PIECEWISE DIFFERENTIABLE MINIMIZATION FOR ILL-POSED INVERSE PROBLEMS

YUYING LI
DEPARTMENT OF COMPUTER SCIENCE
CORNELL UNIVERSITY*

CTC96TR252
8/96

Abstract. Based on minimizing a piecewise differentiable l_p function subject to a single inequality constraint, this paper discusses algorithms for a discretized regularization problem for ill-posed inverse problems. We examine computational challenges of solving this regularization problem. Possible minimization algorithms such as the steepest descent method, iteratively weighted least squares (IRLS) method and a recent globally convergent affine scaling Newton approach are considered. Limitations and efficiency of these algorithms are demonstrated using the geophysical traveltime tomographic inversion and image restoration applications.

Key words. inverse problem, tomographic inversion, image restoration, total variation, steepest descent, projected gradient, IRLS, affine scaling and Newton.

1. Minimization and Ill-posed Inverse Problems. Minimization algorithms have long been used in regulating an ill-posed inverse problem. Assuming that a desired property of a solution is known *a priori*, an ill-posed inverse problem can be regulated by solving a constrained minimization problem. In particular, properties expressed in nondifferentiable form have increasingly been found more appropriate in many applications. Discretization of such a regularization problem often leads to minimizing a large-scale piecewise differentiable function with a single constraint. In this paper, we consider regularization using piecewise differentiable minimization, possibly with a single inequality constraint.

Consider an ill-posed inverse problem,

$$\mathcal{A}u = f, \tag{1.1}$$

where $\mathcal{A}$ is an operator in a Hilbert space. Assume that $||\cdot||_2$ denotes the Euclidean norm and an *a priori* condition (e.g., continuity and boundedness) of the desired solution is given by $||\mathcal{B}u||_2 \le \rho$ for some linear operator

* This paper is written for the proceedings of the workshop *Large-Scale Optimization, with Applications to Inverse Problems, Optimal Control and Design, & Molecular and Structural Optimization* in IMA, Minnesota, July 1995. Research is partially supported by the Applied Mathematical Sciences Research Program (KC-04-02) of the Office of Energy Research of the U.S. Department of Energy under grant DE-FG02-90ER25013.A000, and in part by NSF through grant DMS-9505155 and ONR through grant N00014-96-1-0050, and by the Cornell Theory Center which receives major funding from the National Science Foundation and IBM corporation, with additional support from New York State and members of its Corporate Research Institute.

$\mathcal{B}$ and $\rho > 0$. Depending on applications, a reasonable solution to (1.1) can be computed by solving either

$$\min_{u \in U} \|\mathcal{A}u - f\|_2 \quad \text{subject to} \quad \|\mathcal{B}u\|_2 \le \rho, \tag{1.2}$$

or

$$\min_{u \in U} \|\mathcal{B}u\|_2 \quad \text{subject to} \quad \|\mathcal{A}u - f\|_2 \le \delta. \tag{1.3}$$

The well-known constrained least squares approach, e.g., [14] for image restoration, corresponds to solving either (1.2) or (1.3).

Increasingly, properties involving a nonsmooth piecewise differentiable measurement have been found more appropriate in many applications [20,22,21,9,1,23,15,18,26]. We consider image restoration, for example. Let $u(x,y)$ denote the intensity of an image in a region Ω. Osher et al propose [20,22,21] to minimize total variation $\int_\Omega \sqrt{u_x^2 + u_y^2}\; dxdy$ to achieve image restoration:

$$\begin{aligned} &\min_u \int_\Omega \sqrt{u_x^2 + u_y^2}\; dxdy \\ &\text{subject to} \quad \int_\Omega ((\mathcal{A}u)(x,y) - u_0(x,y))^2\; dxdy = \sigma^2. \end{aligned} \tag{1.4}$$

This is a departure from the popular constrained least squares approach where the functional to be minimized is $\int_\Omega u_x^2 + u_y^2\; dxdy$. The total variation functional is not differentiable everywhere. Let U be a matrix representation of a 2-D image and u be a vector representation, a discretization of the problem (1.4) gives

$$\min_{u \in \Re^n} \sum_l \|B_l u\|_2 = \sum_{ij} \sqrt{(U_{i,j+1} - U_{i,j})^2 + (U_{i+1,j} - U_{i,j})^2} \tag{1.5}$$

$$\text{subject to} \quad \|Au - u_0\|_2 \le \sigma,$$

where $B_l u$ is a discretized approximation to $[u_x; u_y]$ and Au is a discretized approximation to $\mathcal{A}u$. Applying the same total variation minimization principle, Li and Santosa [17] consider a slightly different discretized regularization problem:

$$\begin{aligned} &\min_{u \in \Re^n} \|Bu\|_1 = \sum_{ij} |U_{i,j+1} - U_{i,j}| + |U_{i+1,j} - U_{i,j}| \\ &\text{s.t.} \quad \|Au - u_0\|_2 \le \sigma. \end{aligned} \tag{1.6}$$

Similarly, a component of Bu denotes a discretization of u_x or u_y.

In geophysical traveltime tomographic inversion applications [23,15,18,26], subsurface slowness structure can be determined based on the inverse problem

$$t(ray) = \int_{ray} s(x,y) d\ell.$$

In [23], a discretized linear traveltime inversion problem is considered:

$$\min_{u \in \Re^n} \|Bu - c\|_1, \tag{1.7}$$

here B denotes a distance matrix, c is the recorded traveltime perturbation and $Bu = c$ is an overdetermined system. More generally, a linear l_p problem

$$\min_{u \in \Re^n} \|Bu - c\|_p, \tag{1.8}$$

has been considered [15,31].

The discretized regularization problems (1.6) and (1.8) are examples of a general problem formulation:

$$\begin{aligned} &\min_{u \in \Re^n} \phi(u) = \|Bu - c\|_p \\ &\text{subject to} \quad \|Au - f\|_2 \le \sigma, \qquad 1 \le p < 2, \end{aligned} \tag{1.9}$$

where A and B are matrices of the appropriate sizes and $\sigma > 0$ is a given parameter. Further generalization to a minimization problem using the more complicated objective function $\sum_{i=1}^{m} \|B_i u - c_i\|_p$ with the same inequality constraint of (1.9) (which then includes the formulation (1.5)) is worthy of future research.

Problem (1.9) is a piecewise differentiable minimization possibly with a single inequality constraint (if σ is infinity then the problem is unconstrained). Solving the minimization problem (1.9) is a computationally challenging task, particularly since a problem instance can be very large. However, if a reliable and efficient computational method can be developed to solve (1.9), then this minimization approach has great potential for solving various ill-posed inverse problems in many different applications.

The main objective of this paper is to compare a few possible algorithms for minimizing a piecewise differentiable function subject to a single quadratic inequality constraint. Image restoration and traveltime inversion applications are used to compare different computational algorithms. In §2, we illustrate the usefulness of the piecewise differentiable minimization for regulating ill-posed inverse problem using a traveltime tomographic inversion example. We analyze in §3 applicability of the steepest descent, iteratively reweighted least squares (IRLS), and a recent globally convergent affine scaling Newton approach for minimizing unconstrained piecewise differentiable l_p functions. We emphasize that the steepest descent method is not guaranteed to converge to a solution for a *nondifferentiable* minimization problem and can indeed fail in practice. In §4, the affine scaling Newton approach is adapted to further include the single inequality constraint. This algorithm is demonstrated with the image restoration application. Concluding remarks are given in §5.

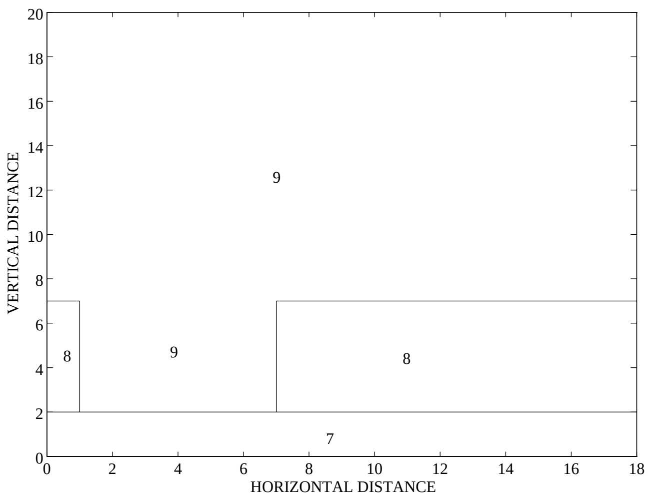

FIG. 2.1. *Geometry for Vertical Seismic Profile*

2. Piecewise Differentiable Measurements. Perhaps due to simplicity and availability of algorithms for smooth minimization, simple smooth convex function has been most widely used for data measurement. Increasingly, however, it has been recognized that a nondifferentiable measurement, e.g., a piecewise differentiable l_p norm, can be more appropriate in many different applications [25,15,26,20,22,21,9,1,17]. Computational algorithms have been developed for problems using these piecewise differentiable measurement [4,2,15,29,11,16,31,32,5]. Specifically, the piecewise linear l_1 measurement has become an increasingly attractive alternative to the 2-norm due to an appealing property of the l_1 solution: a small number of isolated large errors in the data typically do not change the solution of an unconstrained l_1 problem (e.g.,[10,23,26]).

To illustrate the usefulness of the piecewise differentiable minimization problem (1.9), we use a synthetic geophysical tomographic inversion problem described in [23], in which a square velocity anomaly is to be reconstructed.

FIG. 2.1 shows the geometry of this synthetic vertical seismic profile: there exists a square velocity anomaly in the middle layer. FIG. 2.2 shows the ray illumination from 18 sources on the surface to 18 receivers on the left side of the model. Straight rays are assumed and the lower left triangular region is covered with 136 square cells of constant slowness (reciprocal of velocity).

The following unconstrained piecewise differentiable minimization is

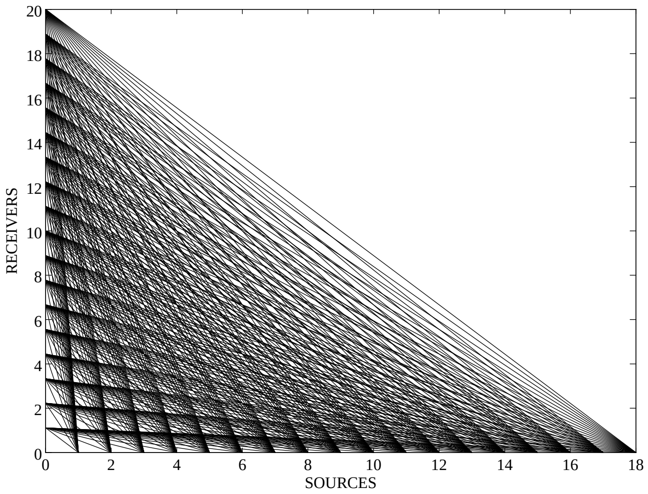

FIG. 2.2. *Ray Illumination from 18 Sources and 18 Receivers*

used in [15,23] for geophysical traveltime tomographic inversion

$$\min_{u \in \Re^n} \phi(u) = \|Bu - c\|_p, \quad 1 \leq p < 2,$$

where B is the distance matrix and c is the measured traveltime. If there is no error in c, there exists a solution satisfying $Bu = c$. This exact l_p solution is given in FIG. 2.3.

In general, however, errors exist in the measured traveltime c and different choices of the norm p lead to different solutions. To demonstrate this, we add both large spike errors and small Gaussian noise in the measured traveltime c, the least squares and l_1 reconstructions are shown in FIG. 2.4. Compared to FIG. 2.3, it is clear that the l_1 solution is less affected by the large spike error and thus is preferable in this context.

The l_p solution with $1 < p < 2$ has been used less often in practice (e.g, [23], [31]). For the synthetic seismic tomographic inversion problem considered, the l_p solutions, with p close to unity, have similar error resistance property, as illustrated in FIG. 3.2.

3. Minimization Algorithms for Unconstrained l_p Problems. Solving a constrained piecewise differentiable problem (1.9) is challenging, especially when a problem is large (which is typical for many applications). We first focus on dealing with piecewise differentiability and consider

$$\min_{u \in \Re^n} \phi(u) = \|Bu - c\|_p, \qquad 1 \leq p < 2, \tag{3.1}$$

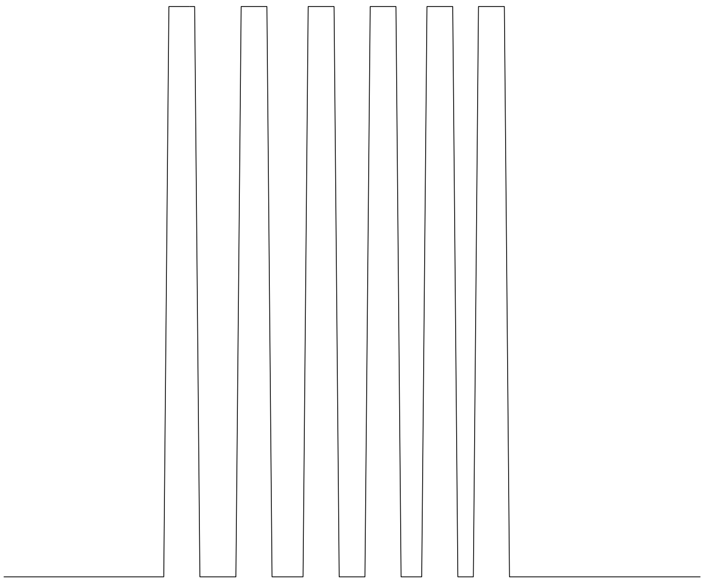

FIG. 2.3. *The Exact Solution for the Synthetic Geophysical Tomographic Inversion Example*

where $B \in \Re^{m\times n}$. We delay discussion on handling of the additional single inequality constraint until §4.

Let $r = Bu - c$ denote the residual vector. The objective function $\phi(u)$ in (3.1) is generally not differentiable everywhere, e.g., the gradient does not exist when a residual component $r_i = 0$. When $p = 1$, the linear l_1 problem can be solved by finite simplex algorithms based on linear programming [28] or a projected gradient method [4,3,6]. In this paper, we do not discuss this type of algorithms because it is less unsuitable for large-scale problems.

Most algorithms for unconstrained smooth nonlinear minimization problem are iteratively descent methods. At each iteration k, given a current approximation u^k to the solution, a descent direction d^k is computed and a new iterate $u^{k+1} = u^k + \alpha^k d^k$ is determined which produces a sufficient decrease of the objective function, i.e., $\phi(u^{k+1})$ is significantly less than $\phi(u^k)$ (The stepsize α^k can be an approximation to the minimizer of $\phi(u^k + \alpha d^k)$). This minimization process can be terminated when decrease in function $\phi(u)$ is sufficiently small, e.g.,

$$\frac{\phi(u^k) - \phi(u^{k+1})}{\phi(u^{k+1})} \le tol \quad \textbf{or} \quad k > itmax,$$

where tol is a small positive number denoting acceptable accuracy and $itmax$ is the maximum number of iterations allowed. In this paper, we consider three types of descent directions including the simple steepest descent, iteratively reweighted least squares, and the more sophisticated affine

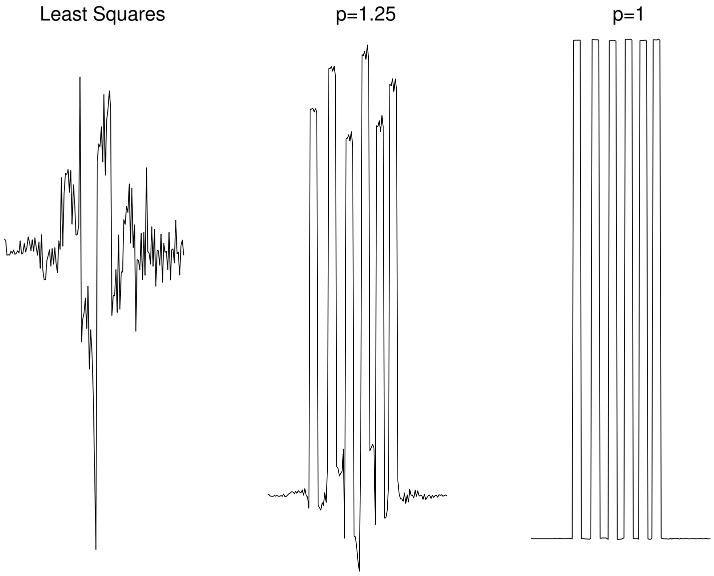

FIG. 2.4. *Comparison of Different Choice of Norms*

scaling Newton directions. Subsequently we demonstrate the effectiveness of each descent direction in the context of the linear l_p problem (3.1).

3.1. Steepest Descent Directions. If $\phi(u)$ is differentiable at u^k, then $d^k = -\nabla\phi(u^k)$ is a descent direction. The most appealing aspect of steepest descent is its low cost. For example, when $p = 1$, the gradient $\nabla\phi(u^k) = B^T \mathrm{sgn}(Bu^k - c)$ (when exists) can be computed with a matrix vector product. (In this paper, we define $\mathrm{sgn}(0) = 1$.) If $\phi(u)$ is not differentiable everywhere, e.g., for $\phi(u) = \|Bu - c\|_1$, then the steepest descent method does not, in general, lead to convergence to a solution of the problem (3.1). Moreover, even if $\{u^k\}$ converges to a solution, the convergence can be extremely slow.

To illustrate this, the reconstruction of the velocity for the traveltime inversion example in FIG.2.1, using steepest descent, is given in FIG.3.1. The descent method using steepest descent directions fails to provide any visible improvement over the starting least squares reconstruction after 500 iterations (maximum number of iterations allowed). Moreover, convergence will not be achieved eventually.

Another fairly common approach for overcoming nondifferentiability, used in many applications, is smoothing. Smoothing eliminates nondifferentiability by slightly perturbing the objective function. For example, one may choose p close to one in (3.1) in the hope of using minimization techniques for a smooth problem, e.g., the steepest descent method. We believe that this is a wrong reason for choosing p. In particular, the steepest de-

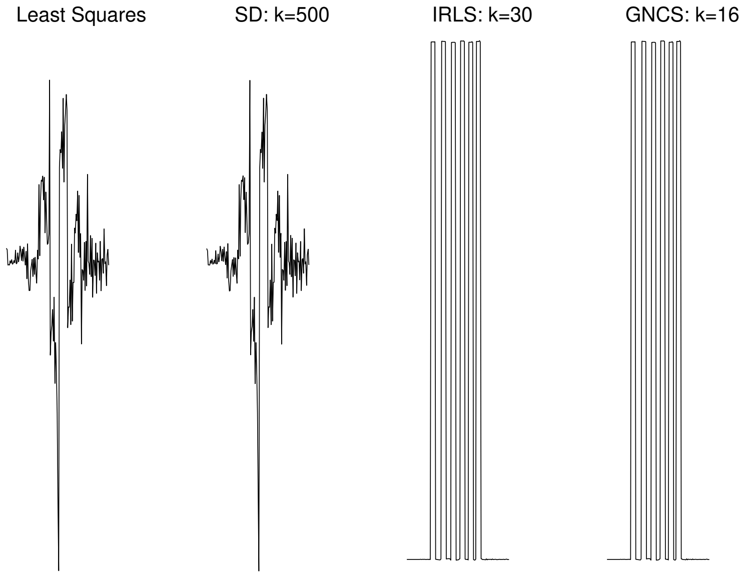

FIG. 3.1. *Comparisons of Different Methods When $p = 1$*

scent method can fail for the slightly perturbed problem, similar to the case when applied directly to the original nondifferentiable problem. To illustrate this, we apply the steepest descent method for minimizing $\|Bu - c\|_p$, with $p = 1.001$. As illustrated in FIG. 3.2, the steepest descent method again is unable to provide significant improvement over the least squares reconstruction after 500 iterations, even though the solution of (3.1) with $p = 1.001$ is essentially the same as the exact solution in FIG. 2.3 (the solution is computed using the affine scaling Newton directions described in §3.3).

3.2. Iteratively Reweighted Least Squares. In order to compute a solution with improved efficiency, an iteratively reweighted least squares (IRLS) approach has often been used for the linear l_p problem (3.1), e.g., [24,7,19]. The IRLS algorithm has become popular due to its simplicity: a weighted least squares problem is solved at the iteration k to get a descent direction and hence an improved approximation u^{k+1}. For some applications, e.g., [8], this algorithm can indeed be useful, particularly if an accurate solution is not necessary. However, there is evidence [19] which does not recommend this method in general due to its linear convergence.

The IRLS algorithm can be derived by considering a Newton step for $\phi(u)$ but ignoring the fact that the objective function may not be sufficiently smooth. Let r denote the residual vector $Bu - c$. Consider u^k and assume that the corresponding $|r^k| > 0$. A Newton step for minimizing $\phi(u) =$

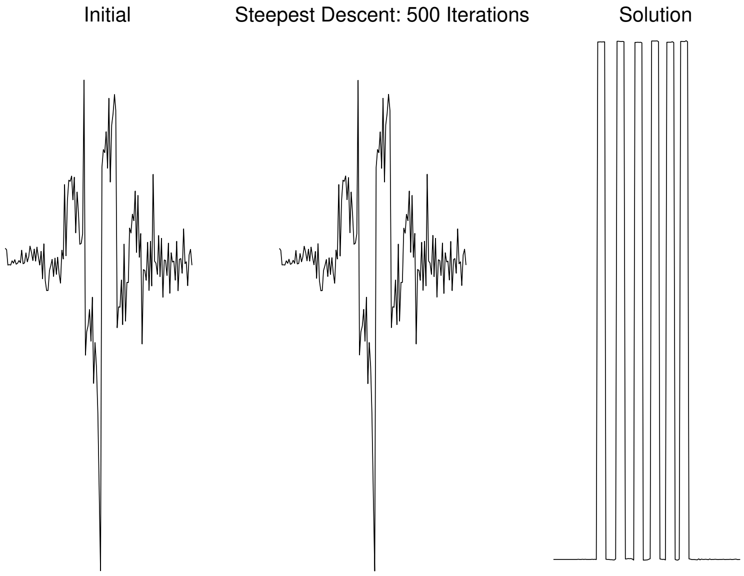

FIG. 3.2. *Reconstruction Using Steepest Descent with Smoothing ($p = 1.001$)*

$||Bu - c||_p$ is

$$B^T \operatorname{diag}(|r^k|^{p-2})Bd = -\frac{1}{p(p-1)} \nabla\phi(u^k),$$

which is the normal equation for the weighted least squares problem

$$\operatorname{diag}(|r^k|^{\frac{p-2}{2}})Bd \stackrel{\text{LS}}{=} -\frac{1}{p(p-1)} \operatorname{diag}(|r^k|^{\frac{2-p}{2}})g^k. \tag{3.2}$$

The direction d^k computed from above equations is descent, and, the iterates $\{u^k\}$ generated by IRLS satisfy $u^{k+1} = u^k + (p-1)d^k$ and converge linearly to a solution [30,19]. A line search acceleration of IRLS can be easily made to achieve a better reduction of $\phi(u)$ along the descent direction d^k. Subsequently, when we refer to the IRLS method, we are referring to this accelerated IRLS.

The reconstruction using the IRLS method is given in FIG. 3.1. The iterates from the IRLS indeed converge in 30 iterations and a good reconstruction is achieved.

3.3. A Globally Convergent Affine Scaling Newton Approach. The IRLS method successfully achieves convergence with a significant cost: solving a weighted least squares problem at each iteration. With increasing computing power and availability of reliable and efficient least squares solver on sequential and parallel computers, the additional cost of IRLS may become more acceptable.

Nonetheless, the IRLS method is only a linearly convergent method. Fast convergence is essential for efficiency if an accurate solution of a large problem (3.1) is desired.

In [16,12], a globally convergent affine scaling Newton method is developed for the problem (3.1). This globalized Newton approach (GNCS) is derived by considering complementarity conditions for the piecewise differentiable minimization (3.1). In particular, a descent direction is defined based on both the current primal iterate u^k and approximations to the dual multipliers λ^k.

Consider $\min_x \|Bu - c\|_p$. Let $r = Bu - c$ and $g = p|r|^{p-1}\mathrm{sgn}(r)$. The optimality conditions for this problem are described in both x and dual multipliers λ: a point u is a solution to (3.1) if there exists λ such that

$$\begin{array}{c} \mathrm{diag}(r)(g-\lambda) = 0 \\ B\lambda = 0, \quad Zr = Zb. \end{array} \tag{3.3}$$

with $|\lambda| \le |g|$. Here Z is a matrix whose rows form a basis for the null space of B, i.e., $Z^T = \mathrm{null}(B)$. We note that *explicit knowledge of Z is not necessary in the actual algorithm.*

Let $D_r = \mathrm{diag}(|r|)$ and $D_\lambda = \mathrm{diag}(\mathrm{sgn}(r))\mathrm{diag}(pg - \lambda)$. A Newton step $d \in \Re^n$ and $d_w \in \Re^{m-n}$ for (3.3) can be computed:

$$[D_\lambda^k B, -D_r^k Z^T] \begin{bmatrix} d \\ d_w \end{bmatrix} = -D_r^k(g^k - Z^T w^k). \tag{3.4}$$

Observing the special structure of the linear system (3.4), it is easy to see that the Newton direction d^k equals

$$d^k = -(B^T D_\lambda^k {D_r^k}^{-1} B)^{-1} B^T g^k. \tag{3.5}$$

Globalization of the Newton step is typically required in the minimization context to ensure proper descent directions are defined everywhere. Let $\theta^k \in \Re^m$ denote the optimality measurement below

$$\begin{array}{l} \theta = \frac{\eta e}{\gamma|g| + \eta e}, \\ \eta = \max(\max(\frac{|D_r(g-\lambda)|}{\phi(r_0)}), \max(\max(|\lambda| - |g|, 0))) \end{array} \tag{3.6}$$

where $0 < \gamma < 1$ is a constant and e is the vector of all ones. We use the Matlab [27] definition for the function max: $\max(x)$ denotes the maximum component of a vector x and the value of $\max(x, y)$ is a vector whose components are the maximum of the corresponding components of x and y.

There are many possible ways to globalize. An example which works well computationally is given in [12,16]. This globalization replaces the diagonal matrix D_λ^k by a combination D_θ^k of $\mathrm{diag}(g^k)$ and D_λ^k using the optimality measurement $\theta^k > 0$:

$$D_\theta^k = \mathrm{diag}(|\theta^k g^k + (e - \theta^k)(pg^k - \lambda^k)|).$$

Hence a globalized descent direction d^k is computed from

$$d^k = -B(B^T D_\theta^k {D_r^k}^{-1} B)^{-1} B^T g^k. \tag{3.7}$$

It is easy to see that this direction d^k can be computed as a weighted least squares solution

$$(D_\theta^k {D_r^k}^{-1})^{\frac{1}{2}} Bd \stackrel{\text{LS}}{=} -(D_\theta^k {D_r^k}^{-1})^{-\frac{1}{2}} g^k. \tag{3.8}$$

Dual multipliers approximations can be updated:

$$\lambda^{k+1} = D_\theta^k {D_r^k}^{-1} Bd^k + g^k.$$

The new iterate is

$$u^{k+1} = u^k + \alpha^k d^k,$$

where α^k is the stepsize which approximately minimizes the piecewise differentiable function $\phi(u^k + \alpha d^k)$ and can be determined easily (we refer a reader to [12,16] for details of the line search).

We observe that the weighted least squares problems (3.2) and (3.8) have the same coefficient matrix B but with different diagonal scalings. Hence the direction d^k (3.8) can be computed with the same cost as that of the IRLS method (3.2). The approximate multipliers can be obtained with negligible cost. FIG.3.1 illustrates the reconstruction for the same example in FIG. 2.1 using the described affine scaling Newton method. This affine scaling Newton method takes roughly half the number of iterations of the IRLS method.

We have performed many computational experiments which indicate that the globally convergent affine scaling Newton approach [12,16] generally takes significantly fewer iterations than the IRLS method (less than half). Table 3.3 illustrates typical behaviors of the two methods in iteration accounts for some randomly generated problems (3.1) with $p = 1$. The stopping tolerance is 10^{-12} and the maximum number of iterations allowed is 50.

4. Solving a Linear l_p Problem with a Single quadratic Constraint. For many ill-posed inverse problems, e.g., image restoration, an *a priori* condition on the desired solution is needed for regularization. Hence it may be appropriate to solve a constrained piecewise differentiable minimization problem (1.9). In this section, we will use image restoration as an example to illustrate how to incorporate the constraint information using the affine scaling Newton method [12] described in §3.3. Specifically, we consider solving (1.6):

$$\begin{aligned} &\min_{u \in \Re^n} \|Bu - c\|_1 \\ \text{subject to} \quad &\|Au - u_0\|_2 \le \sigma. \end{aligned} \tag{4.1}$$

Number of Steps		$m = 200$
n	GNCS	IRLS
10	17	50
30	17	50
50	15	50
70	21	50
90	15	50
110	14	50
130	17	50
150	13	50
170	13	50
190	9	50

It is well recognized that generating iterates to follow nonlinear constraints is difficult; even a single nonlinear constraint adds substantial computational difficulty. The classical approaches for nonlinear constraints include the l_1 penalty and Lagrangian method [13]. The difficulty with these methods is determination of the *a priori* unknown penalty parameter. We believe that a more appealing approach here is to maintain feasibility for the single constraint $||Ax - u_0||_2 \leq \sigma$ explicitly.

For image restoration application, a solution typically appears on the quadratic surface. If the solution lies strictly inside the constraint $||Ax - u_0||_2 < \sigma$, then the problem is essentially unconstrained and can be solved as described in §3.3. In general, a decision can be made as to whether the solution is exactly on the nonlinear constraint surface with available dual multipliers approximation. Employing the active set strategy on the single quadratic constraint, descent directions can be computed to leave or remain on this single constraint surface, e.g. [13]. We assume subsequently that the solution lies on the quadratic constraint surface $||Au - u_0||_2 = \sigma$.

The feasibility to the constraint $||Au - u_0||_2 \leq \sigma$ can be achieved by applying a minimization process (e.g., a conjugate gradient method) to the convex quadratic function $||Au - u_0||_2^2$ until the feasibility $||Au - u_0||_2 \leq \sigma$ is obtained. This feasibility can be easily maintained using a line search with a correction [13,17]. We subsequently concentrate on computing a good descent direction for (4.1).

Let $J = [A^T(Au - u_0), B]$, $\hat{g} = [0; \mathrm{sgn}(Bu - c)]$ and $\hat{r} = [||A^Tu - u_0||^2 - \sigma^2; Bu - c]$. The nonlinear system which captures optimality of the constrained piecewise linear minimization problem (4.1) is

$$(4.2) \qquad \begin{aligned} &J\lambda = 0, \\ &\mathrm{diag}(\hat{r})(\hat{g} - \hat{\lambda}) = 0. \end{aligned}$$

Note that the additional condition, $|\lambda_i| \leq 1$ for any $\hat{r}_i = 0$ and $i > 1$, needs to hold.

Let $\hat{D}^k_\lambda = \mathrm{diag}(\mathrm{sgn}(\hat{r}^k))\mathrm{diag}(\hat{g}^k - \hat{\lambda}^k)$. A Newton step for (4.2) is

$$\begin{bmatrix} \hat{\lambda}^k_1 A^T A & J^k \\ \hat{D}^k_\lambda {J^k}^T & -\mathrm{diag}(\hat{r}^k) \end{bmatrix} \begin{bmatrix} d^k \\ \hat{\lambda}^{k+1} \end{bmatrix} = - \begin{bmatrix} 0 \\ \mathrm{diag}(\hat{r}^k)\hat{g}^k \end{bmatrix}. \tag{4.3}$$

Similar to discussion in §3.3, globalization is needed to ensure that descent directions are properly determined and the globalized Newton directions converge to the Newton direction (4.3) asymptotically. We consider a similar globalization as described before: we replace $\hat{D}^k_\lambda$ by

$$\hat{D}^k_\theta = \mathrm{diag}(|\hat{\theta}^k \hat{g}^k + (e - \hat{\theta}^k)(\hat{g}^k - \hat{\lambda}^k)|)$$

and consider a globalized Newton direction as below:

$$\begin{bmatrix} |\hat{\lambda}^k_1| A^T A & J^k \\ D^k_\theta {J^k}^T & -\mathrm{diag}(|\hat{r}^k|) \end{bmatrix} \begin{bmatrix} d^k \\ \hat{\lambda}^{k+1} \end{bmatrix} = - \begin{bmatrix} 0 \\ \mathrm{diag}(|\hat{r}^k|)\hat{g}^k \end{bmatrix}.$$

Eliminating the dual variables, we have

$$(|\hat{\lambda}^k_1| A^T A + J^k \hat{D}^k_\theta {\hat{D}^k_r}{}^{-1} {J^k}^T) d = -J^k \hat{g}^k.$$

Compared to (3.7), (4.4) has an additional term $|\hat{\lambda}^k_1| A^T A$ which represents the constraint curvature. Let $\hat{D}^k$ denote the diagonal scaling matrix below

$$\hat{D}^k = \begin{bmatrix} (\hat{D}^k_\theta {\hat{D}^k_r}{}^{-1})^{\frac{1}{2}} & 0 \\ 0 & \sqrt{|\hat{\lambda}^k_1| I} \end{bmatrix}. \tag{4.4}$$

A simple algebraic manipulation shows that the globalized direction d^k (3.8) can be computed from a weighted least squares solve

$$\hat{D}^{k\,\frac{1}{2}} \begin{bmatrix} {J^k}^T \\ A \end{bmatrix} d^k \overset{\mathrm{LS}}{=} -\hat{D}^{k\,-\frac{1}{2}} \begin{bmatrix} \hat{g}^k \\ 0 \end{bmatrix}, \tag{4.5}$$

and

$$\hat{\lambda}^{k+1} = (\hat{D}^k_\theta {\hat{D}^k_r}{}^{-1}) d^k + \hat{g}^k.$$

A simple correction technique for maintaining feasibility can be incorporated to stay on the quadratic surface [17]. For efficient computation, the single dense row structure for the least squares problem (4.5) needs to be exploited.

To demonstrate the descent algorithm using the affine scaling Newton direction (4.5) for the constrained minimization problem (1.6), we generate the blurred and noisy image u_0 by a motion blurring function A, and adding a measured amount of random noise ε, i.e.,

$$u_0 = A u_{true} + \varepsilon.$$

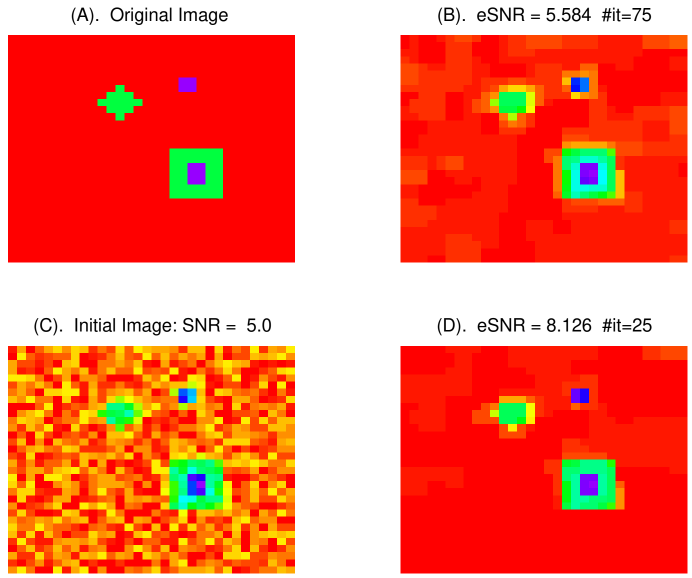

FIG. 4.1. *Comparisons of Affine Scaling Newton Directions and Steepest Descent Directions*

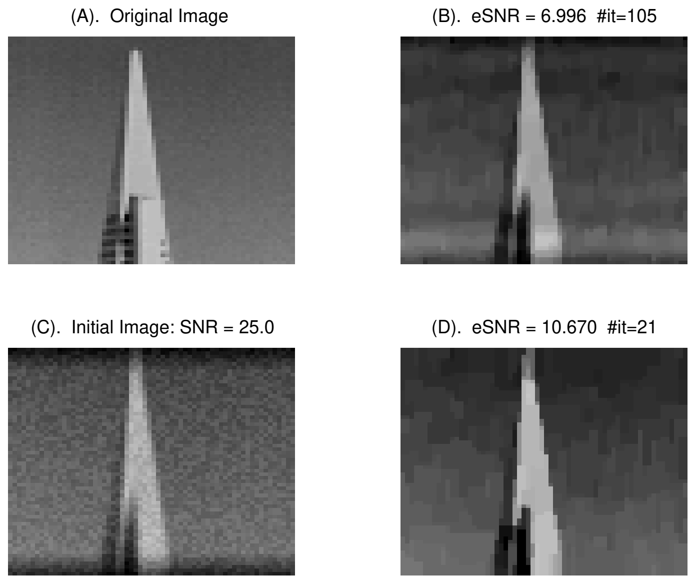

FIG. 4.2. *Comparisons of Affine Scaling Newton Directions and Steepest Descent Directions*

The amount of noise in the noisy and blurred image u_0 is described by the signal-to-noise ratio (SNR):

$$\mathrm{SNR} = 10 \log \frac{\text{variance of the blurred image}}{\text{variance of the noise}} \qquad (dB).$$

When assessing the quality of the restored images, we consider the signal-to-noise ratio improvement e_{SNR}

$$e_{\mathrm{SNR}} = 10 \log \frac{||u_{true} - u_0||_2^2}{||u_k - u_{true}||_2^2} \qquad (dB).$$

When SNR is high, a larger signal-to-noise ratio improvement e_{SNR} suggests a better restoration.

In FIG. 4.1 and 4.2, we demonstrate the quality of image restoration for a 32-by-32 and 64-by-64 blurred and noisy images respectively. We apply both the steepest descent and affine scaling Newton directions for comparison. The stopping criteria for these images are 0.5×10^{-4} with the maximum of 200 iterations allowed. The parameter $\sigma = 0.95 \times ||\varepsilon||_2$ where ε is the random white noise.

Unlike the geophysic traveltime inversion problem, surprisingly good improvements are achieved using steepest descent directions. However, the images computed using the affine scaling Newton directions are consistently better and achieve higher signal-to-noise ratio improvements. Compared to the affine scaling Newton direction, a steepest descent direction is significantly cheaper to compute. Nonetheless, the steepest descent method is more sequential than the Newton approach. In a parallel computing environment with a good parallel least squares solver, the affine scaling Newton method with a dominant work on the least squares solve may prevail if a sparse matrix A is available.

If the matrix A is not available or is very dense, the computation of an affine scaling Newton step (4.5) may be impossible or extremely expensive. We are currently investigating possibility of computing an affine scaling Newton step (4.5) iteratively.

5. Concluding Remarks. Piecewise differentiable minimization with a single quadratic constraint has great potential in regulating ill-posed inverse problems. We analyze three descent methods using steepest descent, iteratively reweighted least squares and the more recent affine scaling Newton directions [12,16]. We illustrate with a geophysic traveltime inversion example, that the steepest descent can fail to achieve convergence for a piecewise differentiable minimization and produce entirely unacceptable results. The alternative iteratively reweighted least squares method is more computational extensive and can be inefficient due to its linear convergence. The globally convergent affine scaling Newton approach achieves quadratic convergence with negligible additional cost, compared to the IRLS approach. The affine scaling Newton approach is further adapted to include a single inequality constraint. Compared to the steepest descent, the affine scaling Newton method produces better reconstruction with higher signal-to-noise improvement ratios in image restoration applications.

The future investigation includes possible iterative solvers for computing the Newton steps for the globalized Newton approach [12,16]. The issue of preconditioner becomes extremely important in this regard. Moreover, algorithms for a more general piecewise differentiable problem of the form

$$\begin{aligned} &\min_{u\in\Re^n} \phi(u) = \sum_{i=1}^{t} \|B_i u - c\|_p \\ &\text{subject to} \quad \|Au - f\|_2 \le \sigma, \end{aligned} \tag{5.1}$$

may be useful for various applications.

Acknowledgements. The author would like to thank Fadil Santosa for introducing image restoration problem to her, Tom Coleman for his comments, and Chunguang Sun for use of his sparse least squares solve routines.

REFERENCES

[1] S. Alliney and S. Ruzisky, *An algorithm for the minimization of mixed l_1 and l_2 norms with applications to bayesian estimation*, IEEE Trans. on Signal Processing, 42 (1994), pp. 618–627.

[2] I. Barrodale and C. Phillips, *An improved algorithm for discrete Chebychev linear approximation*, in Proc. 4th Manitoba Conf. on Numer. Math., U. of Manitoba, Winnipeg, Canada, 1974, pp. 177–190.

[3] I. Barrodale and F. Roberts, *An improved algorithm for discrete l_1 linear approximation*, SIAM J. Num. Anal., 10 (1973), pp. 839–848.

[4] I. Barrodale and A. Young, *Algorithms for best l_1 and l_∞ linear approximations on a discrete set*, Numer. Math., 8 (1966), pp. 256–306.

[5] R. H. Bartels, A. R. Conn, and Y. Li, *Primal methods are better than dual methods for solving overdetermined linear systems in the l_∞ sense?*, SIAM J. Numer. Anal., 26 (1989), pp. 693–726.

[6] R. H. Bartels, A. R. Conn, and J. W. Sinclair, *Minimization techniques for piecewise differentiable functions: the l_1 solution to an overdetermined linear system*, Siam J. Numer. Anal., 15 (1978), pp. 224–240.

[7] A. E. Beaton and J. W. Tukey, *The fitting of power series, meaning polynomials, illustrated on band-spectrographic data*, Technometrics, 16 (1974), pp. 147–185.

[8] P. J. Besl, J. B. Birch, and L. T. Watson, *Robust window operators*, Machine Vision and Applications, 2 (1989), pp. 179–191.

[9] C. Bouman and K. Sauer, *A generalized Gaussian model for edge preserving MAP estimation*, IEEE Trans. on Image Processing, 2 (1993), pp. 298–310.

[10] J. F. Claerbout and F. Muir, *Robust modeling with erratic data*, Geophysics, 38 (1973), pp. 826–844.

[11] T. F. Coleman and Y. Li, *A global and quadratically-convergent method for linear l_∞ problems*, SIAM Journal on Scientific and Statistical Computing, 29 (1992), pp. 1166–1186.

[12] ———, *A globally and quadratically convergent affine scaling method for linear l_1 problems*, Mathematical Programming, 56 (1992), pp. 189–222.

[13] R. Fletcher, *Practical Methods of Optimization: Volume 2, Constrained Optimization*, John Wiley and Sons, 1981.

[14] B. R. Hunt, *The application of constrained least squares estimation to image restoration by digital computer*, IEEE Transactions on Computers, C-22 (1973), pp. 805–812.

[15] A. G. John A. Scales and S. Treitel, *Fast l_p solution of large, sparse, linear systems and application to seismic travel time tomography*, Journal of Computational Physics, 75 (1988), pp. 314–333.

[16] Y. Li, *A globally convergent method for l_p problems*, SIAM Journal on Optimization, 3 (1993), pp. 609–629.

[17] Y. Li and F. Santosa, *An affine scaling algorithm for minimizing total variation in image enhancement*, Tech. Rep. TR 94-1470, Computer Science Department, Cornell University, 1994. To appear in Trans. on Image Processing, June, 1996.

[18] L. R. Lines, *Applications of seismic traveltime tomography–a review*, in Geophysical Inversion, 1992, pp. 333–337. SIAM.

[19] M. R. Osborne, *Finite Algorithms in Optimization and Data Analysis*, Wiley Series in Probability and Mathematical Statistics, 1985.

[20] L. Rudin and S. Osher, *Feature-oriented image enhancement using shock filters*, SIAM J. Numer. Anal., 27 (1990), pp. 919–940.

[21] L. Rudin, S. Osher, and C. Fu, *Total variation based restoration of noisy blurred images*, SIAM J. Numer. Anal., (?), p. to appear.

[22] L. Rudin, S.Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D., 60 (1992), pp. 259–268.

[23] J. A. Scales, *Tomographic inversion via the conjugate gradient method*, Geophysics, 52 (1987), pp. 179–185.

[24] E. J. Schlomacher, *An iterative technique for absolute deviations curve fitting*, J. Am. Stat. Assoc., 68 (1973), pp. 857–865.

[25] H. Späth, *Mathematical Algorithms for Linear Regression*, Academic Press, 1992.

[26] H. L. Taylor, S. C. Banks, and J. F. McCoy, *Deconvolution with the l_1 norm*, Geophysics, (1979), pp. 39–52.

[27] The MathWorks Inc., *Matlab Reference guide*, The MathWorks, Natick, Mass, 1992.

[28] H. M. Wagner, *Linear programming techniques for regression analysis*, J. Am. Stat. Assoc., 54 (1959), pp. 206–212.

[29] G. A. Watson, *On two methods for discrete l_p approximation*, Computing, 18 (1977), pp. 263–266.

[30] J. M. Wolfe, *On the convergence of an algorithm for discrete l_p approximation*, Numer. Math., 32 (1979), pp. 439–459.

[31] R. Yarlagadda, J. B. Bednar, and T. L. Watt, *Fast algorithms for l_p deconvolution*, IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-33 (1985), pp. 174–182.

[32] Y. Zhang, *A primal-dual interior point approach for computing the l_1 and l_∞ solutions of overdetermined linear systems*, Tech. Rep. Technical Report, Department of Mathematics and Statistics, University of Maryland, 1991.