Acceleration on sparse promoting seismic applications

Lina Miao*, University of British Columbia, Vancouver, BC, Canada
linaouc@gmail.com
and
Felix Herrmann, University of British Columbia, Vancouver, BC, Canada

Summary

Sparse promoting oriented problems are never new in seismic applications. Back in 1970s, geophysicists had well exploited the robustness of sparse solutions. Moreover, with the emerging usage of compressed sensing in recent years, sparse recovery have been favored in dealing with 'curse of dimensionality' in various seismic field acquisition, data processing, and imaging applications. Although sparsity has provided a promising approach, solving for it presents a big challenge. How to work efficiently with the extremely large-scale seismic problem, and how to improve the convergence rate reducing computation time are most frequently asked questions in this content. In this abstract, the author proposed a new algorithm -- PQNℓ1, trying to address those questions. One example on seismic data processing is included.

Introduction

"When a traveler reaches a fork in the road, the ℓ1 norm tells him to take either one way or the other, but the ℓ2 norm instructs him to head off into the bushes. " -- John F. Claerbout and Francis Muir, 1973. The road ℓ1 norm succeed telling is one sparse solution, and seeking for a minimum ℓ1 norm solution is a typical way how sparse promoting seismic applications are formulated. The minimum ℓ1 norm formulation has offered quite fruitful results over the years.

Currently, there are tons of algorithms claiming being able to solve ℓ1 norm problem, some of which are listed at http://dsp.rice.edu/cs. One sparse solver -- SPGℓ1 developed by Michael P. Friedlander in 2008(van den Berg and Friedlander) has been proved efficient in various seismic applications (Herrmann, Friedlander and Yilmaz 2012), provided the running feature for SPGℓ is its suitability in solving large-scale problems.

Given the existent efficiency of SPGℓ1, the author wants to push it even further. As in most of our seismic applications, iteratively evaluating the objective function (e.g. data residual) is very computational involving, weeks can be taken for evaluating one objective function in one iteration. The proposed algorithm tries to accelerate the recovery by introducing a second order direction into SPGℓ1, and resulting an improvement in the algorithm convergence behavior.

Theory and/or Method

Logic of SPGℓ1

There are different flavors for sparse promoting seismic application, one of which is called BPDN (minimize $||x||_1$ subject to $||Ax - b||_2 \leq \delta$), seeking a minimum one-norm solution of an undetermined least-squares problem. Trying to probe the optimum trade-off curve (Pareto curve) between the least squares fit and the one-norm of the solution, SPGℓ1 is able to decompose the BPDN problem into several LASSO problem (minimize $||Ax - b||_2$ subject to $||x||_1 \leq \tau$), for each LASSO problem a spectral gradient projection method approximately minimizes a least squares problem with an explicit one norm constrain. In figure 1, the black line shows an example of the Pareto curve, and the black dashed line show the true solution, where each point stands for one true iteration in spectral projected gradient (PSG) subroutine.

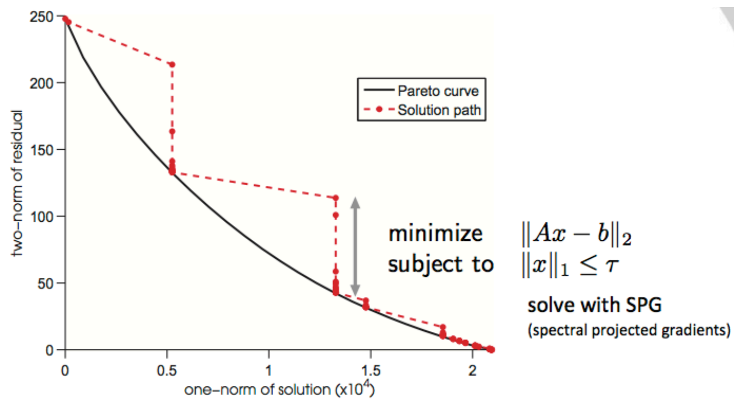


Figure 1: Trade off between data residual and solution sparsity

Our contribution

The contribution of this work is considering solving the individual Lasso problems with a projected quasi newton (PQN) algorithm. With some extra pay on memory constructing an approximate Hessian matrix, resulting in a superlinear convergence rate compared to the linear convergence rate for SPG method. In order to see the difference, pseudo code for SPG and PQN are listed in figure 2. Apart from the difference marked as red we can also see that since there is no closed form solution for finding the PQN direction, a SPG method is used to solve for PQN iteration. It seems to be not worthy at the first glance of doing so since the iterative solving for quasi newton direction could be expensive. However if we look carefully we can find out some extra work on computing a good searching direction makes sense in expensive applications. It is simply because there is no need to evaluate objective function in this subroutine. A better searching direction will bring decrease in the total iteration we need to run for a certain objective data residual, which corresponds to a decrease in computing objective function, a decrease in total computation time. We can also predict this method would benefit more if we have an ill conditioned problem because with a second order correction, gradient direction would suffer less from the problem's ill-posedness.

Objective : $\min_{x \in C} f(x)$

SPG

for $k = 1$ to $iter_{SPG}$ **do**

$$x^{k+1} = \underset{x \in C}{\operatorname{argmin}} f(x^k) + (x - x^k)^T \nabla f(x^k) + \frac{1}{2\alpha} ||(x - x^k)||_2^2$$

Closed form solution :

$$x^{k+1} = P_{x \in C}(x^k - \alpha \nabla f(x^k));$$

end for

PQN

for $K = 1$ to $iter_{PQN}$ **do**

$$x^{K+1} = \underset{x \in C}{\operatorname{argmin}} f(x^K) + (x - x^K)^T \nabla f(x^K) + \frac{1}{2}(x - x^K)^T B^K (x - x^K)$$

Does not have a closed form solution, solved via SPG approach:

for $k = 1$ to $iter_{SPG}$ **do**

...

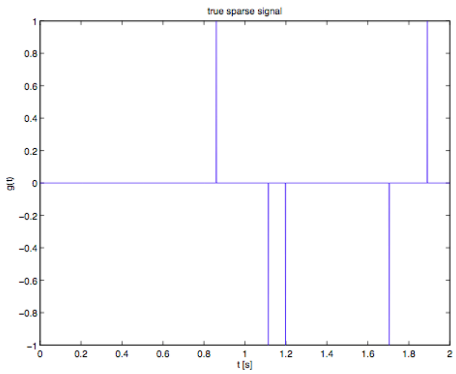
end for

end for

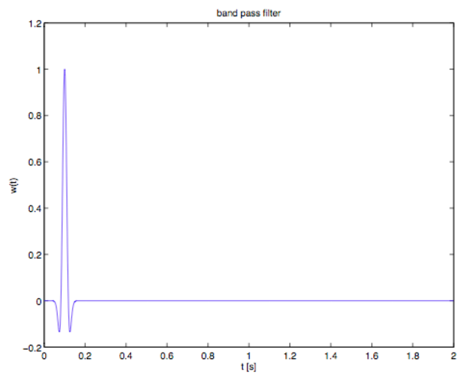
Figure 2: SPG and PQN algorithms

Examples

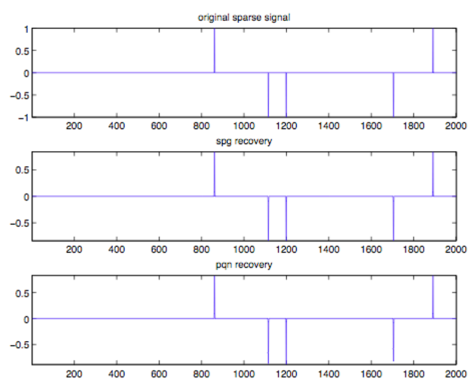
In this section, we're going to show one very simple example, to show acceleration is achieved with proposed PQNℓ1 algorithm. (More examples on data acquisition and inversion are also available) In this example, we have a series of spike reflectors, and convolving it with a low pass filter. In order to recover the reflectors, we run SPGℓ1 and PQNℓ1 to perform a spike de-convolution to the same data residual, the different behaviors are listed in figure 3. a is the original spike reflector, b is the source wavelet, c shows the recovery from SPGℓ1 and PQNℓ1 in the middle and bottom, d shows a variation within PQNℓ1 algorithm, i.e. with 5 10 20 30 gradient updates to form the approximate Hessian matrix. As we can see, acceleration is achieved with the proposed algorithm with the same data residual compared to SPGℓ1.



a



b



c

	ITERATION	MAT-VEC
SPGL1	1950	2907+1951
PQNL1(5)	820	822+822
PQNL1(10)	380	382+382
PQNL1(20)	370	372+372
PQNL1(30)	341	343+343

d

Figure 3: SPG and PQN comparison in one spike de-convolution example

Conclusions

The proposed algorithm PQNℓ1 is being able to accelerate on sparse promoting seismic applications, the advantage is most obvious when the objective function is prohibitive to evaluate, and when the problem is ill conditioned.

References

Albert Tarantola, **Inverse Problem Theory and Methods for Model Parameter Estimation,** 2005

Felix J. Herrmann, Michael P. Friedlander, and Ozgur Yilmaz, "**Fighting the Curse of Dimensionality: Compressive Sensing in Exploration Seismology**", *Signal Processing Magazine, IEEE*, vol. 29, p. 88-100, 2012

E. van den Berg and M. P. Friedlander, **"Probing the Pareto frontier for basis pursuit solutions"** SIAM J. on Scientific Computing, 31(2):890-912, November 2008

Lina Miao, "**Accelerating on sparse promoting recovery and its benefits in seismic application**", *SINBAD consortium talks*. 2012