A NEW MEDIAN FORMULA WITH APPLICATIONS TO PDE BASED DENOISING[*]

YINGYING LI[†] AND STANLEY OSHER[‡]

Abstract. We develop a simple algorithm for finding the minimizer of the function $E(x) = \sum_{i=1}^{n} w_i |x - a_i| + F(x)$, when the w_i are nonnegative and F is strictly convex. If F is also differentiable and F' is bijective, we obtain an explicit formula in terms of a median. This enables us to obtain approximate solutions to certain important variational problems arising in image denoising. We also present a generalization with $E(x) = J(x) + F(x)$ for $J(x)$ a convex piecewise differentiable function with a finite number of nondifferentiable points.

Key words. Convex optimization, ℓ_1 minimization, TV denoising, Bregman iterative method.

AMS subject classifications. 46N10, 94A08.

1. Introduction

Given $a_1, \ldots, a_n \in \mathbb{R}$, it is well-known that

$$\min_{x \in \mathbb{R}} \sum_{i=1}^{n} |x - a_i|^\alpha = \begin{cases} \text{mean}(a_i) & \text{if } \alpha = 2 \\ \text{median}(a_i) & \text{if } \alpha = 1 \\ \text{mode}(a_i) & \text{if } \alpha = 0. \end{cases}$$

More generally, the very early work of Barral [14] investigated

$$\min_{x \in \mathbb{R}} \sum_{i=1}^{n} w_i |x - a_i|^\alpha, \quad w_i \geq 0, \alpha = 0, 1, 2, \infty.$$

Also, [15, 16] developed local M-estimator filters based on such minimizers.

This work was inspired by two variational problems arising in image research. One is soft wavelet thresholding [2, 8] or basis pursuit [3] arising in compressed sensing. The other is the Rudin-Osher-Fatemi (ROF) model [1] of TV-based image denoising and generalizations. The first involves reducing $E(x) = \sum_{i=1}^{n} w_i |x - a_i| + F(x)$ to its simplest terms: the scalar problem

$$\min_{x \in \mathbb{R}} E(x) = |x| + \lambda(x - f)^2, \quad \lambda > 0. \tag{1.1}$$

The solution to (1.1) is obtained from a simple formula: $x_{opt} = \text{shrink}(f, \frac{1}{2\lambda})$, where the shrink operator can be found in figure 1.1

It turns out that

$$\text{shrink}(f, \tfrac{1}{2\lambda}) = \text{median}\big\{ f - \tfrac{1}{2\lambda}, 0, f + \tfrac{1}{2\lambda} \big\}. \tag{1.2}$$

This seems (surprisingly) to be a new result. We generalize it below.

$$\text{shrink}(f,\mu) = \begin{cases} f-\mu, & \text{if } f > \mu; \\ 0, & \text{if } -\mu \leq f \leq \mu; \\ f+\mu, & \text{if } f < -\mu. \end{cases}$$
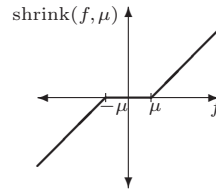
Fig. 1.1. *Shrinkage*

The ROF model [1] for image denoising is the following: find u satisfying

$$u = \arg\min_u \int_\Omega |\nabla u|\, dx + \lambda \|f-u\|_2^2. \tag{1.3}$$

Recently very fast methods for solving the discrete approximation were obtained using graph cuts [6, 12] and split Bregman techniques [10]. If we take a discrete approximation to the anisotropic ROF model with $|\nabla u| = |u_x| + |u_y|$ (or similar generalizations), and fix all the discrete values except one, then we are minimizing the function $E(u)$ which is defined as (2.1), where $F(u) = \lambda(f-u)^2$. We will discuss this procedure below.

2. The main theorem

We consider the following minimization problem:

$$\arg\min_{x\in\mathbb{R}} E(x) = \sum_{i=1}^n w_i |x-u_i| + F(x). \tag{2.1}$$

For notational convenience, denote by W_i the sums

$$W_i = -\sum_{j=1}^i w_j + \sum_{j=i+1}^n w_j, \qquad i = 0,1,\ldots,n$$
$$= (-w_1 - \cdots - w_i) + (w_{i+1} + \cdots + w_n).$$

THEOREM 2.1. *Suppose the w_i are nonnegative, the u_i are sorted as $u_1 \leq u_2 \leq \cdots \leq u_n$, the function F is strictly convex and differentiable, and F' is bijective. Then the minimizer of* (2.1) *is a median:*

$$x_{opt} = \text{median}\{u_1,\ldots,u_n,p_0,\ldots,p_n\}, \tag{2.2}$$

where $p_i = (F')^{-1}(W_i)$.

Proof. By the strict convexity of F, E is strictly convex and hence (2.1) has a unique minimizer. Also by strict convexity, F' and $(F')^{-1}$ are increasing. Therefore, since $W_n \leq \cdots \leq W_1 \leq W_0$, the p_i satisfy

$$p_n \leq p_{n-1} \leq \cdots \leq p_i \leq p_{i-1} \leq \cdots \leq p_0. \tag{2.3}$$

We will prove the theorem by discussing the following two cases.

Case 1: Suppose that the minimizer x_{opt} of (2.1) is in (u_i, u_{i+1}). Consider
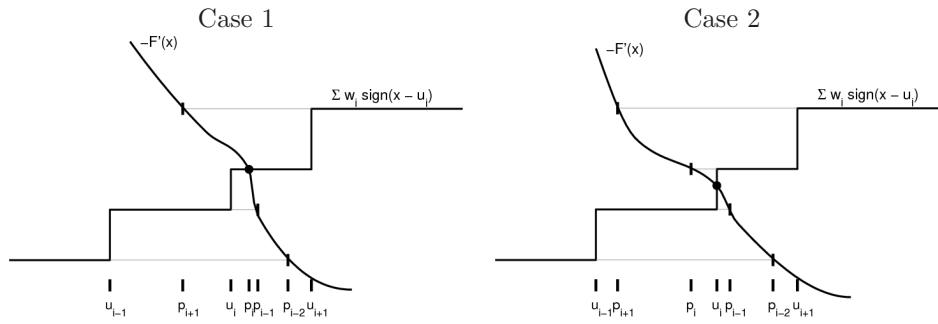
$$\tilde{E}(x) = -W_i x + F(x) + C, \tag{2.4}$$

FIG. 2.1. *The minimizer is the intersection of* $-F'$ *and* $\sum_i w_i \operatorname{sign}(x_{opt} - u_i)$, *and it is the median* (2.2).

where C is a constant, then the associated Euler-Lagrange equation of (2.4) is

$$\partial \tilde{E} = -W_i + F'(x) = 0. \tag{2.5}$$

For $x \in (u_i, u_{i+1})$, functions $\tilde{E}$ and E are equivalent, so x_{opt} is also a local minimizer of $\tilde{E}$. But $\tilde{E}$ is strictly convex, so x_{opt} is necessarily the unique minimizer. Thus

$$x_{opt} = (F')^{-1}(W_i) = p_i. \tag{2.6}$$

Now we have $u_i < p_i = x_{opt} < u_{i+1}$. So

$$\underbrace{u_1 \le \cdots \le u_i}_{i \text{ numbers}} < p_i < \underbrace{u_{i+1} \le \cdots \le u_n}_{(n-i) \text{ numbers}}. \tag{2.7}$$

From the inequality (2.3), we have the ordering:

$$\underbrace{p_n \le \cdots \le p_{i+1}}_{(n-i) \text{ numbers}} \le p_i \le \underbrace{p_{i-1} \le \cdots \le p_0}_{i \text{ numbers}}. \tag{2.8}$$

Altogether, there are n numbers $(u_1, \ldots, u_i$ and $p_n, \ldots, p_{i+1})$ less than or equal to p_i and n numbers $(u_{i+1}, \ldots, u_n$ and $p_{i-1}, \ldots, p_0)$ greater than or equal to p_i, so the minimizer p_i is the median (2.2). Similarly, the same argument applies for $x_{opt} < u_1$ or $x_{opt} > u_n$ with the notations $u_0 = -\infty$ and $u_n = +\infty$.

Case 2: Suppose that the minimizer x_{opt} is equal to u_i, where u values may repeat:

$$\cdots u_{i-1} < x_{opt} = u_i = u_{i+1} = \cdots = u_j < u_{j+1} \cdots$$

Consider

$$\tilde{E}(x) = -W_j x + F(x) + C, \tag{2.9}$$

then we obtain

$$\arg\min_x \tilde{E}(x) = (F')^{-1}(W_j) = p_j.$$

If $p_j \ne u_j$, then both of u_j and u_{j+1} are greater than p_j or less than p_j. Because $E(x) = \tilde{E}(x)$ when $x \in [u_j, u_{j+1}]$ and u_j is the minimizer of $E(x)$, we obtain $\tilde{E}(u_j) = E(u_j) < E(u_{j+1}) = \tilde{E}(u_{j+1})$. Hence, we have

$$\tilde{E}(p_j) < \tilde{E}(u_j) < \tilde{E}(u_{j+1}). \tag{2.10}$$

Thus $p_j < u_j < u_{j+1}$, and we have $p_j \le u_j < u_{j+1}$.
Similarly, we can prove $u_{i-1} < u_i \le p_{i-1}$.
We have

$$\underbrace{u_1 \le \cdots \le u_{i-1}}_{(i-1) \text{ numbers}} \le u_i = \underbrace{u_{i+1} = \cdots = u_j < u_{j+1} \le \cdots \le u_n}_{(n-i) \text{ numbers}}, \qquad (2.11)$$

$$\underbrace{p_n \le p_{n-1} \le \cdots \le p_j \le u_j = \cdots = u_{i+1}}_{(n-i+1) \text{ numbers}} = u_i \le \underbrace{p_{i-1} \le \cdots \le p_0}_{i \text{ numbers}}. \qquad (2.12)$$

There are $(i-1)+(n-i+1)=n$ numbers less than or equal to u_i and $(n-i)+i=n$ numbers greater than or equal to u_i. Therefore, u_i is the median (2.2).   □

We now generalize Theorem 2.1 by replacing the ℓ^1 term with an arbitrary convex piecewise differentiable function.

THEOREM 2.2. *Consider*

$$\min_x J(x) + F(x). \qquad (2.13)$$

Suppose J is convex and piecewise differentiable with finitely many nondifferentiable points $u_1 < u_2 < \cdots < u_n$, and F is strictly convex. Then the unique solution to (2.13) is a median, as defined in (2.15).
    Define

$$J(x) = \begin{cases} J_0(x), & \text{when } x \in (-\infty, u_1), \\ J_i(x), & \text{when } x \in (u_i, u_{i+1}), \\ J_n(x), & \text{when } x \in (u_n, \infty). \end{cases} \qquad (2.14)$$

Let $\tilde{J}_i$ be a differentiable and convex extension of J_i to the whole domain $\mathbb{R}$ such that the $\tilde{J}_i$ satisfy $\tilde{J}_i' \le \tilde{J}_{i+1}'$. Denote $p_i = \arg\min_{x \in \mathbb{R}} \left(\tilde{J}_i(x) + F(x) \right)$ (here, p_i is allowed to be infinite), then the solution to (2.13) is

$$x_{opt} = \text{median} \{u_1, u_2, \ldots, u_n, p_0, p_1, \ldots, p_n\}. \qquad (2.15)$$

    Proof. There is a simple way to extend J_i to satisfy all the requirements. When $x \in (u_i, u_{i+1})$, J_i is differentiable and convex, so J_i' is continuous and increasing. Denote $a = \lim_{x \to u_i^+} J_i'(x)$ and $b = \lim_{x \to u_{i+1}^-} J_i'(x)$, then define

$$\tilde{J}_i = \begin{cases} ax + c, & \text{when } x \in (-\infty, u_i], \\ J_i, & \text{when } x \in (u_i, u_{i+1}), \\ bx + c', & \text{when } x \in [u_{i+1}, \infty), \end{cases} \qquad (2.16)$$

where c and c' are computed as to ensure continuity of J. It is easy to show that $\tilde{J}_i$ is convex and differentiable in $\mathbb{R}$ and

$$\tilde{J}_i' = \begin{cases} a, & \text{when } x \in (-\infty, u_i], \\ J_i', & \text{when } x \in (u_i, u_{i+1}), \\ b, & \text{when } x \in [u_{i+1}, \infty). \end{cases} \qquad (2.17)$$

$J'(x)$ is increasing in $\mathbb{R}\backslash S$, so $\lim_{x\to u_i^+} J'(x) \le \lim_{x\to u_{i+1}^+} J'(x)$ and $\lim_{x\to u_i^-} J'(x) \le \lim_{x\to u_{i+1}^-} J'(x)$. Thus $\tilde{J}_i' \le \tilde{J}_{i+1}'$.

If p_i is the minimizer of $(\tilde{J}_i + F)$, then $(\tilde{J}_i' + F')(p_i) = 0$. So $(\tilde{J}_{i+1}' + F')(p_i) \ge (\tilde{J}_i' + F')(p_i) = 0$, then $p_{i+1} \le p_i$ since the function $(\tilde{J}_{i+1}' + F')$ is increasing. Thus we have the following inequality:

$$p_n \le p_{n-1} \le \cdots \le p_1 \le p_0. \tag{2.18}$$

The remainder of the proof is the same as in the previous theorem. □

In numerical implementation, the bottleneck computation in formula (2.2) is sorting the u_i and finding the median value among all the numbers in the median formula. We can make the computation cheaper by taking advantage of the structure of the numbers in the median formula. In particular, we know that

$$p_0 \ge p_1 \ge \cdots \ge p_{n-1} \ge p_n,$$
$$u_1 \le u_2 \le \cdots \le u_{n-1} \le u_n.$$

THEOREM 2.3. *Suppose that the u_i are sorted, then the following algorithm will find*

$$x_{opt} = \mathrm{median}\{u_1, u_2, \ldots, u_n, p_0, p_1, \ldots, p_n\}$$

with an additional $2n$ comparisons. If we define $T(p,q,u) = \min\{p, \max\{q,u\}\}$ and let $T_1 = T(p_0, p_1, u_1)$ and $T_m = T(T_{m-1}, p_m, u_m)$, then

$$\mathrm{median}\{u_1, \ldots, u_n, p_0, \ldots, p_n\} = T_n.$$

Proof.   For $n=1$, since $p_1 \le p_0$, we have

$$\mathrm{median}\{u_1, p_0, p_1\} = \begin{cases} p_1 & \text{if } u_1 \le p_1 \\ u_1 & \text{if } p_1 \le u_1 \le p_0 \\ p_0 & \text{if } p_0 \le u_1 \end{cases}$$
$$= T(p_0, p_1, u_1) = T_1.$$

Suppose that the formula holds for $n=m$, then we want to show it is true for $n = (m+1)$. Since $p_{m+1} \le p_m \le \cdots \le p_0$, then p_{m+1} is less than $(m+1)$ numbers in the set $\{u_1, \ldots, u_m, p_0, \ldots, p_m\}$, necessarily

$$p_{m+1} \le T_m = \mathrm{median}\{u_1, \ldots, u_m, p_0, \ldots, p_m\}.$$

Now we want to show that

$$\mathrm{median}\{u_1, \ldots, u_{m+1}, p_0, \ldots, p_{m+1}\}$$
$$= \mathrm{median}\{\mathrm{median}\{u_1, \ldots, u_m, p_0, \ldots, p_m\}, p_{m+1}, u_{m+1}\}$$
$$= \mathrm{median}\{T_m, p_{m+1}, u_{m+1}\}.$$

We sort $\{u_1, \ldots, u_m, p_0, \ldots, p_m\}$ and denote its elements as $s_1 \ge \cdots \ge s_m \ge T_m \ge t_1 \ge \cdots \ge t_m$. Consider the set

$$S = \{u_1, \ldots, u_{m+1}, p_0, \ldots, p_{m+1}\} = \{s_1, \ldots, s_m, T_m, t_1, \ldots, t_m, u_{m+1}, p_{m+1}\},$$

we have $t_m \leq u_m \leq u_{m+1}$, so t_m is less than $(m+1)$ elements in S. We also have $s_1 \geq p_m \geq p_{m+1}$, so s_1 is also bigger than $(m+1)$ elements in S. Thus, in finding the median of S, we can remove s_1 and t_m together,

$$\text{median}\{u_1, \ldots, u_{m+1}, p_0, \ldots, p_{m+1}\} = \text{median}\{s_2, \ldots, s_m, T_m, t_1, \ldots, u_{m+1}, p_{m+1}\}.$$

Inductively, we can remove pairs $\{s_2, t_{m-1}\}, \{s_3, t_{m-2}\}, \ldots, \{s_m, t_1\}$ without changing the median of S. That means

$$\text{median}\{u_1, \ldots, u_{m+1}, p_0, \ldots, p_{m+1}\} = \text{median}\{T_m, p_{m+1}, u_{m+1}\}$$
$$= T(T_m, p_{m+1}, u_{m+1}) = T_{m+1}. \qquad \square$$

3. Applications

3.1. Solving ROF. Let $F(x) = \lambda |x - f|^\alpha$ with $\alpha > 1$, $\lambda > 0$. Then the minimizer of (2.1) is found as follows:

$$\arg\min_x \sum_{i=1}^n w_i |x - u_i| + \lambda |x - f|^\alpha$$
$$= \text{median}\{u_1, u_2, \ldots, u_n, f + |w_n + w_{n-1} + \cdots + w_1|^p \mu,$$
$$f + \text{sign}(w_n + w_{n-1} + \cdots + w_2 - w_1)|w_n + w_{n-1} + \cdots + w_2 - w_1|^p \mu,$$
$$f + \text{sign}(w_n + w_{n-1} + \cdots + w_3 - w_2 - w_1)|w_n + w_{n-1} + \cdots + w_3 - w_2 - w_1|^p \mu,$$
$$\ldots, f + \text{sign}(w_n - w_{n-1} - \cdots - w_1)|w_n - w_{n-1} - \cdots - w_1|^p \mu,$$
$$f - |w_n + w_{n-1} + \cdots + w_1|^p \mu\}, \tag{3.1}$$

where $p = \frac{1}{\alpha - 1}$ and $\mu = (\frac{1}{\alpha \lambda})^p$. A computational savings is that the p_i do not depend on the u_i, so with all else fixed they only need to be computed once.

As $\lambda \searrow 0$, we can replace μ by infinity, which gives us an expression for

$$x_{opt} = \arg\min_x \sum_{i=1}^n w_i |x - u_i|. \tag{3.2}$$

We get the solution by simply counting occurrences of $+\infty$ and $-\infty$ terms on the right of (3.1) and arrange all terms in increasing order.

Formula (3.1) is very useful in certain denoising problems. For example, returning to the anisotropic ROF model with $|\nabla u| = |\nabla_x u| + |\nabla_y u|$, we can try to solve a discrete approximation on a Cartesian grid point by point. Let the unknown value at a fixed point be u, and its upper, lower, left and right values to be u_u, u_d, u_l and u_r (refer to figure 3.1). We fix all the pixel values except u and minimize:

$$\min_{u \in \mathbb{R}} E(u) = |u - u_l| + |u - u_r| + |u - u_u| + |u - u_d| + \lambda(u - f)^2. \tag{3.3}$$

We have

$$u_{opt} = \arg\min_{u \in \mathbb{R}} E(u) = \text{median}\left\{u_\ell, u_r, u_u, u_d, f + \frac{2}{\lambda}, f + \frac{1}{\lambda}, f, f - \frac{1}{\lambda}, f - \frac{2}{\lambda}\right\}. \tag{3.4}$$

We apply the median formula (3.4) pixel-by-pixel over the whole image until convergence. During the iteration process, since a pixel directly affects only its four neighbors, we can apply the median formula in parallel on multiple pixels at one time.
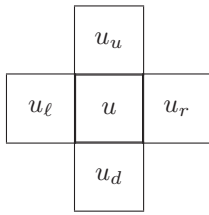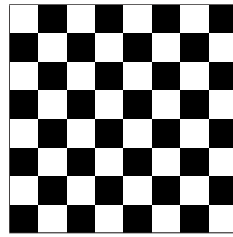
FIG. 3.1. u_l, u_r, u_u, u_d and u.

FIG. 3.2. *After dividing pixels into two groups by their color, pixels are in the same group are not neighbors.*

We divide the image into two groups in a checkerboard pattern such that pixels of the same group are not neighbors (see figure 3.2). Then we can update all pixels in one group at one time while keeping the pixels in the other group constant.

The algorithm is as follows:

while not converged,

   apply (3.4) to pixels in the black pattern;

   apply (3.4) to pixels in the white pattern.

For the numerical implementation, we use the difference between the current solution and the last solution to test convergence.

LEMMA 3.1. *Consider function* $E(x) = \sum_{i=1}^{n} |x - a_i| + \lambda(x - f)^2$, *where* $x \in \mathbb{R}$, $a_1 \le a_2 \le \cdots \le a_n$. *According to Theorem 2.1,*

$$x_{opt} = \arg\min E(x) = \text{median}\left\{a_1, a_2, \ldots, a_n, f + \frac{n}{2\lambda}, f + \frac{n-2}{2\lambda}, \ldots, f - \frac{n}{2\lambda}\right\}, \quad (3.5)$$

then, we claim $E(x) - E(x_{opt}) \ge \lambda|x - x_{opt}|^2$ *for any* x.

Proof. First, observe that

$$(x - f)^2 - (x_{opt} - f)^2 = (x - x_{opt})^2 + 2(x_{opt} - f)(x - x_{opt}). \quad (3.6)$$

Suppose $a_k \le x_{opt} \le a_{k+1}$ (or similarly, $x_{opt} \le a_1$ or $a_n \le x_{opt}$). Then x_{opt} satisfies the Euler-Lagrange equation

$$0 \in \sum_{i=1}^{n} \text{sign}(x_{opt} - a_i) + 2\lambda(x_{opt} - f). \quad (3.7)$$

Case 1: Suppose that $a_k < x_{opt} < a_{k+1}$. Then $\text{sign}(x_{opt} - a_i)$ can only be 1 or -1, so we have

$$k - (n - k) + 2\lambda(x_{opt} - f) = 0$$
$$2\lambda(x_{opt} - f) = n - 2k.$$

Therefore, applying (3.6),

$$E(x) - E(x_{opt})$$

$$= \lambda(x - x_{opt})^2 + 2\lambda(x_{opt} - f)(x - x_{opt}) + \sum_{i=1}^{n} |x - a_i| - \sum_{i=1}^{n} |x_{opt} - a_i|$$

$$= \lambda(x - x_{opt})^2 + (n - 2k)(x - x_{opt}) + \sum_{i=1}^{n} |x - a_i| - \sum_{i=1}^{k} (x_{opt} - a_i) + \sum_{i=k+1}^{n} (x_{opt} - a_i)$$

$$\geq \lambda(x-x_{opt})^2 + (n-2k)(x-x_{opt}) + \sum_{i=1}^{k}\{(x-a_i)-(x_{opt}-a_i)\}$$

$$+ \sum_{i=k+1}^{n}\{(a_i-x)-(a_i-x_{opt})\}$$

$$= \lambda(x-x_{opt})^2 + (n-2k)(x-x_{opt}) + (k-(n-k))(x-x_{opt})$$

$$= \lambda(x-x_{opt})^2.$$

Case 2: If x_{opt} is one a_i, that is, it satisfies $a_l < a_{l+1} = \cdots = a_m = x_{opt} < a_{m+1}$ $(l < m)$, then

$$\text{sign}(x_{opt}-a_i) = \begin{cases} 1 & \text{if } i \leq l, \\ [-1,1] & \text{if } l+1 \leq i \leq m, \\ -1 & \text{if } i > m. \end{cases} \tag{3.8}$$

According to the Euler-Lagrange equation, we have

$$0 \in l + (m-l)\text{sign}(x_{opt}-a_m) - (n-m) + 2\lambda(x_{opt}-f). \tag{3.9}$$

So for some $p = \text{sign}(x_{opt}-a_m) \in [-1,1]$,

$$2\lambda(x_{opt}-f) = (n-m-l) - (m-l)p.$$

$$F(x)-F(x_{opt})$$

$$= \lambda(x-x_{opt})^2 + 2\lambda(x_{opt}-f)(x-x_{opt}) + \sum_{i=1}^{n}|x-a_i| - \sum_{i=1}^{n}|x_{opt}-a_i|$$

$$= \lambda(x-x_{opt})^2 + ((n-m-l)-(m-l)p)(x-x_{opt}) + \sum_{i=1}^{n}|x-a_i| - \sum_{i=1}^{n}|x_{opt}-a_i|$$

$$\geq \lambda(x-x_{opt})^2 + ((n-m-l)-(m-l)p)(x-x_{opt})$$

$$+ \sum_{i=1}^{l}[(x-a_i)-(x_{opt}-a_i)] + (m-l)|x-x_{opt}| + \sum_{i=m+1}^{n}[(a_i-x)-(a_i-x_{opt})]$$

$$= \lambda(x-x_{opt})^2 + (x-x_{opt})((n-m-l)-(m-l)p) + (l-(n-m))(x-x_{opt})$$

$$+ (m-l)|x-x_{opt}|$$

$$= \lambda(x-x_{opt})^2 - (m-l)(x-x_{opt})p + (m-l)|x-x_{opt}|$$

$$\geq \lambda(x-x_{opt})^2 \qquad (\text{since } -1 \leq p \leq 1).$$

Hence we conclude that $F(x)-F(x_{opt}) \geq \lambda|x-x_{opt}|^2$ for any x. □

THEOREM 3.2. *The algorithm defined by repeatedly applying (3.4) converges,* $u_j^{(k+1)} = \text{argmin}_{u_j \in \mathbb{R}} E^{(k)}(u_j)$,

$$u^{(k)} \to \text{argmin}_{u} E(u).$$

Proof. First, we denote the candidate solution at the kth step as $u^{(k)}$. At the kth step, we update only the jth coordinate, $u_j^{(k)}$. From (3.4), we have

$u_j^{(k+1)} = \arg\min_{u_j \in \mathbb{R}} E^{(k)}(u_j)$, so $E^{(k)}(u_j^{(k+1)}) \leq E^{(k)}(u_j^{(k)})$. This implies that the energy over all pixels decreases, $E(u^{(k+1)}) \leq E(u^{(k)})$, and since it is also bounded from below, the sequence $\big(E(u^{(k)})\big)$ converges. From Lemma 3.1, we have $|u^{(k)} - u^{(k+1)}|_\infty \leq \big(\frac{E(u^{(k)}) - E(u^{(k+1)})}{\lambda}\big)^{1/2}$, so $\big(u^{(k)}\big)$ converges. $\qquad\qquad\square$

REMARK 3.1. Lemma 3.1 can easily be generalized to $E(x) = \sum_{i=1}^n w_i |x - a_i| + \lambda(x - f)^2$, where $w_i \geq 0$, so we have the convergence for the weighted TV denoising and non-local median denoising algorithms, which will be introduced in the following sections.

Figure 3.3 compares the proposed algorithm with the graph-cuts method of Darbon and Sigelle. The proposed method strictly decreases the energy function every step, so it will definitely converge. However, it may stagnate at a nondifferentiable point; the result obtained may not be a minimizer. The graph-cuts method does give the exact minimizer. From the figure, we see the results from the proposed method and graph-cuts are similar, though those with graph-cuts have slightly better *SNR*. The two methods cost similar computational time in C++, but the proposed method is much easier to code. If we use MATLAB for the implementation, the proposed method is as efficient as it is in C++. In contrast, implementation of the graph-cuts method is much more challenging.[1]

FIG. 3.3. *ROF denoising using a median formula (3.4). The test image is a* 512 *by* 512 Barbara. First column: *inputs with different noise levels.* Second column: *The denoised results using the proposed method. (Parameters:* $\lambda = 0.061$, *22 iterations in the top row;* $\lambda = 0.026$, *29 iterations in the bottom row.)* Third column: *The denoised results using the graph-cut algorithm under the same λs.*

---

[1]The experiments are done in C++ on an AMD Athlon 64x2 Dual Core 5600+ and 2GB RAM.

REMARK 3.2. The algorithm uses coordinate descent to approach the minimizer of ROF. Combined with our median formula, it is very efficient. However, it is known that the coordinate descent method can get stuck at a non-stationary point if the objective function is nondifferentiable [9], which is the case for ROF. So the obtained denoising result is not necessarily the optimal ROF solution. To remedy this problem, we can add small perturbations or alternate with another descent method to overcome these sticking points. Another solution is the multilevel approach of Tony Chan and Ke Chen [7].

3.2. Solving ROF with more neighbors.   The anisotropic discretized TV term with only 4 nearest neighbors has metrification artifacts (see figure 3.3). To avoid those artifacts, a simple method is to use more neighbors, like adding the 4 diagonal neighbors, or using 16 weighted neighbors, see figure 3.4. We will use different sizes of neighbors to discretize $|\nabla u|$ in the following. We propose discretizing $|\nabla u|$ by using
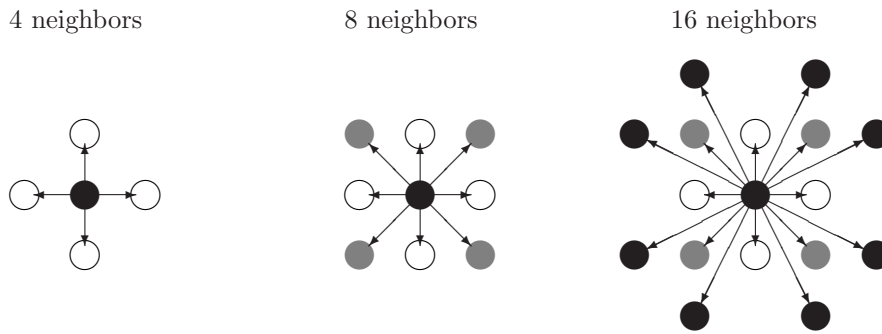
4 neighbors   8 neighbors   16 neighbors

FIG. 3.4. *This diagram shows how we choose different sizes of neighborhoods for discretizing $|\nabla u|$.*

4 neighbors [12],

$$|\nabla u| = \sum_{i=1}^{4} \left| u - u_i^{white} \right|;$$ (3.10)

using 8 weighted neighbors [12],

$$|\nabla u| = \sum_{i=1}^{4} \left| u - u_i^{white} \right| + \frac{1}{\sqrt{2}} \sum_{i=1}^{4} \left| u - u_i^{gray} \right|;$$ (3.11)

and using 16 weighted neighbors [12, 4],

$$|\nabla u| = 0.26 \sum_{i=1}^{4} \left| u - u_i^{white} \right| + 0.19 \sum_{i=1}^{4} \left| u - u_i^{gray} \right| + 0.06 \sum_{i=1}^{8} \left| u - u_i^{black} \right|.$$ (3.12)

With weights on the ℓ^1 terms in the energy function, the solution is still a median.

Suppose that the w_i are nonnegative, $u_1 \leq u_2 \leq \cdots \leq u_n$, and $\lambda > 0$, then

$$\operatorname*{argmin}_x \sum_i w_i |x - u_i| + \lambda (x - f)^2$$

$$= \operatorname{median}\Big\{ u_1, u_2, \ldots, u_n, f + \tfrac{1}{2\lambda}(w_n + w_{n-1} + \cdots + w_1),$$

$$f + \tfrac{1}{2\lambda}(w_n + w_{n-1} + \cdots + w_2 - w_1), f + \tfrac{1}{2\lambda}(w_n + w_{n-1} + \cdots + w_3 - w_2 - w_1),$$

$$\ldots, f + \tfrac{1}{2\lambda}(w_n - w_{n-1} - \cdots - w_1), f + \tfrac{1}{2\lambda}(-w_n - w_{n-1} - \cdots - w_1) \Big\}. \qquad (3.13)$$

The formula still holds in the limit $\lambda \searrow 0$. In this case, $\frac{1}{\lambda}$ goes to infinity.

$$\operatorname*{argmin}_x \sum w_i |x - u_i|$$

$$= \operatorname{median}\Big\{ u_1, u_2, \ldots, u_n, f + (w_n + w_{n-1} + \cdots + w_1)\infty,$$

$$f + (w_n + w_{n-1} + \cdots + w_2 - w_1)\infty, f + (w_n + w_{n-1} + \cdots + w_3 - w_2 - w_1)\infty,$$

$$\ldots, f + (w_n - w_{n-1} - \cdots - w_1)\infty, f + (-w_n - w_{n-1} - \cdots - w_1)\infty \Big\}.$$

From the equations, we can get the median formula for minimizing an energy only containing weighted ℓ^1 terms. Now, we can use the median formula (3.13) to implement the three different kinds of discretized TV method. We will compare the numerical results in figure 3.5.
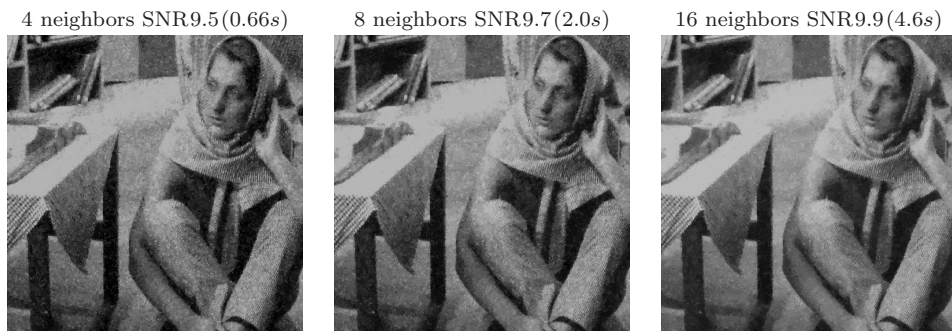
4 neighbors SNR$9.5(0.66s)$ 8 neighbors SNR$9.7(2.0s)$ 16 neighbors SNR$9.9(4.6s)$

FIG. 3.5. *Denoising with the median formula (3.13) using 4, 8, and 16 neighbors. The input is the noisy* Barbara *with SNR 3.1 used in the previous experiment. Each experiment uses 10 iterations (Parameters: $\lambda = 0.026$).*

3.3. Nonlocal median. The nonlocal mean method was introduced by Buades, Coll and Morel [5]; the method works extremely well for image denoising. The idea is to find similar patches inside the image, then average them to get rid of the noise.

We start by defining weights of any two pixels in the image f. Every two pixels x and y have a weight $w(x,y)$ used to evaluate the similarity of their patches. Define

$$w(x,y) = \exp\Big(-\frac{1}{h^2} \int_\Omega G_a(t) |f(x+t) - f(y+t)|^2 \, dt \Big), \qquad (3.14)$$

where G_a is a Gaussian with standard deviation a.

For computational efficiency, the support of $w(x,y)$ is often restricted to a "search window" $|x-y|_\infty \le R$ and set to zero outside.

The nonlocal denoising filter [5] is

$$NL(u)(x) = \frac{1}{c(x)} \int_\Omega w(x,y) f(y)\, dy, \qquad c(x) = \int_\Omega w(x,y)\, dy. \qquad (3.15)$$

Here, we come up with the Nonlocal-ROF model as in [11],

$$\arg\min_u E(u,f) = \int w(x,y) |u(x) - u(y)|\, dx\, dy + \lambda \int (u-f)^2\, dx. \qquad (3.16)$$

In this objective function, we use the weighted ℓ^1 norm such that the result will be sharper than using the weighted ℓ^2 norm, which is the nonlocal filter case. The discrete version is

$$\arg\min_u E(u,f) = \sum_{i,j} w_{ij} |u_i - u_j| + \lambda \sum_i (u_i - f_i)^2. \qquad (3.17)$$

We solve this by our weighted median formula and coordinate descent method. Consider the optimization for u_k with all other pixels $u_i (i \ne k)$ fixed, then (3.17) reduces to the weighted subproblem (3.13), for which we have a median formula for its solution. We call this the nonlocal median. We apply it pixel-by-pixel and repeat for few times to get a denoised result. The numerical examples are shown in Fig. (3.6).
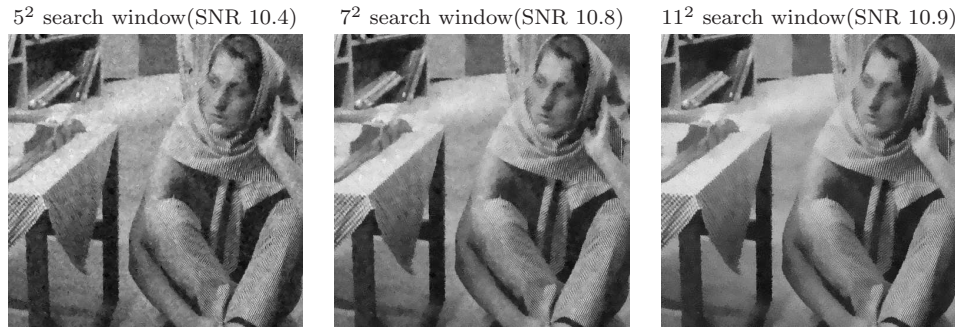
5^2 search window(SNR 10.4) 7^2 search window(SNR 10.8) 11^2 search window(SNR 10.9)

FIG. 3.6. *Denoising by approximately solving (3.17) with a median formula (3.13) using different search windows. The input is the noisy* Barbara *with SNR* 3.1 *used in the previous experiment. Each experiment uses* 2 *iterations. The runtimes are* 14s *with the* 5×5 *search window and* 20s *with the* 7×7 *search window and* 34s *with* 11×11 *search window. (Parameters:* $\lambda = 5 \times 10^{-3}$, $a = 1.25$, $h = 10.2$.)

4. Numerical results

Figure 3.3 shows two denoising experiments with the 4-neighbor ROF median formula[2]. The first column is the clean image, the second column is two noisy inputs with different noise levels, and the last column shows the denoised results. Figure 3.5 shows the improvement by using more neighbors in the TV discretization. The

---

[2]The experiments are done on an AMD Athlon 64x2 Dual Core 5600+ and 2GB RAM. Figure 3.3 and figure 3.5 are run in C++, the others are in MATLAB R2007b.

4-neighbor method is the most efficient, but the 8-neighbor and 16-neighbor methods produce better results. Figure 3.6 shows denoising results using the nonlocal median (3.13) with different sizes of search windows. Again we observe a trade off between computational efficiency and quality with the different search window sizes.

5. Conclusion

We have developed a simple and very general formula for finding solutions to functions that arise in ℓ^1 minimization problems in image research and elsewhere. The formula involves simple expressions using the median and leads to interesting and fast denoising algorithms.

REFERENCES

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

[2] D. Donoho, *Denoising by soft thresholding*, IEEE. Trans. Inform. Theory, 41(3), 613–627, 1995.

[3] S. Chen, D. Donoho and M.A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput., 20, 33–61, 1998.

[4] H.T. Nguyen, M. Worring and R. van den Boomgaard, *Watersnakes: energy-driven watershed segmentation*, IEEE Trans. PAMI, 25(3), 330–342, 2003.

[5] A. Buades, B. Coll and J.M. Morel, *On image denoising methods*, SIAM Multiscale Modeling and Simulation, 4(2), 490–530, 2005.

[6] J. Darbon and M. Sigelle, *A fast and exact algorithm for total variation minimization*, 2nd Iberian Conference on Pattern Recognition and Image Analysis, LNCS, 3522, 351–359, 2005.

[7] T. Chan and K. Chen, *An optimization-based multilevel algorithm for total variation denoising*, Multiscale Model. Simul., 5(2), 615–645, 2006.

[8] J. Xu and S. Osher, *Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising*, IEEE Transactions on Image Processing, 16, 534–544, 2006.

[9] J. Friedman, T. Hastie, H. Hofling and R. Tibshirani, *Pathwise coordinate optimization*, The Annals of Applied Statistics, 1, 302–332, 2007.

[10] T. Goldstein and S. Osher, *The split Bregman algorithm for ℓ^1 regularized problems*, UCLA CAM Report, April, 08–29, 2008.

[11] G. Gilboa and S. Osher, *Nonlocal linear image regularization and supervised segmentation*, SIAM Multiscale Modeling and Simulation, 6(2), 595–630, 2007.

[12] D. Goldfarb and W. Yin, *Parametric maximum flow algorithms for fast total variation minimization*, submitted, 2008.

[13] W. Yin, S. Osher, D. Goldfarb and J. Darbon, *Bregman iterative algorithm for ℓ^1 minimization with applications to compressed sensing*, SIAM J. Imaging Sciences, 143–168, 2008.

[14] J. Barral Souto, *El modo y otras media, casos particulares de una misma expresión matemática*, Tech. Rep. 3, Instituto de Biometria, Universidad Nacional de Buenos Aires, Argentina, 1938.

[15] G. Winkler, V. Aurich, K. Hahn and A. Martin, *Noise reduction in images: some recent edge-preserving methods*, Pattern Recognition and Image Analysis, 9, 749–766, 1999.

[16] P. Mrazek, J. Weickert and A. Bruhn, *On robust estimations and smoothing with spatial and tonal kernels*, R. Klette, R. Kozera, L. Noakes, J. Weickert (Eds.), Geometric Properties from Incomplete Data, Springer, Dordrecht, 335–352, 2006.