AN ACTIVE SET STRATEGY BASED ON THE AUGMENTED LAGRANGIAN FORMULATION FOR IMAGE RESTORATION

Kazufumi Ito[1] and Karl Kunisch[2]

Abstract. Lagrangian and augmented Lagrangian methods for nondifferentiable optimization problems that arise from the total bounded variation formulation of image restoration problems are analyzed. Conditional convergence of the Uzawa algorithm and unconditional convergence of the first order augmented Lagrangian schemes are discussed. A Newton type method based on an active set strategy defined by means of the dual variables is developed and analyzed. Numerical examples for blocky signals and images perturbed by very high noise are included.

Résumé. On analyse les méthodes de lagrangien et de lagrangien augmenté pour des problèmes d'optimisation non différentiable, provenant de la formulation de variation totale bornée en restauration d'images. La convergence conditionnelle de l'algorithme d'Uzawa et la convergence inconditionnelle des schémas de premier ordre de lagrangien augmenté sont discutées. Une méthode de type Newton basée sur une statégie d'ensemble actif, définie au moyen de variables primales et duales, est développée et analysée. Des exemples numériques sont donnés pour des signaux discontinus et des images pertubées par très fort bruit.

1. Introduction

In this paper we present an augmented Lagrangian technique to numerically solve the nonsmooth convex optimization problem that arises in image restoration. The image restoration problem is formulated as the minimization of a least squares fit criterion and of a regularization of total bounded variational (BV-)type. BV-regularization terms have the advantage over square of gradient-type regularization in that they minimize undesirable smearing of corners and of discontinuities in the image. Due to the BV-term the cost functional becomes nondifferentiable, however, and an efficient numerical technique needs to be employed which allows a reliable minimization. Here we present such a technique.

To briefly describe the formulation of the problem let u denote the grey values of a noisy image z occupying a two-dimensional bounded region Ω. To restore (or denoise) the image $\bar{u}$ one considers the variational problem

$$\min \int_\Omega (\tfrac{\mu}{2}|\nabla u|^2 + g|\nabla u|)\, dx + \tfrac{1}{2}\int_\Omega |u - z|^2 dx \text{ over } u \in H_0^1(\Omega). \tag{1.1}$$

In the infinite dimensional part of our study we focus on the case where the constant μ is positive and small with respect to g. For the discretized problems we may take $\mu \geq 0$. From [7] we recall that $\int_\Omega |\nabla u| dx$ is equivalent to the BV-seminorm on $H_0^1(\Omega)$-functions.

---

[1] Department of Mathematics, North Carolina State University, Raleigh, NC 27695, USA.
[2] Institut für Mathematik, Universität Graz, 8010 Graz, Austria. e-mail: `karl.kunisch@kfunigraz.ac.at`

While there is a large literature on image restoration the analysis of nondifferential optimization techniques started only rather recently. In [10] the authors propose the BV-formulation (1.1) with $\mu = 0$ and describe a gradient type algorithm for numerical implementation. In [11] the BV-term in (1.1) is replaced by $\int_\Omega \sqrt{\varepsilon + |\nabla u|^2}$, for small positive ε, thus circumventing the problem of lack of differentiability of the cost functional. The requirement $\varepsilon > 0$ was eliminated in [2] where image deblurring problems are considered, *i.e.* the data z are related to Ku with K a compact operator. The numerical methods in [2, 11] are, however, completely different from those proposed in the present work.

In [5] the authors give an interesting numerical account of the advantages and disadvantages of the BV-formulation for the image restoration problem. The optimality condition related to (1.1) leads to a nonlinear elliptic equation. In this context we mention the recent work of nonlinear parabolic diffusion models [1, 4] which, in particular guarantee that smoothing is enhanced parallel to edges in the original picture and its intensity depends on the radius of curvature of the edges.

In the present paper the approach to solve (1.1) is based on abstract results on Lagrangian and augmented Lagrangian techniques for nonsmooth optimization [8]. These techniques allow to substitute the nondifferentiable term by a new variable, called the Lagrangian or dual variable. At the solution the primal variable u and the dual variable satisfy a nonlinear system of equations, called the optimality system. This approach is comparable (and on an abstract level equivalent) to realizing equality and inequality constraints in optimization problems by Lagrangian terms in the cost functional. In Section 2 we develop the duality framework and describe first order methods for solving the optimality system. Specifically, the Uzawa algorithm and the first order augmented Lagrangian method are compared. The former can be considered as explicit algorithm in the dual variable which is conditionally convergent. The latter on the other hand is an implicit method that converges unconditionally.

Section 3 is devoted to the development of a Newton type algorithm that is based on the duality theory of Section 2 for solving the optimality system. Difficulties arise due to the fact that the dual variable is not uniquely defined on the active set, *i.e.* on $\{x : \nabla u(x) = 0\}$. To circumvent this problem an active set strategy based on the first order augmented Lagrangian update of the dual variable is employed. A similar technique was proposed for finite dimensional inequality constrained optimization problems in [3]. The convergence analysis of the method is impeded by the fact that $|\nabla u(x)|$ can become zero arbitrarily slowly as x varies and that the gradient operator is not surjective. We therefore restrict ourselves to the discretized formulation of (1.1) in that section. Section 3 closes with a brief account of the case when Ω is only one-dimensional. The practical implications of this case are given by voice detection problems. Numerical results for both, the image reconstruction and the voice detection cases are given in Section 4.

2. Convex optimization and necessary optimality condition

Consider the variational problem

$$\min \int_\Omega \left(\frac{\mu}{2} |\nabla u|^2 + g |\nabla u| \right) dx + \frac{1}{2} \int_\Omega |u - z|^2 \, dx \text{ over } u \in H_0^1(\Omega), \tag{2.1}$$

for restoring the original image u from a noisy image $z \in L^2(\Omega)$. Here $\mu > 0$ and $g > 0$ are fixed but can be adjusted to the statistics of noise. If $\mu = 0$ then this problem is equivalent to the image enhancement in [10] based on minimization of the BV semi-norm formulation $\int_\Omega |\nabla u|$ subject to the standard deviation constraint $\int_\Omega |u - z|^2 \, dx = \sigma^2$. In fact $1/g$ is the Lagrange multiplier associated with the equality constraint $\int_\Omega |u - z|^2 \, dx = \sigma^2$. To describe and analyze algorithms to solve (2.1) in a systematic manner we shall consider (2.1) as a special case of the following optimization problem:

$$\text{minimize } f(x) + \varphi(\Lambda x). \tag{2.2}$$

Here $f : X \to R$ and $\varphi : H \to R$ are convex functions, $\Lambda : X \to H$ is a bounded linear operator, and X, H are Hilbert spaces. It is assumed that f is a continuously differentiable, coercive function and that φ is a proper, lower semicontinuous, convex function. The functional φ represents the nonsmooth component of the costfunctional.

For the image restoration problem (2.1), we choose

$$X = H_0^1(\Omega), H = L^2(\Omega) \times L^2(\Omega) \text{ and } \Lambda = \text{ grad},$$

and define f and φ by

$$f(u) = \tfrac{1}{2} \int_\Omega (\mu |\nabla u|^2 + |u - z|^2) \, dx,$$

and

$$\varphi(v_1, v_2) = g \int_\Omega \sqrt{v_1^2 + v_2^2} \, dx.$$

We note that there are no essential changes if X is chosen to be $H^1(\Omega)$. In either case f is coercive on X.

2.1. Lagrange multiplier

We discuss the necessary optimality condition for (2.1) in terms of the Lagrange multiplier rule for the nonsmooth minimization problem (2.2). For this purpose H is identified with its dual $H^*, (\cdot, \cdot)$ denotes the inner product on H and $\langle \cdot, \cdot \rangle$ denotes the duality product of $X^* \times X$. We assume that f and φ are bounded below by zero and that

$$\langle f'(x_1) - f'(x_2), x_1 - x_2 \rangle \geq \sigma |x_1 - x_2|_X^2 \tag{2.3}$$

for a constant $\sigma > 0$ and all $x_1, x_2 \in X$. The subdifferential $\partial\varphi(v)$ of φ is defined by

$$\partial\varphi(v) = \{v^* \in H : \varphi(y) - \varphi(v) \geq (v^*, y - v) \text{ for all } y \in H\}.$$

That is, the differential is no longer single-valued. Since φ is proper there exists an element $v_0 \in D(\varphi) = \{v \in H : \varphi(v) < \infty\}$ such that

$$\varphi(\Lambda x) - \varphi(v_0) \geq (v_0^*, \Lambda x - v_0)_H \text{ for all } v_0^* \in \partial\varphi(v_0). \tag{2.4}$$

Hence, $\lim f(x) + \varphi(\Lambda x) \to \infty$ as $|x|_X \to \infty$ and it follows from [6] that there exists a unique minimizer $x^* \in X$ for (2.2). The necessary and sufficient condition for $x^* \in X$ to be the minimizer of (2.2) is given by

$$\langle f'(x^*), x - x^* \rangle + \varphi(\Lambda x) - \varphi(\Lambda x^*) \geq 0 \text{ for all } x \in X. \tag{2.5}$$

Suppose that φ is differentiable at Λx^*. Then (2.5) is equivalent to $f'(x) + \Lambda^* \varphi'(\Lambda x^*) = 0$. For nondifferentiable, proper, lower semicontinuous, convex function φ (2.5) implies the existence of $\lambda^* \in H$ such that

$$\lambda^* \in \partial\varphi(\Lambda x^*) \text{ and } f'(x^*) + \Lambda^* \lambda^* = 0, \tag{2.6}$$

provided, for example, that φ is finite and continuous at Λx^*. Conversely, suppose $(x^*, \lambda^*) \in X \times H$ satisfies (2.6). Then, $\varphi(\Lambda x) - \varphi(\Lambda x^*) \geq (\lambda^*, \Lambda(x - x^*))$ for all $x \in X$. Thus, (2.6) implies (2.5). It then follows from (2.4) that x^* minimizes (2.2).

For the image restoration problem $\varphi(v)$ is finite and continuous at each $v \in H$ and $\lambda \in \partial\varphi(v)$ is given by (*e.g.*, see [6,8])

$$\lambda \in C \text{ and } \lambda(x) \cdot v(x) = |v(x)| \text{ a.e. } x \in \Omega$$

where

$$C = \{\lambda \in H : |\lambda(x)| \leq 1 \text{ a.e. } x \in \Omega\}.$$

Thus, the necessary and sufficient optimality condition for (2.1) is given by the existence of $(u, \lambda) \in X \times H$ satisfying

$$\begin{aligned}
-\mu \Delta u + u - g \operatorname{div} \lambda &= z \\
\lambda(x) \cdot \nabla u(x) &= |\nabla u(x)| \text{ a.e. } x \in \Omega.
\end{aligned} \tag{2.7}$$

Remark 2.1. The analysis of this section requires Hilbert space structure and for this reason $\mu > 0$ is assumed. Let us briefly comment, however, on the limit problem as $\mu \to 0^+$. We consider

$$\min g \int_\Omega |\nabla u| + \frac{1}{2} \int_\Omega |u - z|^2 \, dx \text{ over } u \in BV(\Omega) \text{ with } \tau_\Gamma u = 0 \tag{2.8}$$

where $BV(\Omega)$ denotes the space of functions of bounded variation and τ_Γ denotes the trace in the sense of $BV(\Omega)$ on the boundary Γ of Ω, as defined in [7], for example. Let us, within this remark, denote the solutions to (2.1) by u_μ. We shall argue that the sequence $\{u_\mu\}$ converges to the solution u of (2.8) as $\mu \to 0^+$. In fact $\{u_\mu : \mu \in (0,1]\}$ is bounded in $BV(\Omega)$ and hence in $L^2(\Omega)$ [7]. It follows that there exists a subsequence, denoted by the same symbols, and $\bar{u} \in BV(\Omega)$, such that $u_\mu \to \bar{u}$ weakly in $L^2(\Omega)$ and strongly in $L^1(\Omega)$, [7]. Moreover $\int_\Omega |\nabla \bar{u}| \leq \liminf_{\mu \to 0} \int_\Omega |\nabla u_j|$, [7] and hence we may take the lim inf with respect to $\mu > 0$ in

$$\int_\Omega \left(\frac{\mu}{2} |\nabla u_\mu|^2 + g |\nabla u_\mu|\right) dx + \frac{1}{2} \int_\Omega |u_\mu - z|^2 \, dx \leq \int_\Omega \left(\frac{\mu}{2} |\nabla u|^2 + g |\nabla u|\right) dx + \frac{1}{2} \int_\Omega |u - z|^2 \, dx$$

for every $u \in H_0^1(\Omega)$ to obtain

$$g \int_\Omega |\nabla \bar{u}| + \frac{1}{2} \int_\Omega |\bar{u} - z|^2 \, dx \leq g \int_\Omega |\nabla u| + \frac{1}{2} \int_\Omega |u - z|^2 \, dx,$$

for all $u \in H_0^1(\Omega)$. Since for every $u \in BV(\Omega)$ with $\tau_\Gamma u = 0$ (in the sense of $BV(\Omega)$) there exists $\{u_j\} \in H_0^1(\Omega)$ with $\lim_{j \to \infty} |u_j - u|_{L^2} = 0$ and $\lim_{j \to \infty} \int_\Omega |\nabla u_j| = \int_\Omega |\nabla u|$, ([7], p. 38), it follows that $\bar{u}$ is a solution to (2.8). $\qquad\square$

2.2. Augmented Lagrangian formulation

In this section we apply a regularization method based on augmented Lagrangians [3,8] to obtain a regularized optimality condition that is equivalent to the complementary condition $\lambda^* \in \partial\varphi(\Lambda x^*)$ in (2.17) below. This regularization is a generalization of the well-known Moreau-Yosida regularization. It is clear that problem (2.2) is equivalent to

$$\text{minimize } f(x) + \varphi(\Lambda x - u) \text{ subject to } u = 0 \text{ in } H.$$

The equality constraint $u = 0$ is treated by the augmented Lagrangian method. That is, we consider the minimization over $x \in X$ and $u \in H$ of the form

$$\text{minimize } f(x) + \varphi(\Lambda x - u) + (\lambda, u)_H + \frac{c}{2}|u|_H^2 \tag{2.9}$$

where $\lambda \in H$ is a multiplier and c is a positive scalar penalty parameter [3]. Equivalently, problem (2.9) is written as

$$\text{minimize } L_c(x, \lambda) = f(x) + \varphi_c(\Lambda x, \lambda) \tag{2.10}$$

where $\varphi_c(v, \lambda)$ is defined by

$$\varphi_c(v, \lambda) = \inf_{y \in H} \{\varphi(v - y) + (\lambda, y)_H + \tfrac{c}{2}|y|_H^2\} \tag{2.11}$$

for $(c, \lambda) \in R^+ \times H$. Here, $\varphi_c(v, \lambda)$ is a C^1 approximation of φ in the following sense [8].

Lemma 2.2. *For every $v, \lambda \in H$ the infimum in (2.11) is attained at a unique point $y_c(v, \lambda)$ and $\varphi_c(v, \lambda)$ is convex, (Lipschitz-) continuously Fréchet differentiable in v and*

$$\varphi_c'(v, \lambda) = \lambda + c\, y_c(v, \lambda).$$

$\square$

If $\lambda = 0$ then $\varphi_c(v, 0)$ is the proximal approximation of φ. Moreover, if $\lambda \in \partial\varphi(v)$ then $y_c(v, \lambda) = 0$ and $\varphi_c'(v, \lambda) = \lambda$. The proof of the lemma is based on the observation that (2.11) is equivalent to

$$\varphi_c(v, \lambda) = \inf_{\tilde{y} \in H} \{\varphi(\tilde{y}) + \tfrac{c}{2}|v + c^{-1}\lambda - \tilde{y}|^2\} - \tfrac{1}{2c}|\lambda|^2 \tag{2.11$'$}$$

where $\tilde{u} = v - u$. The following dual representation of $\varphi_c(v, \lambda)$ plays an important role in our discussion. Define the conjugate function φ^* of φ by

$$\varphi^*(v^*) = \sup_{v \in H} \{(v, v^*) - \varphi(v)\} \text{ for } v^* \in H.$$

Then, we have the following lemma [8].

Lemma 2.3. *For $x, \lambda \in H$*

$$\varphi_c(v, \lambda) = \sup_{y^* \in H} \{(v, y^*) - \varphi^*(y^*) - \tfrac{1}{2c}|y^* - \lambda|^2\} \tag{2.12}$$

where the supremum is attained at a unique point $\lambda_c(v, \lambda)$ and we have

$$\lambda_c(v, \lambda) = \lambda + cy_c(v, \lambda) = \varphi_c'(v, \lambda). \tag{2.13}$$

$\square$

In many of the applications including the image restoration problem, the conjugate function φ^* is given by

$$\varphi^*(y^*) = \chi_{C^*}(y^*), \tag{2.14}$$

where C^* is a closed convex set in H. Then, it follows from Lemma 2.3 that for $v, \lambda \in H$

$$\varphi_c(v, \lambda) = \sup_{y^* \in C^*} \{-\tfrac{1}{2c}|y^* - (\lambda + c\,v)|_H^2\} + \tfrac{1}{2c}(|\lambda + c\,v|_H^2 - |\lambda|_H^2).$$

Hence, the supremum is attained at

$$\varphi_c'(v, \lambda) = Proj_{C^*}(\lambda + c\,v) \tag{2.15}$$

where $Proj_{C^*}(\phi)$ denotes the projection of $\phi \in H$ onto C^*.

The following theorem provides an equivalent characterization of $\lambda \in \partial\varphi(x)$.

Theorem 2.4. *If* $\lambda \in \partial\varphi(v)$ *for* $v, \lambda \in H$, *then* $\lambda = \varphi'_c(v, \lambda)$ *for all* $c > 0$. *Conversely, if* $\lambda \in H$ *satisfies* $\lambda = \varphi'_c(v, \lambda)$ *for some* $c > 0$, *then* $\lambda \in \partial\varphi(v)$. $\qquad\square$

Proof. If $\lambda \in \partial\varphi(x)$ then for all $u \in H$

$$\varphi(x - u) - \varphi(x) \geq (\lambda, -u)$$

and thus

$$\varphi(x - u) + (\lambda, u) + \tfrac{c}{2}|u|^2 \geq \varphi(x).$$

Hence, the infimum of (2.11) is attained at $u_c(x, \lambda) = 0$ and it follows from Lemma 2.2 that $\lambda = \varphi'(x, \lambda)$. Conversely, if $\lambda \in H$ satisfies $\lambda = \varphi'(x, \lambda)$ for some $c > 0$, then $u_c(x, \lambda) = 0$ by Lemma 2.2. Hence, it follows from Lemma 2.2, (2.11), and Lemma 2.3 that

$$\varphi(x) = \varphi_c(x, \lambda) = (x, \lambda) - \varphi^*(\lambda).$$

Since $\lambda \in \partial\varphi(x)$ if and only if

$$\varphi(x) + \varphi^*(\lambda) = (\lambda, x)$$

(*e.g.*, see [6]), $\lambda \in \partial\varphi(x)$. $\qquad\square$

It follows from Theorem 2.4 that the complementarity condition $\lambda^* \in \partial\varphi(\Lambda x^*)$ can be equivalently expressed as

$$\lambda^* = \varphi'_c(\Lambda x^*, \lambda^*). \tag{2.16}$$

Thus, the optimality condition (2.6) is equivalently written as

$$f'(x^*) + \Lambda^*\lambda^* = 0 \text{ and } \lambda^* = \varphi'_c(\Lambda x^*, \lambda^*), \tag{2.17}$$

for some and equivalently all $c > 0$.

The multiplier λ^* in (2.6, 2.17) is the Lagrange multiplier in the following sense. Define the Lagrangian $L : X \times H \to R$ by

$$L(x, \lambda) = f(x) + (\Lambda x, \lambda) - \varphi^*(\lambda).$$

Then, (2.2) is equivalent to the min-max problem

$$\min_{x \in X} \max_{\lambda \in H} L(x, \lambda)$$

and we have the following result [6, 9].

Theorem 2.5. *The pair* $(x^*, \lambda^*) \in X \times H$ *satisfies (2.6) (equivalently, (2.17)) if and only if* (x^*, λ^*) *is a saddle point of* L, *i.e.,*

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$$
$$\text{for all } x \in X, \lambda \in H.$$

$$\square$$

For the image restoration problem we have

$$\varphi^*(v) = g\,\chi_{\mathcal{C}}(v)$$

where $\mathcal{C}$ is defined by (2.7) and from (2.12, 2.15)

$$\varphi_c(v, \lambda) = g \int_{\Omega\setminus A} \left(\tfrac{1}{c}(|\lambda + c\,v| - 1) + \tfrac{1}{2c}(1 - |\lambda|^2)\right) dx + \tfrac{g}{2c} \int_A (|\lambda + c\,v|^2 - |\lambda|^2)\, dx,$$

where $A = \{x : |\lambda(x) + c\,v(x)| \le 1\}$, and

$$\varphi_c'(v, \lambda) = g \frac{\lambda(x) + c\,v(x)}{\max(1, |\lambda(x) + c\,v(x)|)}, \quad \text{a.e. } x \in \Omega.$$

The necessary and sufficient optimality condition (2.17) is given by

$$-\mu\Delta u + u - g \operatorname{div} \lambda = z$$

$$\lambda(x) = \frac{(\lambda + c\nabla u)(x)}{\max(1, |\lambda + c\nabla u|(x))} \tag{2.18}$$

for some and equivalently every $c > 0$.

2.3. First order algorithms

In this section we present the first order iterative algorithms for the necessary and sufficient optimality condition (2.17). It follows from (2.3) that given $\lambda \in H$ the first equation of (2.17)

$$f'(x) + \Lambda^*\lambda = 0$$

has a unique solution $x(\lambda) \in X$. Then (2.17) can be written as $\lambda^* = \varphi_c'(\Lambda x(\lambda^*), \lambda^*)$. That is, $\lambda^* \in H$ is a fixed point of $\Phi(\lambda) = \varphi_c'(\Lambda x(\lambda), \lambda)$. In this setting the Uzawa algorithm [6] is the fixed point iteration given by

$$\lambda_{k+1} = \varphi_c'(\Lambda x(\lambda_k), \lambda_k). \tag{2.19}$$

In particular the Uzawa algorithm is an explicit iteration method for solving $\Phi(\lambda) = \lambda$. It is proved in [9] that the Uzawa algorithm converges provided that there exists a Lagrange multiplier λ^* satisfying (2.6) and $2c\sigma - c^2\|\Lambda\|^2 > 0$.

The first order augmented Lagrangian method [3, 8] can be written as the fixed point iteration

$$\lambda_{k+1} = \varphi_c'(\Lambda x_{k+1}, \lambda_k) \tag{2.20a}$$

where $x_{k+1} \in X$ satisfies

$$f'(x_{k+1}) + \Lambda^*\lambda_{k+1} = 0. \tag{2.20b}$$

The first order augmented Lagrangian method is implicit and it is proved in [8] that differently from the Uzawa algorithm it converges unconditionally (in $c > 0$) provided that there exists a Lagrange multiplier λ^* satisfying (2.6). For the image restoration problem the Uzawa algorithm is expressed as

$$-\mu\Delta u_k + u - g \operatorname{div} \lambda_k = z$$

$$\lambda_{k+1} = \frac{(\lambda_k + c\nabla u_k)(x)}{\max(1, |\lambda + c\nabla u_k|(x))}. \tag{2.21}$$

3. Discretized problem and active set strategy

Consider the discretization of (2.1) by the finite difference approximation on $\Omega = (0,1) \times (0,1)$:

$$
\begin{aligned}
\min \quad & \sum_{0 < i,\, j < n} \frac{\mu}{2} \quad \left(\left|\frac{u_{i,j} - u_{i-1,j}}{h}\right|^2 + \left|\frac{u_{i+1,j} - u_{i,j}}{h}\right|^2 \right. \\
& \left. + \left|\frac{u_{i,j} - u_{i,j-1}}{h}\right|^2 + \left|\frac{u_{i,j+1} - u_{i,j}}{h}\right|^2\right) + \frac{1}{2}\, |u_{i,j} - z_{i,j}|^2 \\
& + \sum_{0 < i,\, j < n} g \left(\left|\tfrac{u_{i,j} - u_{i-1,j}}{h}\right|^2 + \left|\tfrac{u_{i,j} - u_{i,j-1}}{h}\right|^2\right)^{1/2}
\end{aligned}
\tag{3.1}
$$

over $u_{i,j}, 0 < i < n, 0 < j < n$, where $h = \frac{1}{n}, u_{i,j} = u(ih, jh), u_{0,j} = u_{n,j} = 0$ and $u_{i,0} = u_{i,n} = 0$. Numerical experiments showed that the discretization of $|\nabla u|$ by one-sided finite differences gives significantly better results than by central finite differences. We can write (3.1) as

$$
\min \frac{\mu}{2}\, u^t H u + \frac{1}{2}|u - z|^2 + g|Du|_1 \text{ over } u \in X = R^m
\tag{3.2}
$$

where $m = (n-1)^2, u = \text{col}\,(u_{\cdot,1}, u_{\cdot,2}, \cdots, u_{\cdot,n-1}) \in R^m$. The matrices H, D are defined by $H = H_0 \otimes I + I \otimes H_0$ and

$$
D = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} \text{ with } D_1 = D_0 \otimes I \text{ and } D_2 = I \otimes D_0.
$$

The matrix product $\otimes$ denotes the Kroneker product. The matrix $H_0 \in R^{n-1,n-1}$ is symmetric tri-diagonal with $(H_0)_{i,i} = \frac{2}{h^2}, (H_0)_{i,i-1} = (H_0)_{i-1,i} = -\frac{1}{h^2}$, the matrix D_0 is a lower triangular matrix with $(D_0)_{i,i} = \frac{1}{h}$ and $(D_0)_{i,i-1} = -\frac{1}{h}$ and zero otherwise and I is the $(n-1) \times (n-1)$ identity matrix. Further we use the following notations

$$
|Du|_1 = \sum_i |(Du)_i|, \text{ where } (Du)_i = \begin{pmatrix} (D_1 u)_i \\ (D_2 u)_i \end{pmatrix},
$$

$$
|(Du)_i| = \sqrt{(D_1 u)_i^2 + (D_2 u)_i^2}
$$

and

$$
\mathcal{C}^m = \{\lambda = (\lambda_1, \lambda_2) \in R^{2m} : |((\lambda_1)_i, \lambda_2)_i)|_{R^2} \le 1, 1 \le i \le m\}.
$$

Throughout this section $|x|$ denotes the Euclidean norm of the vector x, $\cdot$ denotes the dot product in R^2 and $(\cdot)_i$ is an element in R^2.

It is possible to consider more general costfunctionals $f(u)$ provided that f is C^2 and there exists a $\sigma > 0$ such that

$$
f''(u)(h,h) \ge \sigma|h|^2 \text{ for all } u, h \in X.
$$

In particular, in this section we do not require $\mu > 0$, but $\mu \ge 0$ suffices.

In (3.2)

$$
f(u) = \tfrac{\mu}{2} u^t H u + \tfrac{1}{2}|u - z|^2.
$$

The necessary and sufficient optimality condition for the optimality pair (u^*, λ^*) for (3.2) is given by

$$\mu Hu + u + gD^t\lambda = z \text{ and } \lambda_i = \frac{\lambda_i + c(Du)_i}{\max(1, |\lambda_i + c(Du)_i|)}, 1 \leq i \leq m \tag{3.3}$$

for every $c > 0$. Again the second condition is equivalent to

$$\lambda_i \cdot (Du)_i = |(Du)_i| \text{ and } |\lambda_i| \leq 1$$

for all i and

$$\lambda_i = \frac{(Du)_i}{|(Du)_i|} \quad \text{if} \quad |(Du)_i| \neq 0.$$

Hence, (3.3) is equivalent to

$$\mu\,Hu + u + g\,D^t\lambda = z, \quad \begin{cases} \lambda_i &= \frac{(Du)_i}{|(Du)_i)|} &\text{for} \quad i \in I, \\ (Du)_j &= 0 &\text{for} \quad j \in J. \end{cases} \tag{3.4}$$

where the index sets $I = \{i : |(Du)_i| \neq 0\}, J = \{j : |(Du)_j| = 0\}$.

3.1. Active set strategy

Both the Uzawa algorithm and the first order augmented Lagrangian method are of first order. In order to develop a Newton-like (formally second order) iterate for the nonsmooth minimization problem (3.2) we propose an "active set strategy" which is based on the dual variable λ. It is motivated by the work in [3,8]. From (3.4) it follows that

$$\lambda_i = \frac{(Du)_i}{|(Du)_i|}, \quad i \in I^* = (J^*)^c \tag{3.5}$$

where J^* is the active (index) set defined by

$$J^* = \{j : |(Du)_j| = 0\}$$

at the minimizer $u = u^*$. This terminology is suggested by inequality constraints of the type $u \leq 0$, in which the complementarity conditions $(u, \lambda)_{\mathbb{R}^m} = 0$ and $\lambda \geq 0$ imply that on the active set $\{j : u_j = 0\}, \lambda_j \geq 0$ is unknown and otherwise $\lambda = 0$. The active set strategy that we propose involves the update of the active index set J according to the complementarity condition

$$\lambda_i = \frac{\lambda_i + c\,(Du)_i}{\max\,(1, |\lambda_i + c\,(Du)_i|)}, \quad 1 \leq i \leq m,$$

i.e.,

$$J^k = \{j : |\lambda_j^k + c\,(Du^k)_j| \leq 1\} \quad \text{and} \quad I^k = \{i : |\lambda_i^k + c\,(Du^k)_i| > 1\}, \tag{3.6}$$

where (u^k, λ^k) is the solution update. Define the matrices $B_1^k = (D_1)_{i,\cdot}, B_2^k = (D_2)_{i,\cdot}$ for $i \in I^k$ and $C_1^k = (D_1)_{j,\cdot}, C_2^k = (D_2)_{j,\cdot}$ for $j \in J^k$ and

$$B^k = \begin{pmatrix} B_1^k \\ B_2^k \end{pmatrix} \quad \text{and} \quad C^k = \begin{pmatrix} C_1^k \\ C_2^k \end{pmatrix}.$$

Then, we consider the minimization problem

$$\min J(u) = f(u) + g\,|Du\,|_1 \quad \text{subject to} \quad C^k u = 0 \tag{3.7}$$

which is equivalent to

$$\min f(u) + g\,|\,B^k u\,|_1 \text{ subject to } C^k u = 0.$$

Here we use the same notation for the matrices B^k, C^k as for the matrix D. In the statement and proof of the following lemma we suppress the index k in the notation of B^k and C^k.

Lemma 3.1. *The minimization (3.7) has a unique solution u and there exist λ satisfying*

$$\lambda_i \cdot (Bu)_i = |(Bu)_i| \ \text{ and } \ |\lambda_i| \leq 1 \ \text{for } i \in I^k \tag{3.8}$$

and $\mu \in range\ (C)$ such that

$$f'(u) + g\,B^t \lambda + g\,C^t \mu = 0. \tag{3.9}$$

$\square$

Proof. First, note that the necessary and sufficient optimality condition for (3.7) is given by

$$f'(u)(h) + g\,|B(u + h)|_1 - g\,|Bu|_1 \geq 0 \quad \text{for all} \quad h \in \ker(C). \tag{3.10}$$

The C^1 approximation $\varphi_c(v, 0)$ of $|v|$ in the sense of (2.11) is given by

$$\varphi_c(v, 0) = \begin{cases} |v| - \frac{1}{2c} & \text{if} \quad |v| \geq \frac{1}{c} \\[2mm] \frac{c}{2}\,|v|^2 & \text{if} \quad |v| < \frac{1}{c}. \end{cases}$$

Consider the regularized problem

$$\min J_c(u) = f(u) + g\,\tilde{\varphi}_c(Bu, 0) \quad \text{subject to} \quad Cu = 0, \tag{3.11}$$

where $\tilde{\varphi}_c(Bu, 0) = \sum_{i \in I} \varphi_c((Bu)_i, 0)$. The necessary and sufficient optimality condition for (3.11) is given by

$$f'(u_c)(h) + g\,B^t \tilde{\varphi}'_c(Bu_c, 0)(h) = 0 \quad \text{for all} \quad h \in \ker(C).$$

Since $X = range\ (C^t) + \ker(C)$, this implies that there exists a unique $\mu_c \in range\ (C)$ such that

$$f'(u_c) + g\,B^t \tilde{\varphi}'_c(Bu_c, 0) + g\,C^t \mu_c = 0. \tag{3.12}$$

From (3.10)

$$0 \geq J_c(u_c) - J_c(u) \geq f(u_c) - f(u) - f'(u)(u_c - u)$$

$$+g\,(\tilde{\varphi}_c(Bu_c, 0) - \tilde{\varphi}_c(Bu, 0) - |Bu_c| + |Bu|).$$

Since

$$\varphi_c((Bu)_i, 0) - |(Bu)_i| \leq 0$$

we have

$$\varphi_c((Bu_c)_i, 0) - |(Bu_c)_i| - (\varphi_c((Bu)_i, 0) - |(Bu)_i|) \geq -\tfrac{1}{2c}$$

and thus

$$\frac{\sigma}{2}|u_c - u|^2 \leq \frac{g}{2c} \to 0 \quad \text{as} \quad c \to \infty.$$

If we define

$$(\lambda_c)_i = \varphi_c'((Bu_c)_i, 0) = \frac{c\,(Bu_c)_i}{\max(1, |c\,(Bu_c)_i|)}$$

then $|(\lambda_c)_i| \leq 1$ and

$$(\lambda_c)_i \cdot (Bu_c)_i = \begin{cases} |(Bu_c)_i| & \text{if} \quad |(Bu_c)_i| \geq \tfrac{1}{c} \\[2mm] c\,|(Bu_c)_i|^2 & \text{if} \quad |(Bu_c)_i| \leq \tfrac{1}{c}. \end{cases}$$

From (3.12) we deduce that $|\mu_c|$ is uniformly bounded in $c > 0$. Hence, it follows from (3.12) that for any cluster point (λ, μ) of the sequence $\{(\lambda_c, \mu_c)\}_{c>0}$ (3.8, 3.9) are satisfied. $\qquad\square$

The second order update method of (u^k, λ^k) based on the active set strategy is given as follows.

Algorithm 1.
Step 1: Choose $c > 0, u^0 \in \mathbb{R}^m$ and $\lambda^0 \in \mathbb{R}^{2m}$ satisfying $|\lambda_i^0| \leq 1$ and set $k = 0$.
Step 2: Set the index sets $J^k = \{j : |(\lambda^k + c\,Du^k)_j| \leq 1\}$ and $I^k = \{i : |(\lambda^k + c\,Du^k)_i| > 1\}$. Let (u, λ, μ) be the solution triple (as defined in Lemma 3.1) to

$$\min f(u) + g\,|B^k u|_1 \text{ subject to } C^k u = 0.$$

Step 3: Set $u^{k+1} = u$, $\lambda^{k+1} = \lambda$ on I^k and $\lambda^{k+1} = \mu$ on J^k and $k = k + 1$. Go to Step 2.
In the following theorem we establish the fact that the algorithm converges in finitely many steps. $\qquad\square$

Theorem 3.2. *Suppose that the iterates generated by the Algorithm satisfy $u^{k+1} \neq u^k$ if $J^k \neq J^{k-1}$ for $k \geq 1$. Then the algorithm converges in finitely many steps to the solution of (3.2).* $\qquad\square$

Proof. If $J^k = J^{k+1}$ for some $k \geq 0$, then $|\mu_i| \leq 1$ for all $i \in J^k$. Since by Lemma 3.1 $|\lambda_i^{k+1}| \leq 1$ for all $i \in I^k$ it follows that (u^{k+1}, λ^{k+1}) satisfies the optimality condition (3.4) and u^{k+1} is the minimizer of (3.2). It remains to show that $J^k \neq J^{k+1}$ for all k is impossible.

We show that $u^k \in \ker(C^k)$ for every $k \geq 1$. This claim is equivalent to $(Du^k)_j = 0$ for all $j \in J^k$ and $k \geq 1$. If $j \in J^{k-1}$ then $(Du^k)_j = 0$ by definition of u^k. Suppose that $(Du^k)_j \neq 0$ for some $j \in I^{k-1}, j \in J^k$. Then, see (3.3),

$$|(\lambda^k + c\,B^k u^k)_j| = |\frac{(Du^k)_j}{|(Du^k)_j|} + c\,(Du^k)_j| > 1$$

and therefore $j \in I^k$ which is impossible.

Next we show that $J(u^k)$ is strictly decreasing as long as $J^{k-1} \neq J^k$. Since $u^k, u^{k+1} \in \ker(C^k)$ for every $k \geq 1$, it follows from (3.8, 3.9) that

$$J(u^k) - J(u^{k+1}) = f(u^k) - f(u^{k+1}) - f'(u^{k+1})(u^k - u^{k+1})$$
$$+g\left(|B^k u^k|_1 - |B^k u^{k+1}|_1 - (\lambda^{k+1}, B^k u^k - B^k u^{k+1})\right) \geq \tfrac{\sigma}{2}|u^k - u^{k+1}|^2,$$

which by assumption implies that $J(u^k)$ is strictly decreasing as long as $J^{k-1} \neq J^k$. Since there are only finitely many combinations of possible active indices this completes the proof. $\qquad\square$

Let us turn to giving a sufficient condition for the assumption of Theorem 3.2. From the proof of Theorem 3.2 it follows that $u^k \in \ker(C^{k-1}) \cap \ker(C^k)$. If $u^k = u^{k+1}$ then the minimization of $J(u)$ over the two different subspaces $\ker(C^{k-1})$ and $\ker(C^{k-1}) \cap \ker(C^k)$ have the same solution. A sufficient condition that this cannot happen and that the assumption of Theorem 3.2 holds is that D is surjective. We give the details of this statement in the following lemma. D is surjective in the case of the voice detection problem which is the one dimensional version of (3.2) and is described in Section 3.3 below.

Lemma 3.3. *Suppose D is surjective. Then $J^{k-1} \neq J^k$ implies $u^k \neq u^{k+1}$ for every $k \geq 1$.* $\qquad\square$

Proof. Suppose that $j \in J^{k-1}$ and $j \in I^k$ for some $k \geq 1$. This implies that $(Du^k)_j = 0$ and $|\mu_j| > 1$. Then, for h satisfying $(Dh)_j = \mu_j$ and $(Dh)_i = 0, i \neq j$ we find using (3.9, 3.10) that

$$J(u^k + t\,h) - J(u^k) = f(u^k + t\,h) - f(u^k) - t\,f'(u^k)(h) + gt\left(|\mu_j| - |\mu_j|^2\right) < 0$$

for appropriately chosen $t > 0$, since $|\mu_j| > 1$. It follows that $u^{k+1} \neq u^k$. $\qquad\square$

3.2. Modified algorithm

The implementation of the Algorithm for the active set strategy is as difficult as that for the original minimization. Thus, we propose the following modification of the Algorithm 1 which can be readily implemented.

Algorithm 2 (implemented).
Step 1: Choose $c > 0, \epsilon > 0, u^0 \in \mathbb{R}^m$ and $\lambda^0 \in \mathbb{R}^{2m}$ satisfying $|\lambda_i^0| \leq 1$ and set $k = 0$.
Step 2: Set the index sets $J^k = \{j : |(\lambda^k + c\,Du^k)_j| \leq 1\}$ and $I^k = \{i : |(\lambda^k + c\,Du^k)_i| > 1\}$. Define the corresponding matrices $B = D_{i,\cdot}, i \in I^k$ and $C = D_{j,\cdot}, j \in J^k$ and solve

$$\left(\mu\,H + g\left(B^t A(u^k)B + \frac{1}{\epsilon}C^tC\right)\right)u + u = z - g\,B^t\frac{(Bu^k)_i}{|(Bu^k)_i|} \tag{3.13}$$

for $u \in R^m$, where the nonzero elements of the matrix $A = A(u^k)$ are given by

$$\begin{pmatrix} A_{i,i} & A_{i,i+l} \\ A_{i+l,i} & A_{i+l,i+l} \end{pmatrix} = \frac{1}{|(Bu^k)_i|^3}\begin{pmatrix} (q_i)_2^2 & -(q_i)_1(q_i)_2 \\ -(q_i)_1(q_i)_2 & (q_i)_1^2 \end{pmatrix} =: A_i$$

for $1 \leq i \leq l$ with $l = $ cardinality of the index set I^k and $q_i = (Bu^k)_i$. Then set

$$\lambda_i = \frac{(Bu^k)_i}{\max(\epsilon, |(Bu^k)_i|)} \quad \text{for} \quad i \in I^k$$

$$\lambda_j = \frac{1}{\epsilon}(Cu)_j \qquad \text{for} \quad j \in J^k.$$

Step 3: Set $u^{k+1} = u, \lambda_i^{k+1} = \dfrac{\lambda_i}{\max(1, |\lambda_i|)}$ and $k = k + 1$. Go to Step 2. $\qquad\square$

In what follows we describe the derivation of the modified algorithm and analyze its convergence properties. Algorithm 2 (implemented) is based on treating the constraint $Cu = 0$ in (3.7) by the penalty method. Thus for $\epsilon > 0$ we consider the unconstrained problem

$$\min J_\epsilon(u) = f(u) + g\,|Bu|_1 + \frac{g}{2\epsilon}|Cu|^2. \tag{3.14}$$

It can be proved using the same argument as in the proof of Theorem 2.4 and Lemma 3.1 that the necessary and sufficient optimality condition for (3.14) is given by

$$f'(u_\epsilon) + g\, B^t \lambda_\epsilon + \frac{g}{\epsilon} C^t C u_\epsilon = 0 \tag{3.15}$$

and

$$(\lambda_\varepsilon)_i \cdot (Bu_\epsilon)_i = |(Bu_\epsilon)_i| \quad \text{and} \quad |(\lambda_\varepsilon)_i| \le 1 \quad \text{for} \quad i \in I^k. \tag{3.16}$$

Equation (3.13) is derived for the case when $(Bu_\epsilon)_i \ne 0$ for all $i \in I^k$. In this case

$$(\lambda_\epsilon)_i = \frac{(Bu_\epsilon)_i}{|(Bu_\epsilon)_i|},$$

and (3.13) is one step of the Newton method applied to (3.15). Here we use the fact that $B^t A(u^k) Bu^k = 0$.

In the following results we justify Algorithm 2 (implemented) as far as possible. The ultimate test will be its numerical behavior.

Lemma 3.4. *If for a given index pair (I^k, J^k) u is the solution to (3.7) (see, Lemma 3.1), then $u_\varepsilon \to u$ as $\epsilon \to 0^+$ and for any cluster point λ of the sequence $\{\lambda_\epsilon\}$ and μ of the sequence $\{\mu_\epsilon\}$ defined by*

$$\mu_\epsilon = \frac{1}{\epsilon} C u_\epsilon, \tag{3.17}$$

(3.8, 3.9) are satisfied. □

Proof. Since

$$0 \ge J_\epsilon(u_\epsilon) - J_\epsilon(u) = f(u_\epsilon) - f(u) - f'(u)(u_\epsilon - u)$$
$$+ g\left(|Bu_\epsilon|_1 - |Bu|_1 - (\lambda, Bu_\epsilon - Bu)\right) + \frac{g}{2\epsilon} |Cu_\epsilon|^2 - g\left(\mu, Cu_\epsilon\right)$$

we have

$$\frac{\sigma}{2} |u_\epsilon - u|^2 + \frac{g}{2\epsilon} |Cu_\epsilon|^2 \le g\left(\mu, Cu_\epsilon\right)$$

which implies that

$$\frac{\sigma}{2} |u_\epsilon - u|^2 + \frac{g\epsilon}{2} |\mu_\epsilon - \mu|^2 \le \frac{g\epsilon}{2} |\mu|^2.$$

Thus $|u_\epsilon - u| = O(\sqrt{\epsilon})$ and $|\mu_\epsilon|$ is bounded uniformly in $\epsilon > 0$. The last assertion follows from (3.15, 3.16) by taking the limit $\epsilon \to 0$ for any clustering sequence. □

Corollary 3.5. *Suppose that $|(Bu)_i| \ne 0$ for all $i \in I^k$. Then λ and $\mu \in$ range C in Lemma 3.1 are unique and $(\lambda_\epsilon, \mu_\epsilon) \to (\lambda, \mu)$.* □

Proof. It follows from Lemma 3.4 that $|(Bu_\epsilon)_i| \ne 0$ for all $i \in I^k$ provided that ϵ is sufficiently small. Thus,

$$\lambda_\epsilon = \frac{Bu_\epsilon}{|Bu_\epsilon|} \to \lambda = \frac{Bu}{|Bu|}$$

as $\epsilon \to 0$. Hence the corollary follows from (3.9, 3.15). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

In this section we henceforth assume that at the minimizer u^* of (3.2) the Lagrange multiplier λ^* satisfies the condition:

$$\{\lambda_j^*\}_{j \in J^*} \in range\,(C) \quad \text{and} \quad |\lambda_j^*| < 1 \text{ for } j \in J^*, \qquad\qquad (3.18)$$

where C corresponds to the index set J^*. The second part of (3.18) is referred to as the strict complementarity condition. Due to (3.4) a multiplier λ^* that satisfies the range condition is necessarily unique. The range condition is satisfied if D is surjective. Moreover, if the active set strategy of the Algorithm 1 converges in finitely many steps then the range condition is satisfied. With (3.18) holding we have the following proposition.

Proposition 3.6. *Suppose C corresponds to J^*; i.e., $J^* = J^k$. Then, there exists a constant $\epsilon_0 > 0$ such that if $\epsilon \le \epsilon_0$ then $|\mu_\epsilon| < 1$.* $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof. First, note that the minimizer u^* of (3.2) is the solution to (3.7) with C corresponding to J^*. It follows from Lemma 3.4 and Corollary 3.5 that $u_\epsilon \to u^*$ and μ_ϵ converges to λ^* on J^*, and thus there exist an $\epsilon_0 > 0$ such that $\epsilon \le \epsilon_0$ for $|\mu_\epsilon| < 1$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Corollary 3.5 and Proposition 3.6 imply that the penalty treatment (3.13) of the constraint $Cu = 0$ in (3.7) maintains the property of the active set J^*.

By construction $|(Bu^k)_i| \neq 0$ for $i \in I^k$ and hence the modified algorithm is well-posed. Furthermore the sequence generated by the modified algorithm is uniformly bounded. Note that the matrix

$$\mu\,H + g\left(B^t A(u^k)B + \frac{1}{\epsilon}\,C^t C\right)$$

has the same sparsity as H does. The following proposition supports the second order convergence property of Algorithm 2 (implemented).

Proposition 3.7. *Assume that $0 < \epsilon \le \epsilon_0$ and let u_ϵ be the solution to (3.14) where C corresponds to J^* as in Proposition 3.6. Suppose that $J^k = J^*$. Then there exist $\delta > 0$ and $\bar{c} > 0$ such that for $0 < c \le \bar{c}$ if $|u^k - u_\epsilon| \le \delta$ then $J^{k+1} = J^*$ and*

$$|u^{k+1} - u_\epsilon| \le \min\left(\delta, M\,|u^k - u_\epsilon|^2\right)$$

for some $M > 0$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof. Note that

$$F(u) = \mu\,Hu + u - z + g\,B^t\,\frac{Bu}{|Bu|} + \frac{g}{\epsilon}\,C^t Cu$$

is locally C^2 at u_ϵ and that (3.13) can equivalently be written as

$$F'(u^k)(u^{k+1} - u_\epsilon) = F(u_\epsilon) - F(u^k) - F'(u^k)(u_\epsilon - u^k).$$

Thus, there exist constants $\delta_0 > 0$ and $M > 0$ such that if $|u^k - u_\epsilon| \le \delta_0$ then

$$|u^{k+1} - u_\epsilon| \le M\,|u^k - u_\epsilon|^2.$$

We choose $\delta \le \delta_0$ such that $\delta \le 1/M$ and $M\delta^2\,|C| < \epsilon\,(1 - |\mu_\epsilon|)$, which is possible due to Proposition 3.6. Then $|u^{k+1} - u_\epsilon| \le \delta$ for $|u^k - u_\epsilon| \le \delta$ and there exists $\eta \in [0,1)$ such that

$$|\frac{1}{\epsilon}\,(Cu^{k+1})_j| \le \eta \quad \text{for all } j \in J^*.$$

Moreover δ can be chosen sufficiently small such that

$$(Bu^k)_i \cdot (Bu^{k+1})_i > 0 \quad \text{for all } i \in I^*.$$

From the above two inequalities it follows that there exists $\bar{c} > 0$ such that $J^{k+1} = J^*$ for all $c \in (0, \bar{c}]$. □

In the following proposition it is asserted that under the assumption of convergence of the iterates $\{u^k\}$ of Algorithm 2 (implemented) and with strict complementarity holding, the limit is the solution to the penalized problem (3.14) corresponding to the true active set J^*.

Proposition 3.8. *Suppose that for some $\varepsilon > 0$ the sequence $\{u^k\}$ of Algorithm 2 (implemented) converges to $\hat{u}$. Then $\hat{u}$ is the solution to*

$$\min_u J_{1/\varepsilon}(u) = f(u) + g\tilde{\varphi}_{1/\varepsilon}(Du, 0),$$

where $\tilde{\varphi}_{1/\varepsilon}(Du, 0) = \sum_i \varphi_{1/\varepsilon}((Du)_i, 0)$. Moreover, if ε is sufficiently small and strict complementarity (in the sense of the second condition in (3.18)) holds, then $\hat{u} = u_\varepsilon$ with u_ε as in Proposition 3.7. □

Proof. Suppose that $|(D\hat{u})_j| < \epsilon$. Then there exists k_1 such that for $k \geq k_1$ $j \in J^k$. If, on the other hand $|(D\hat{u})_i| > \epsilon$, then there exists k_2 such that $i \in I^k$ for $k \geq k_2$. Finally suppose that $|(D\hat{u})_\ell| = \epsilon$. Then the index ℓ can belong to either I^k or J^k. But, for either case we have

$$\lambda_\ell^{k+1} \to \frac{1}{\epsilon}(D\hat{u})_\ell = \frac{(D\hat{u})_\ell}{|(D\hat{u})_\ell|}.$$

For $c = \frac{1}{\epsilon}$ consider the cost functional

$$J_c(u) = f(u) + g\,\tilde{\varphi}_c(Du, 0) \tag{3.19}$$

where $\varphi_c(v, \lambda)$ is the C^1 approximation of $|v|_1$. Hence the limit $\hat{u}$ satisfies

$$J_c'(\hat{u}) = 0. \tag{3.20}$$

Using the optimality condition for $\hat{u}$ and for u_ε from Proposition 3.7, one can argue that $\hat{u} = u_\varepsilon$ provided that $\varepsilon > 0$ is sufficiently small and that strict complementarity holds. □

3.3. **One dimensional problem**

Consider the one-dimensional problem on the interval $(0, 1)$

$$\min \int_0^1 \tfrac{\mu}{2}|u_x|^2 + \tfrac{1}{2}|u(x) - z(x)|^2 \, dx + g\int_0^1 |u_x(x)|\, dx.$$

Then, the corresponding discretized problem is given by

$$\min \tfrac{\mu}{2}u^t H_0 u + \tfrac{1}{2}|u - z|_2^2 + g|D_0 u|_1 \quad \text{over} \quad u \in R^{n-1},$$

and the necessary and sufficient optimality condition is given by

$$\mu\, H_0 u + u + g\, D_0^t \lambda = z \quad \text{and} \quad \lambda_i = \frac{\lambda_i + c\,(D_0 u)_i}{\max(1, |\lambda_i + c\,(D_0 u)_i|)}.$$

Note that D_0 is surjective. Hence, in view of Lemma 3.3, Theorem 3.2 is applicable. The algorithm corresponding to Algorithm 2 (implemented) is given as follows.

Algorithm 3.
Step 1: Choose $c > 0$, $\epsilon > 0$, u^0 and λ^0 satisfying $|\lambda_i^0| \leq 1$ and set $k = 0$.
Step 2: Set the index sets $J^k = \{j : |(\lambda^k + c\,D_0 u^k)_j| \leq 1\}$ and $I^k = \{i : |(\lambda^k + c\,D_0 u^k)_i| > 1\}$. Define the matrices $B = (D_0)_{i,\cdot}$, $i \in I^k$ and $C = (D_0)_{j,\cdot}$, $j \in J^k$. Let u be a solution to

$$\mu\, H_0 u + u + \frac{g}{\epsilon}\, C^t C u = z - g\, B^t\, \frac{(Bu^k)_i}{|(Bu^k)_i|}$$

and let

$$\lambda_i = \frac{(Bu^k)_i}{\max(\epsilon, |(Bu^k)_i|)} \quad \text{for} \quad i \in I^k$$

$$\lambda_j = \frac{1}{\epsilon}\, (Cu)_j \qquad \text{for} \quad j \in J^k.$$

Step 3: Set $u^{k+1} = u$, $\lambda_i^{k+1} = \dfrac{\lambda_i}{\max(1, |\lambda_i|)}$ and $k = k + 1$. Go to Step 2.   □

4. Numerical experiments

Numerical experiments are carried out to demonstrate the efficiency of the proposed algorithm for the image denoising and the voice detection problem. For the two-dimensional results that we report on here, the domain Ω was chosen to be the unit square with 100×100 pixels. The noisy image z was produced by adding uniformly distributed random numbers from the interval $[-\delta, \delta]$ to the noise-free image at the nodal points. Here we only show results with high relative noise. In order to ensure global convergence we consider the merit function $J_c(\cdot)$ defined by (3.19) with $c = \frac{1}{\epsilon}$. We modify Step 3 of the proposed algorithm by

Step 3′: Find w such that $J_c(\tilde{u}) < J_c(u^k)$, $\tilde{u} = (1 - w)\, u^k + w\, u$ and set $u^{k+1} = \tilde{u}$, λ^{k+1} as in Step 3 and $k = k + 1$. Go to Step 2.

The following choices were made for the parameters appearing in the algorithm:

$$\mu = 10^{-10},\ c = 0.1,\ w = 0.5,$$

with $g = 0.02$ for example 4.1 and $g = 0.01$ for example 4.2.

The typical number of iterations was 8. The code was written in MATLAB and the runs made on IBM RISC 6000 workstation. Upon request the code can be made available.

Example 4.1. *Here we consider the reconstruction of two square humps of height one and two lying inside each other. The noise was $\delta = 1.5$. Figure 1 shows the noisy image and a corresponding histogram, Figure 2 gives the reconstructed image.*   □

Example 4.2. *The second image to be reconstructed consists of flat as well as oblique surfaces. Figures 3 and 4 show the noise free and noisy image, respectively. Figure 5 shows the reconstructed image. The choice of g was a heuristic one. It can be observed that choosing g too large smears the corners, whereas undesirable oscillations may occur if g is too small.*   □

Example 4.3. *This is a one dimensional problem with the noise free signal z^* given by*

$$z^*(\frac{i}{n}) = \begin{cases} 2 & \text{for } 201 \leq i \leq 300 \text{ or } 651 \leq i \leq 800 \\ 1 & \text{for } 500 \leq i \leq 650 \\ 0 & \text{otherwise,} \end{cases}$$

$n = 1000$ *and* $\delta = 2$. *The choices for the parameters in the one dimensional example were*

$$\mu = 10^{-10}, \; g = 0.02, \; c = 10^{-4}, \; w = 0.$$

The result after 12 iterations is shown in Figure 6, where the solid line represents z^*, *the broken line represents* z *and the stars depict the reconstruction* u^{12}. *Decreasing the value of* c *slows down the convergence, whereas increasing* c *can cause the solution to chatter.*   □

Concerning the update of the Lagrange multiplier λ in Algorithm 2 (implemented) the Newton step would be of the form

$$\lambda_i = \frac{(Bu^k)_i}{|(Bu^k)_i|} + A_i(Bu)_i \quad \text{for } i \in I^k. \tag{4.1}$$

We gave our preference to the update stated in Algorithm 2 (implemented) since (4.1) does not enhance the numerical results and requires a more restrictive choice of the relaxation parameter w. For the one dimensional problem of Section 3.3 both the updates coincide.

References

[1] L. Alvarez, P.L. Lions and J.M. Morel, Image selective smoothing and edge detection by nonlinear diffusion, II. *SIAM J. Numer. Anal.* **29** (1992) 845-866.

[2] R. Acar and C.R. Vogel, Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems* **10** (1994) 1217-1229.

[3] D.P. Bertsekas, *Constraint Optimization and Lagrange Multiplier Methods*. Academic Press, Paris (1982).

[4] F. Catte, P.L. Lions, J.M. Morel and T. Colle, Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.* **29** (1992) 182-193.

[5] D.C. Dobson and F. Santosa, Recovery of blocky images from noisy and blurred data. Preprint.

[6] I. Ekeland and T. Turnbull, *Infinite Dimensional Optimization and Convexity*. The University of Chicago Press, Chicago (1983).

[7] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*. Birkhäuser, Boston (1984).

[8] K. Ito and K. Kunisch, Augmented Lagrangian methods for nonsmooth convex optimization in Hilbert spaces. To appear in *Nonliner Analysis, Theory, Methods and Applications*.

[9] K. Ito and K. Kunisch, Augmented Lagrangian Formulation of Nonsmooth, Convex Optimization in Hilbert Spaces, in *Lecture Notes in Pure and Applied Mathematics. Control of Partial Differential Equations and Applications*, E. Casas. Marcel Dekker Eds., **174** (1995) 107-117.

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

[11] C.R. Vogel, Total variation regularization for ill-posed problems. Technical Report, Department of Mathematical Sciences, Montana State University.
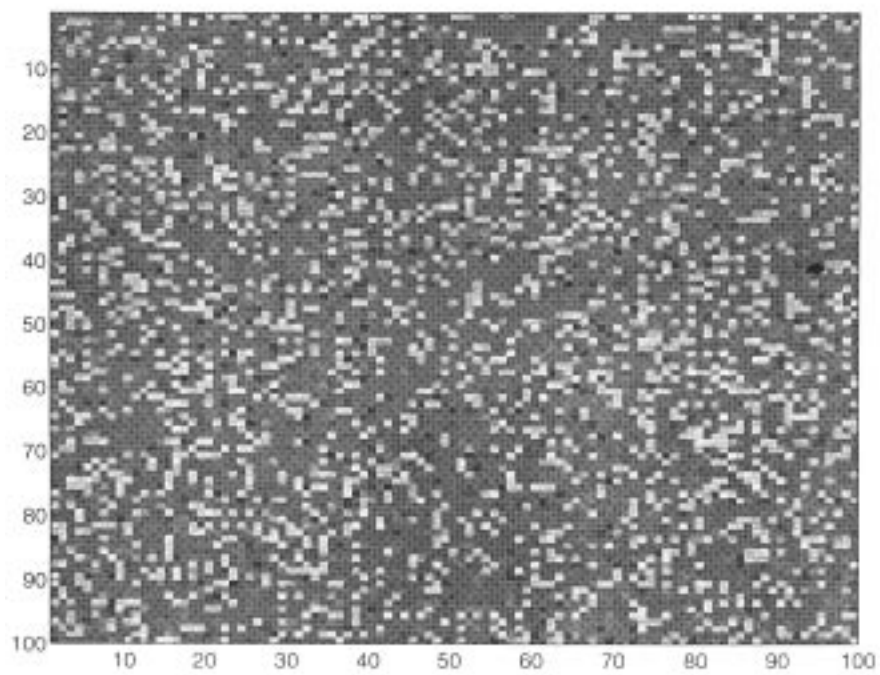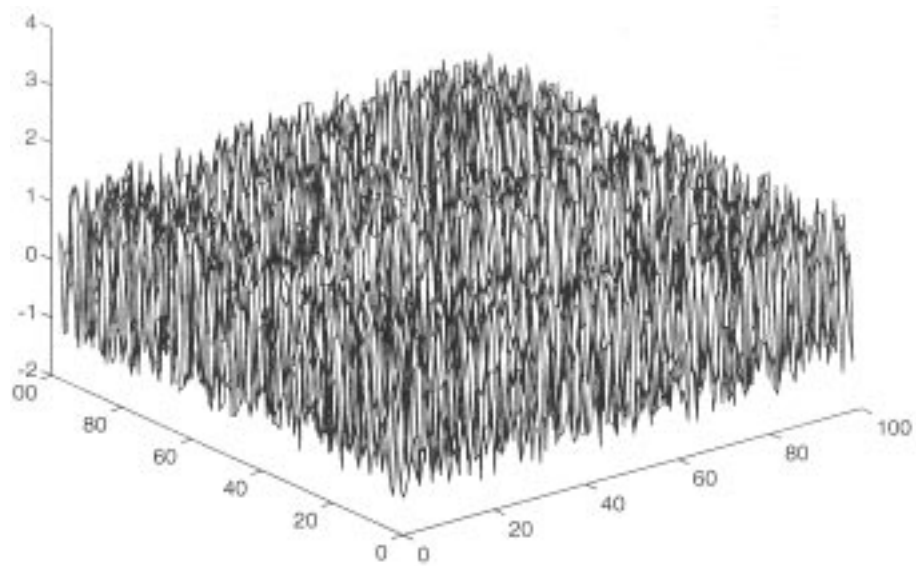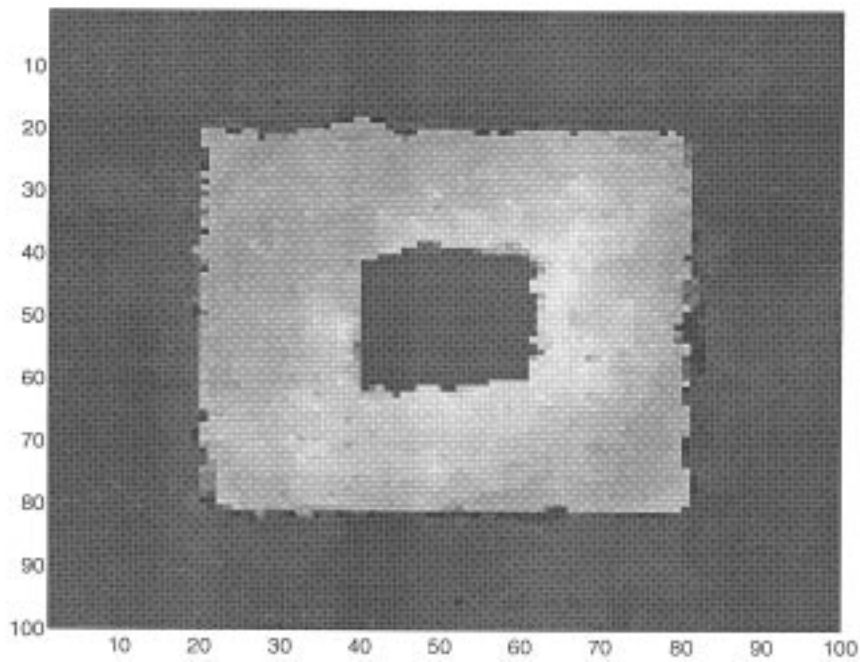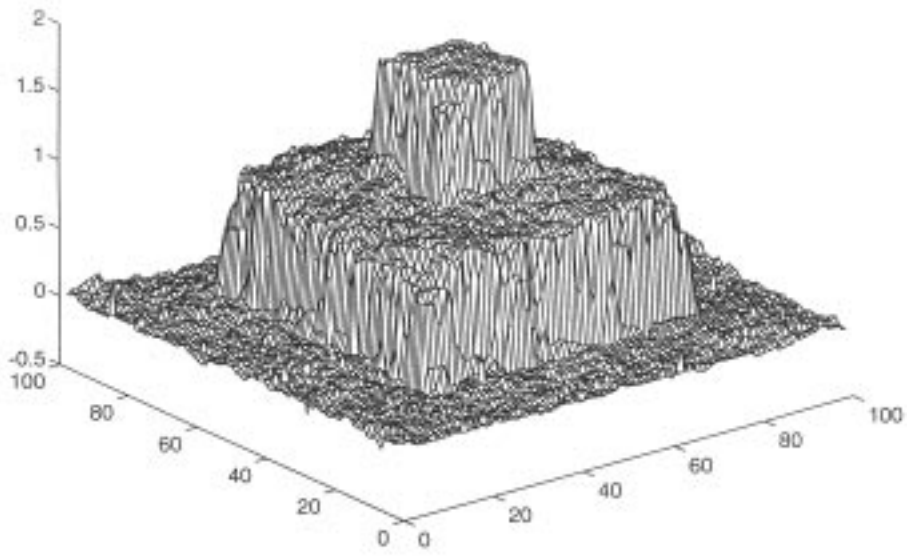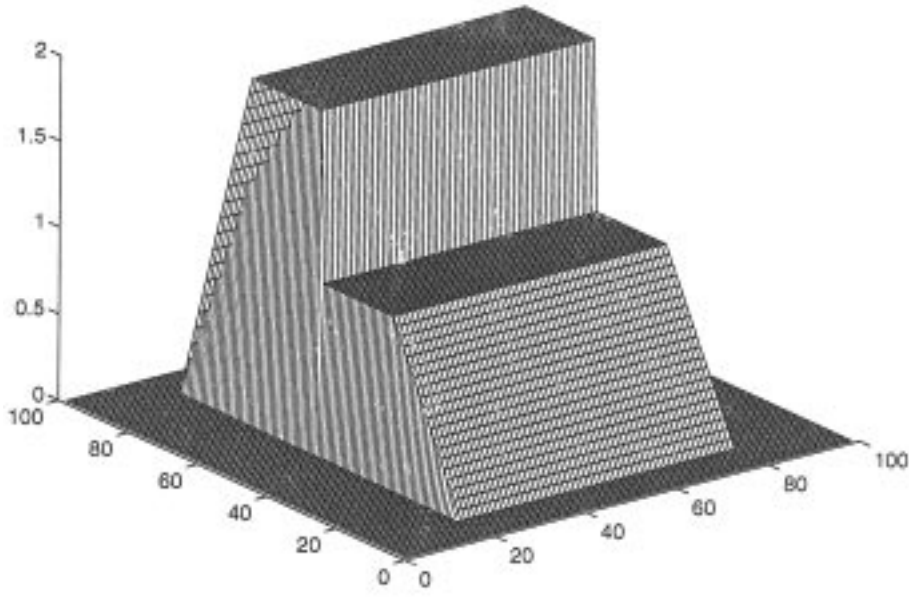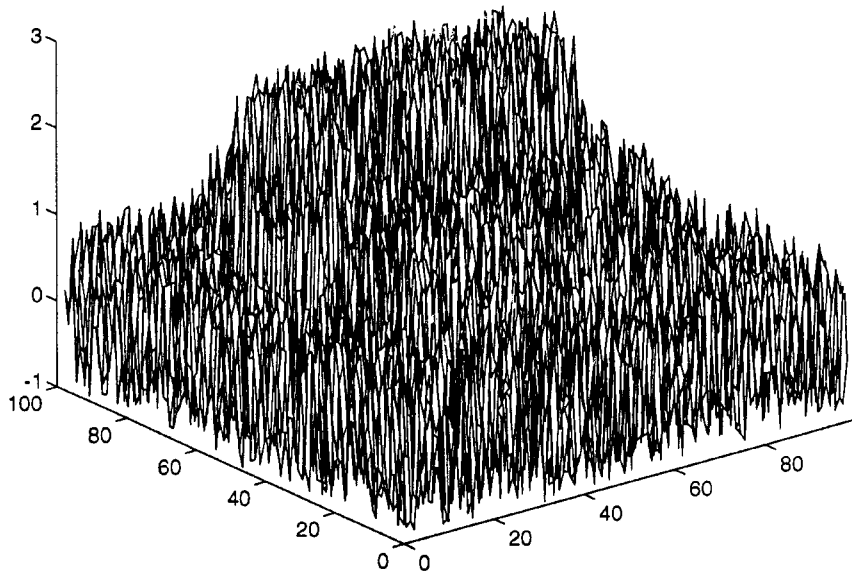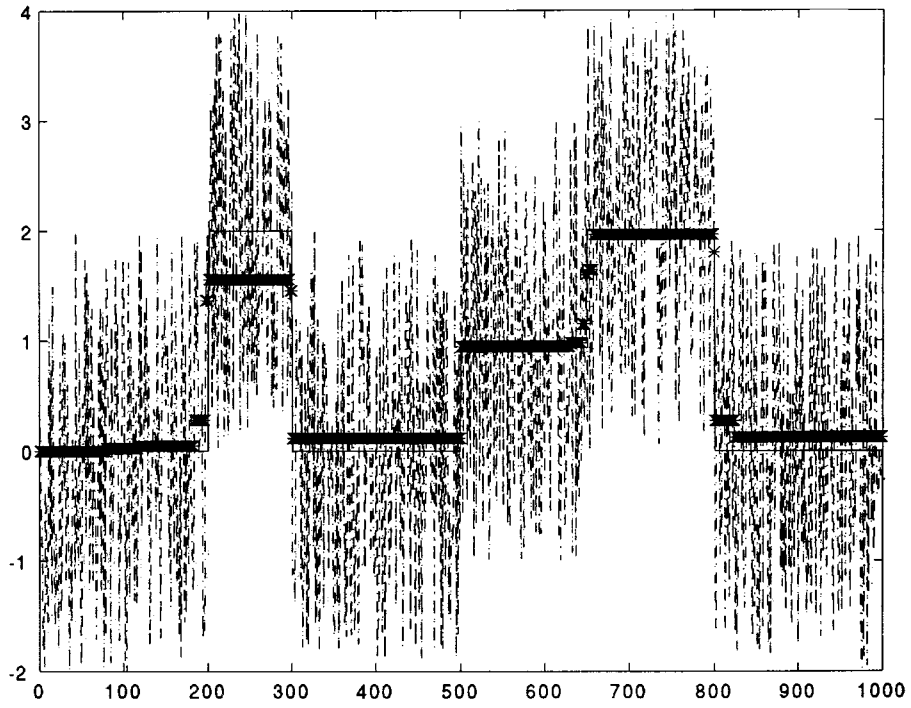
FIGURE 1

FIGURE 2

FIGURE 3

FIGURE 4

FIGURE 5

FIGURE 6