Digital Signal Processing 48 (2016) 178–187

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Fast and exact unidimensional L2–L1 optimization as an accelerator for iterative reconstruction algorithms

Marcelo V.W. Zibetti [a,*,1], Daniel R. Pipa [a,1], Alvaro R. De Pierro [b,1]

[a] Federal University of Technology – Paraná (UTFPR), Curitiba, 80230-901, Brazil
[b] CNPq, Brazil

ARTICLE INFO

Article history:
Available online 21 September 2015

Keywords:
L2–L1 problems
Line search
Compressed sensing
Sparse reconstruction
Iterative algorithms

ABSTRACT

This paper studies the use of fast and exact unidimensional L2–L1 minimization as a line search for accelerating iterative reconstruction algorithms. In L2–L1 minimization reconstruction problems, the squared Euclidean, or L2 norm, measures signal-data discrepancy and the L1 norm stands for a sparsity preserving regularization term. Functionals as these arise in important applications such as compressed sensing and deconvolution. Optimal unidimensional L2–L1 minimization has only recently been studied by Li and Osher for denoising problems and by Wen et al. for line search. A fast L2–L1 optimization procedure can be adapted for line search and used in iterative algorithms, improving convergence speed with little increase in computational cost. This paper proposes a new method for exact L2–L1 line search and compares it with the Li and Osher's, Wen et al.'s, as well as with a standard line search algorithm, the method of false position. The use of the proposed line search improves convergence speed of different iterative algorithms for L2–L1 reconstruction such as iterative shrinkage, iteratively reweighted least squares, and nonlinear conjugate gradient. This assertion is validated experimentally in applications to signal reconstruction in compressed sensing and sparse signal deblurring.

1. Introduction

Line search is an important accelerating step for many multidimensional optimization methods [1,2]. Since a significant group of iterative reconstruction algorithms are based on optimization methods [3–5], a nonexpensive line search can be readily added to improve their performance. In general, the line search step can be described as:

$$\alpha_k = \arg\min_{\alpha} \Psi(\mathbf{f}_k + \alpha \mathbf{d}_k), \tag{1}$$

where $\Psi(\mathbf{f})$ is the multidimensional cost function to be minimized, $\mathbf{f}_k$ is the current estimate or solution at iteration k, $\mathbf{d}_k$ is a search direction, normally pointing to a lower value of the cost function, and α is the step size, a complementary information needed to move the current solution to a better new one, $\mathbf{f}_{k+1} = \mathbf{f}_k + \alpha_k \mathbf{d}_k$, with a lower value of $\Psi(\mathbf{f})$. Generally, $\alpha > 0$. Most optimization algorithms follow this recipe, with some common examples being the classical steepest descent (SD) and conjugate gradient (CG) [1].

Another group of algorithms follows a fixed point iteration scheme, which has the form $\mathbf{f}_{k+1} = F(\mathbf{f}_k)$, and can also be expressed as $\mathbf{f}_{k+1} = \mathbf{f}_k + (F(\mathbf{f}_k) - \mathbf{f}_k)$. This last expression suggests a search direction given by $\mathbf{d}_k = (F(\mathbf{f}_k) - \mathbf{f}_k)$. In this case, no search is usually performed and $\alpha_k = 1$. Examples of such methods are: iteratively reweighted least squares (IRLS) [6,7] as well as some iterative shrinkage algorithms (ISA) [3,8].

The step size α_k in (1) should essentially guarantee the convergence of the algorithm, sufficiently reducing the value of the cost function. For some algorithms, choosing the exact minimum point, instead of an approximation along the search direction, compensates for the additional computational cost per iteration of the line search. If the cost to find the minimum point is relatively low, then it may accelerate the whole algorithm, reducing the total number of iterations for convergence, as noted in [8].

Some successful examples in the field of signal reconstruction techniques are the minimum of quadratic cost functions (least squares problems) [1] and of absolute value functions (least absolute problems) [9,10], where the optimal step is easily computed. For quadratic problems, the optimal step is a standard part of the steepest descent and (linear) conjugate gradient algorithms [1,5]. For pure L1 problems, the optimal step is calculated using a weighted median [10], which may cost slightly more than the least squares step (that can be computed as a weighted mean), though still overall advantageous.

* Corresponding author.
E-mail addresses: marcelozibetti@utfpr.edu.br (M.V.W. Zibetti), danielpipa@utfpr.edu.br (D.R. Pipa), alvaro@ime.unicamp.br (A.R. De Pierro).

1 The work of these authors was supported by CNPq Grants 306030/2014-4 and 475553/2013-6.

http://dx.doi.org/10.1016/j.dsp.2015.09.009
1051-2004/

A problem of great interest for the signal reconstruction community is the combination of L2 and L1 norms in the cost function, i.e.:

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \Psi(\mathbf{f}), \tag{2}$$

with

$$\Psi(\mathbf{f}) = \frac{1}{2}\|\mathbf{g} - \mathbf{Hf}\|_2^2 + \lambda\|\mathbf{Rf}\|_1, \tag{3}$$

where $\|\mathbf{y}\|_2^2 = \sum_{j=1}^{M} |y_j|^2$ is the squared L2 norm and $\|\mathbf{x}\|_1 = \sum_{i=1}^{N} |x_i|$ is the L1 norm. $\mathbf{H}$ is the system matrix, of size $M \times P$, $\mathbf{g}$ is the captured data, ordered in an $M \times 1$ vector, $\mathbf{f}$ is the signal to be estimated, in a $P \times 1$ vector, and $\mathbf{R}$ is the regularization matrix, of size $N \times P$. λ is a positive regularization parameter.

This type of functional appears in several signal reconstruction problems, such as compressed sensing [11,12], sparse signal deblurring [8], super-resolution [13] and medical image reconstructions [14,15]. To solve (3), iterative methods based on descent directions or fixed point iteration schemes are commonly used and, therefore, a study of fast alternatives for exact line search is of relevance.

In this paper, we first explore the subject describing existing fast methods that can be used for L2–L1 line search. In Section 2 we describe with details the L2–L1 optimization problem and the one dimensional function that should be minimized for the line search. Then we describe the Method of False Position (MFP), that is very fast to approximate the minimum, but could need too many iterations to calculate its "exact" value, followed by the methods proposed by Li and Osher [16] (LO) and by Wen, Yin, Goldfarb and Zhang [17] (WYGZ). In Section 3 we present our own fast line search method and at the end of the section all the methods are compared through their computational cost (number of Floating Point Operations) as well as their computational CPU times, highlighting its advantages and disadvantages. Our main goal was, not only to improve the existing line search methods for the L2–L1 problem, but to show that exact line search could substantially improve the results of many existing algorithms when applied to important signal processing problems. So, in Section 4 the four procedures are evaluated when used as accelerators for three typical iterative algorithms, namely: Iterative Shrinkage Algorithm (ISA), Iteratively Reweighted Least Squares (IRLS), and nonlinear Conjugate Gradients (NLCG). The chosen problems were Compressed Sensing and Sparse Deblurring. Section 5 is dedicated to some conclusions and further research directions.

2. L2–L1 optimal line search problem

For the multidimensional optimization problem given in (3), the line search problem can be written as:

$$\alpha^* = \arg\min_{\alpha} \Psi(\alpha), \tag{4}$$

with

$$\Psi(\alpha) = \frac{1}{2}\|\mathbf{g} - \mathbf{H}(\mathbf{f} + \alpha\mathbf{d})\|_2^2 + \lambda\|\mathbf{R}(\mathbf{f} + \alpha\mathbf{d})\|_1, \tag{5}$$

where the k index is omitted for visual convenience.

Expanding (5):

$$\begin{aligned}\Psi(\alpha) &= \sum_{j=1}^{M} \frac{1}{2}|g_j - \mathbf{h}_j^T(\mathbf{f} + \alpha\mathbf{d})|^2 + \lambda\sum_{i=1}^{N}|\mathbf{r}_i^T(\mathbf{f} + \alpha\mathbf{d})| \\ &= \sum_{j=1}^{M} \frac{1}{2}|g_j - \mathbf{h}_j^T\mathbf{f} - \alpha\mathbf{h}_j^T\mathbf{d}|^2 + \lambda\sum_{i=1}^{N}|\mathbf{r}_i^T\mathbf{f} + \alpha\mathbf{r}_i^T\mathbf{d}| \\ &= \sum_{j=1}^{M} \frac{1}{2}|\mathbf{h}_j^T\mathbf{d}|^2 \left|\alpha - \frac{g_j - \mathbf{h}_j^T\mathbf{f}}{\mathbf{h}_j^T\mathbf{d}}\right|^2 \\ &\quad + \sum_{i=1}^{N}\lambda|\mathbf{r}_i^T\mathbf{d}| \left|\alpha - \left(-\frac{\mathbf{r}_i^T\mathbf{f}}{\mathbf{r}_i^T\mathbf{d}}\right)\right|, \end{aligned} \tag{6}$$

where $\mathbf{h}_j$ is the j-th row of $\mathbf{H}$, and $\mathbf{r}_i$ is the i-th row of $\mathbf{R}$, it can be written as:

$$\Psi(\alpha) = \sum_{j=1}^{M} \frac{q_j}{2}|\alpha - x_j|^2 + \sum_{i=1}^{N}\omega_i|\alpha - \alpha_i| \tag{7}$$

with:

$$\begin{aligned} x_j &= \frac{g_j - \mathbf{h}_j^T\mathbf{f}}{\mathbf{h}_j^T\mathbf{d}}, \quad q_j = |\mathbf{h}_j^T\mathbf{d}|^2 = (\mathbf{h}_j^T\mathbf{d})^2, \\ \alpha_i &= -\frac{\mathbf{r}_i^T\mathbf{f}}{\mathbf{r}_i^T\mathbf{d}}, \qquad \omega_i = \lambda|\mathbf{r}_i^T\mathbf{d}| = \lambda\sqrt{(\mathbf{r}_i^T\mathbf{d})^2}. \end{aligned} \tag{8}$$

One can easily verify that this problem is a sum of multiple unidimensional quadratic functions and magnitude, or absolute value functions. Each individual function has its own minimum point, x_j and α_i, and growth rate, g_j and ω_i.

Although (7) is a convex nondifferentiable function,[2] we can write its subdifferential [18] as:

$$\partial\Psi(\alpha) = \sum_{j=1}^{M} q_j(\alpha - x_j) + \sum_{i=1}^{N}\omega_i \operatorname{sign}(\alpha - \alpha_i), \tag{9}$$

where the sign function is defined as:

$$\operatorname{sign}(\alpha) = \begin{cases} 1, & \text{if} \quad \alpha > 0 \\ -1, & \text{if} \quad \alpha < 0 \\ [-1, 1] & \text{if} \quad \alpha = 0 \end{cases}. \tag{10}$$

Remark 1. It is worth noting that the equality above, as well as those that follow, are an abuse of notation, because the right hand side represents not a single number but a set. However, all the necessary calculations are easy to be verified in the general scene point-to-set setting. In the same way $\Psi'(\alpha)$ will denote one element of the set $\partial\Psi(\alpha)$. The fact that we are dealing with one dimensional functions strongly simplifies the analysis.

Observing that (9) is a sum of linear terms plus sign functions, we can still develop it as:

$$\partial\Psi(\alpha) = \left(\sum_{j=1}^{M} q_j\right)\alpha - \left(\sum_{j=1}^{M} q_j x_j\right) + \sum_{i=1}^{N}\omega_i \operatorname{sign}(\alpha - \alpha_i)$$

$$\partial\Psi(\alpha) = Q\alpha - Qx_q + \sum_{i=1}^{N}\omega_i \operatorname{sign}(\alpha - \alpha_i), \tag{11}$$

where $Q = \sum_{j=1}^{M} q_j$, and $x_q = \frac{1}{Q}\sum_{j=1}^{M} q_j x_j$. Note that x_q is the weighted mean of the x_j points with the respective q_j weights. The two first terms in the right-hand side of (11) are a sum of lines that can be converted to one line.

A plot of the $\partial\Psi(\alpha)$, shown in bold continuous line in Fig. 1, illustrates that the subdifferential in (11) can be represented as

[2] Since the magnitude function, $|\alpha|$, is not differentiable at $\alpha = 0$, we need to resort to the subderivative and subdifferential concepts, which establish that the subdifferential of $|\alpha|$ at $\alpha = 0$ is the closed interval $[-1, 1]$. Each element of the set is called subgradient (subderivative in this case because we are dealing with single variables).