High-dimensional convex optimization via optimal affine subgradient algorithms

Masoud Ahookhosh
Faculty of Mathematics
University of Vienna
masoud.ahookhosh@univie.ac.at

Arnold Neumaier
Faculty of Mathematics
University of Vienna
Arnold.Neumaier@univie.ac.at

Abstract: This study is concerned with some algorithms for solving high-dimensional convex optimization problems appearing in applied sciences like signal and image processing, machine learning and statistics. We improve an optimal first-order approach for a class of objective functions including costly affine terms by employing a special multidimensional subspace search. We report some numerical results for some imaging problems including nonsmooth regularization terms.

Keywords: Convex optimization, Optimal subgradient algorithm, Multidimensional subspace search

1 Introduction and Basic Idea

Lots of applications in signal and image processing, machine learning, statistics, geophysics, etc include affine terms in their structure that are the most costly part of function evaluations. Therefore, we consider the following unconstrained convex optimization problem

$$\text{minimize } f(x) := \sum_{i=1}^{n_1} f_i(x, A_i(x)), \tag{1}$$

where $f_i(x, v_i)$ $(i = 1, \cdots, n_1)$ are convex functions defined for $x \in \mathbb{R}^n$ and $v_i = A_i(x) \in \mathbb{R}^{m_i}$ and the A_i are linear operators from $\mathbb{R}^n$ to $\mathbb{R}^{m_i}$. The aim is to derive an approximating solution x just using function values $f(x)$ and subgradients $g(x)$. Consider the following generic first-order descent algorithm:

Generic first-order descent algorithm

Input: $x_0 \in \mathbb{R}^n$; $\epsilon > 0$;
Output: x_b;
Begin
1 choose x_b;
2 calculate $f_{x_b} \leftarrow f(x_b)$; $g_{x_b} \leftarrow g(x_b)$;
3 **Repeat**
4 generate x'_b by an extended algorithm;
5 generate $\bar{x}_b$ by a heuristic algorithm;
6 $x_b \leftarrow \operatorname{argmin}_{\{x_b, x'_b, \bar{x}_b\}} f(x)$;
7 **Until** a stopping criterion holds;
End

On the basis of high-dimensional data used in applications, the most costly part of function and subgradient evaluations relates to applying the linear operator and its transpose. Then we should decrease the number of linear operator calculations as much as possible. To find an appropriate $\bar{x}_b$ in line 5 of this algorithm, we use in each step of this algorithm a multidimensional subspace search as inner algorithm. The main idea is a generalization of line search techniques by saving $M \ll n$ columns u and the previously computed vectors $v_i = A_i(u)$ in the matrices U and V_i $(i = 1, \cdots, n_1)$, find $t^* \approx \operatorname{argmin}_{t \in \mathbb{R}^M} \tilde{f}(t)$ where

$$\tilde{f}(t) := f(x_b + Ut) = \sum_{i=1}^{n_1} f_i(x + Ut, v_i + V_i t) \tag{2}$$

can be calculated without further calls to the A_i.

Multidimensional subspace search procedure

Input: $x_b \in \mathbb{R}^n$; U; $v_i \in \mathbb{R}^{m_i}$; $V_i (i = 1, \cdots, n_1)$;
Begin
approximately solve the M-dimensional problem

$$t^* \approx \operatorname*{argmin}_{t \in \mathbb{R}^M} \tilde{f}(t); \tag{3}$$

$\bar{x}_b = x_b + Ut$;
End

2 Specific version of the algorithm

The general idea is made specific by choosing a particular descent algorithm. We use the OSGA algorithm, [3], which monotonically reduces a bound on the error $f(x_b)$ of the function value of the best known point x_b. The modified version of OSGA including affine scaling and the multidimensional subspace search, as defined above, is called ASGA which is suitable to deal with high-dimensional convex optimization appearing in applications. In ASGA, the subproblem (3) is also solved by OSGA.

The OSGA algorithm generates and updates linear relaxations

$$f(z) \geq \gamma + h^T z \quad \forall z \in \mathbb{R}^n,$$

where $\gamma \in \mathbb{R}$, $h \in \mathbb{R}^n$. In line 3 of the algorithm we set $x_b' = x_b + \alpha(u - x_b)$, where $u = U(\gamma, h) \in C$ solves a minimization problem of the form

$$E(\gamma, h) := -\inf_{x \in \mathbb{R}^n} \frac{\gamma + h^T z}{Q_0 + \frac{1}{2}\|z - z_0\|^2}. \qquad (4)$$

For details see [3]. It is proved in [3] that OSGA achieves the optimal complexity bounds $O(1/\sqrt{\epsilon})$ for the optimization of smooth convex functions and $O(1/\epsilon^2)$ for nonsmooth convex functions [2], no matter which heuristic choice $\bar{x}_b$ is made, so ASGA is an optimal complexity algorithm.

3 Numerical Results

This section reports some numerical results to show the efficiency of the proposed algorithms for solving practical problems arising in applications. On the basis of the fact that OSGA and ASGA just need function and subgradient evaluations, they can be employed in solving wide range of problems including regularization terms. Examples include *lasso, basis pursuit denoising, l_1 and l_2 decoding, isotropic and anisotropic total variation, group regularizations, elastic net, nuclear norm, linear support vector machine, kernel-based models.*

Here, we consider the $l_2^2 - ITV$ regularization problem defined by

$$f(x) = \frac{1}{2}\|A(x) - b\|_2^2 + \lambda\|x\|_{ITV},$$

where $\|.\|_{ITV}$ denotes the isotropic TV norm and A is a linear operator chosen as in the TwIST package for reconstruction of the 512×512 blurry-noisy Lena image.

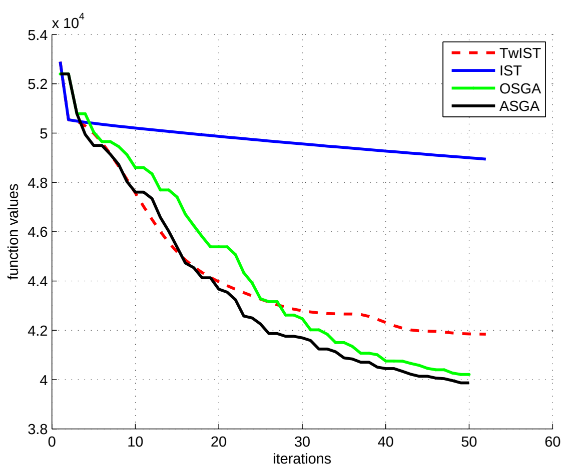

Fig. 1: A comparison of function values for IST, TwIST, OSGA and ASGA

If we count the images of Fig. 2 row by row, the first image is the original image, and the second image shows a blurry-noisy image constructed by adding noise and uniform 9×9 blur with BSNR $= 40$ dB. The rest of the images are restored by the minimization problem (3) where x_0 is given by a Wiener filter for all considered algorithms. The third image is produced by the IST algorithm, while the fourth image is recovered by TwIST [4]. Also the fifth image restored by OSGA and the sixth images was reconstructed by ASGA. It is clear that the final function value of the proposed algorithms is less than those of IST and TwIST and the restored image visually looks good. This shows that the proposed algorithms can effectively reconstruct blurry and noisy images at a reasonable cost.

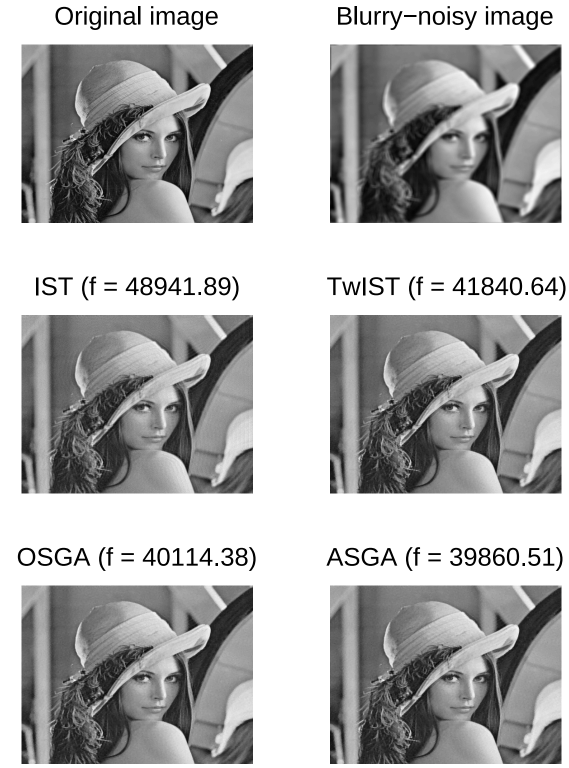

Fig. 2: Isotropic TV-based image reconstruction from 40% missing sample with the considered algorithms: IST, TwIST, OSGA and ASGA

References

[1] M. Ahookhosh, A. Neumaier, An affine subgradient algorithm for large-scale convex optimization. *Manuscript*, University of Vienna, 2013.

[2] Y. Nesterov, Introductory lectures on convex optimization: A basic course, Kluwer, Dordrecht 2004.

[3] A. Neumaier, Convex optimization with optimal complexity. *Manuscript*, University of Vienna, 2013.

[4] TwIST website: http://www.lx.it.pt/ bioucas/TwIST/TwIST.htm