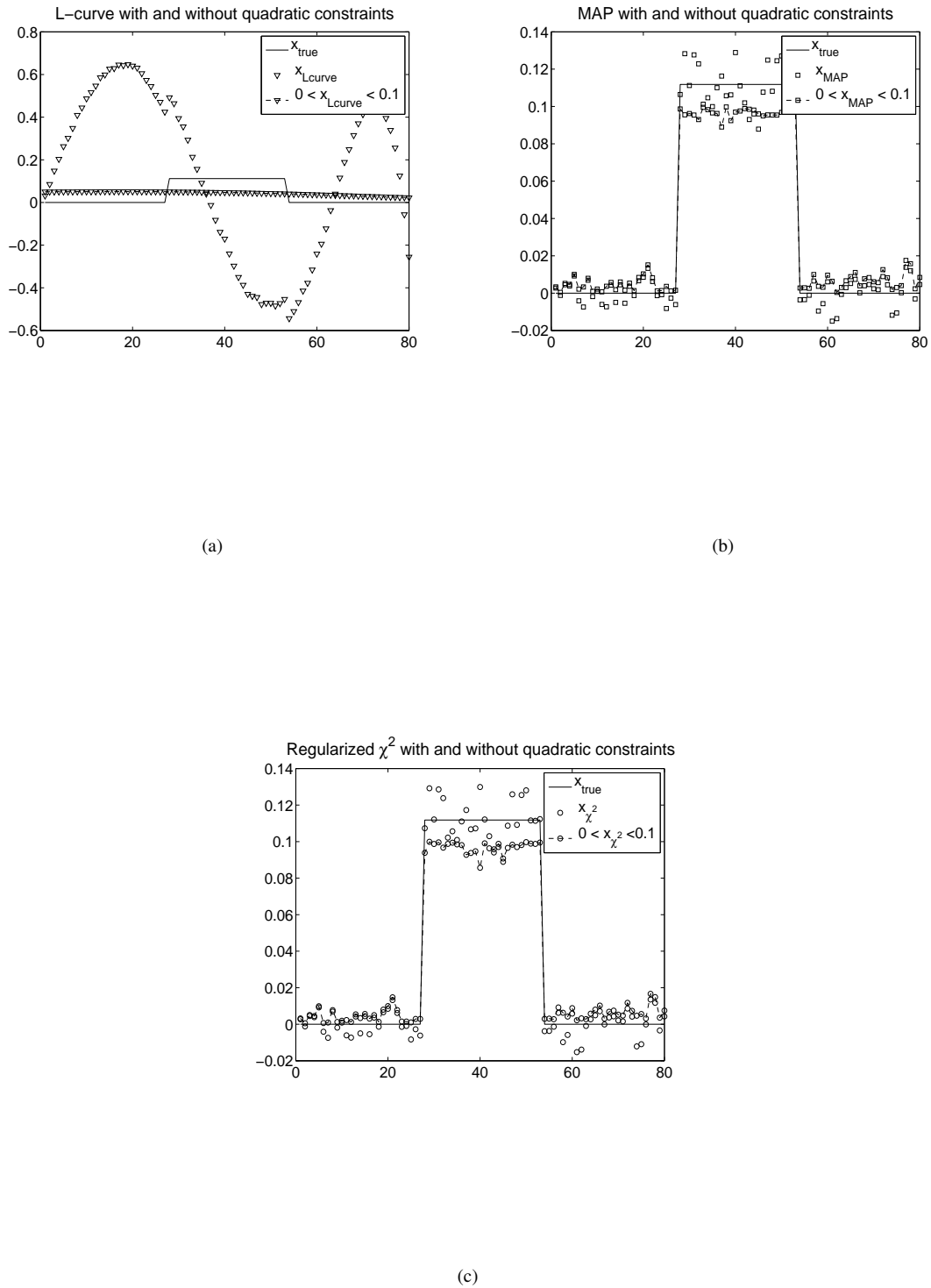


Figure 8: Wing test problem of (a) L-curve , (b) maximum a posteriori estimation (c) regularized χ^2 method.

These parameters depend on the soil type: sand, silt, clay, loam and combinations of them. Extensive studies have been done to determine the parameter values based on soil type. These values can be found in [23]. In particular, lower and upper bounds have been given for each soil type, thus each parameter is assumed to lie within prescribed intervals. The second column of Table 1 gives parameter ranges (or constraints) in the form of soil class averages found in [23]. These ranges are used to form \mathbf{C}_ϵ .

This is a severely overdetermined problem, and we used constrained least squares (1) to find the best parameters \mathbf{x} . The matrix \mathbf{A} is given by van Genuchten's equation, while the data \mathbf{b} are the field measurements described above. The box constraints in Table 1 were implemented as quadratic constraints with the penalty approach, and are used to form \mathbf{C}_ϵ .

Since initial parameter estimates \mathbf{x}_0 and covariance \mathbf{C}_x is found by repeated measurements in the laboratory, the χ^2 method is used to find the standard deviation σ_b on field measurements \mathbf{b} , and form $\mathbf{C}_b = \sigma_b^2 \mathbf{I}$. In other words, the regularization parameter or initial parameter misfit weight is taken from laboratory measurements while the data weight is obtained by the χ^2 method.

Table 1 gives parameter values for both pits, in both the constrained and unconstrained cases. For both pits, the only unconstrained parameter that did not fit into the appropriate range is θ_s . The constrained parameters did fit into the ranges given by [23]. However, after further investigation, we began to question the validity of these ranges. The parameter θ_s represents soil moisture when the ground is nearly saturated. In the semi-arid environment of the Dry Creek Watershed, the soil does not typically come near saturation. The fact that Algorithm 2 correctly implemented the constraints showed us that that the soil class averages, with a minimum value of $\theta_s = 0.3010$, do not reflect the soils found in this region. A more realistic minimum would be $\theta_s = 0.2$.

Parameter	Ranges	NU10_15		SU5_15	
		Unconstrained	Constrained	Unconstrained	Constrained
$\log_{10} \alpha$	$[-2.86, -0.9060]$	-1.6109	-0.9567	-2.0109	-1.0978
$\log_{10} n$	$[0.004, 0.6820]$	0.1732	0.2182	0.5239	0.1182
θ_s	$[0.3010, 0.5680]$	0.2271	0.3522	0.2222	0.3409
θ_r	$[-0.0150, 0.2310]$	-0.0080	0.0493	0.1032	-0.0109

Table 1: Hydrological Parameters

Figure 9 shows the constrained and unconstrained results with θ representing soil moisture on the horizontal axis, and ψ representing pressure head on the vertical. The van Genuchten equation is typically plotted in this manner, and the curve is called the *soil moisture retention curve*. Near saturation, i.e. for $|\psi|$ near 0, the soil moisture falls below 0.25 further indicating the the soil class averages found in [23] are not appropriate for this region and should not be used as constraints.

4 Conclusions

In this paper we introduced an implementation of box constraints as quadratic constraints for linear least squares problems. Because the original least squares problems may be ill-conditioned, the quadratic constraints are added to

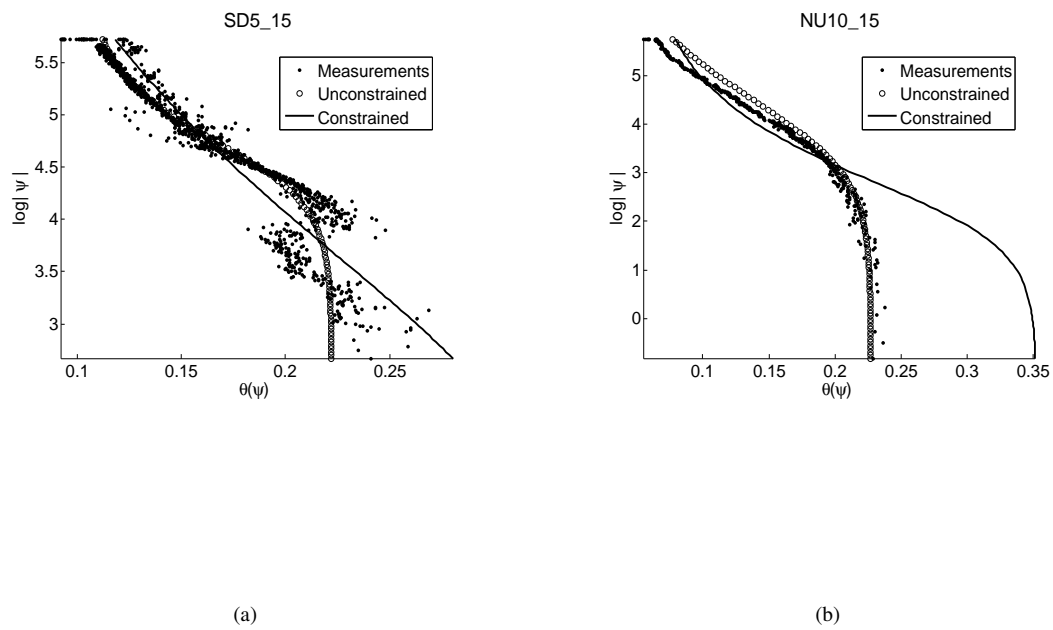


Figure 9: Unconstrained and constrained soil moisture retention curves for (a) SD5_15 and (b) NU10_15 .

the objective function of the regularized problem. Thus there is a unique solution to the problem for any choice of ϵ used in the iteration to the solution of the box-constrained problem. The quadratic constraints circumscribe an ellipsoid around the box constraints, and the radius of the ellipsoid is iteratively reduced until the constraints are satisfied.

The quadratic constraint approach was used with regularization via the L-curve, maximum a posteriori estimation (MAP) and the χ^2 method [11], [12]. Constrained results were compared to those found by the Matlab function *lsqlin*. Results from *lsqlin* stayed within the constraints, but did not maintain the correct shape of the parameter solution curve. The L-curve gave the poorest unconstrained results, which were sometimes improved upon by implementing constraints. The MAP and χ^2 estimates gave the best results but the MAP estimate requires a priori information about the parameters which is typically not available. Thus the method of choice for constrained least squares problems is χ^2 regularization method with box constraints implemented as quadratic constraints. This approach was also used to solve a problem in Hydrology.

The quadratic constraint approach can be implemented with any regularized least squares method with box constraints. It is simple to implement and is preferred over the Matlab function *lsqlin* because the constrained solution keeps the shape of the unconstrained solution, while the *lsqlin* solution merely stays at the bounds of the constraints.

Acknowledgements

Professor Jim McNamara, Boise State University, Department of Geosciences and Professor Molly Gribb, Boise State University, Department of Civil Engineering supplied the field and laboratory data, respectively, for the Hydrological example.

References

- [1] Aster, R.C., Borchers, B. and Thurber, C., 2005, *Parameter Estimation and Inverse Problems*, Academic Press, p 301.
- [2] Bennett, A., 2005 *Inverse Modeling of the Ocean and Atmosphere* (Cambridge University Press) p 234.
- [3] Bierlaire, M. , Toint, Ph.L. and Tuytens, D., 1991, *On iterative algorithms for linear least squares problems with bound constraints*, Lin. Alg. Appl., Vol. 143 **1**, 111-143.
- [4] Björck, Å, 1996, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, p 408.
- [5] Hansen, P. C., 1994, *Regularization Tools: A Matlab Package for Analysis and Solution of Discrete Ill-posed Problems*, Numerical Algorithms **6**, 1-35.
- [6] Huyer, W. and Neumaier, A., 1999, *Global optimization by multilevel coordinate search*, *J. Global Optimization* **14** , 331-355.
- [7] Lawson, C.L. and Hanson, R.J., 1974, *Solving Least Squares Problems*, Prentice-Hall, p 340.

- [8] Lin, C-J and Moreé, J., 1999, *Newton's method for large bound-constrained optimization problems*, SIAM Journal on Optimization, Vol. 9, **4**, 1100-1127.
- [9] Lötstedt, P., 1984, *Solving the minimal least squares problem subject to bounds on the variables*, BIT **24**, 206-224.
- [10] McNamara, J. P., Chandler, D. G., Seyfried, M., and Achet, S. 2005, *Soil moisture states, lateral flow, and streamflow generation in a semi-arid, snowmelt-driven catchment*, Hydrological Processes, **19**, 4023-4038.
- [11] Mead J.L., 2008, *Parameter estimation: A new approach to weighting a priori information*, J. Inv. Ill-posed Problems, **16**, 2, 175-194.
- [12] Mead J.L. and Renaut, R.A., *A Newton root-finding algorithm for estimating the regularization parameter for solving ill-conditioned least squares problems*, submitted to Inverse Problems.
- [13] Michalewicz, Z. and Janikow, C. Z., 1996, *GENOCOP: a genetic algorithm for numerical optimization problems with linear constraints*, Comm. ACM, Volume 39 , Issue 12es, Article No. 175.
- [14] Mockus J., 1989, *Bayesian Approach to Global Optimization*, Kluwer Academic Publishers , Dordrecht.
- [15] Morigi, S, Reichel, L., Sgallari, F., and Zama, F., 2007, *An iterative method for linear discrete ill-posed problems with box constraints*, J. Comp. Appl. Math. **198**, 505-520.
- [16] Morozov, V.A. , 1966, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl. 7, 414-417.
- [17] Nocedal, J. and Wright S, 1999, *Numerical Optimization*, Springer-Verlag, New York, p 636.
- [18] Pierce, J.E. and Rust, B.W. 1985, *Constrained Least Squares Interval Estimation*, SIAM Sci. Stat. Comput., Vol. 6, No. 3, 670-683.
- [19] Rojas M. and Sorensen, D.C., 2000, *A Trust-Region Approach to the Regularization of Large-Scale Discrete Forms of Ill-Posed Problems*, SISC , Vol. 23, No. 6, 1842-1860.
- [20] Stark P.B. and Parker, R.L., 1995, *Bounded-Variable Least-Squares: An Algorithm and Applications*, Computational Statistics, 10:129-141.
- [21] Tikhonov, A.N., 1963 *Regularization of incorrectly posed problems*, Soviet Math., 4, 1624-1627.
- [22] van Genuchten, M.Th. 1980, *A closed-form equation for predicting the hydraulic conductivity of unsaturated soils*, Soil Sci. Soc. Am. J. 44:892-898.
- [23] Warrick, A.W., 2003, *Soil Water Dynamics*, Oxford University Press, p 391.