A Hybrid Hard Thresholding Algorithm for Compressed Sensing

Fengmin Xu, Shanhe Wang, Zongben Xu
Department of Mathematics and Institute for Information and System Science
Xi'an Jiaotong University
Xi'an, China
fengminxu@mail.xjtu.edu.cn
wangshanhe.137@stu.xjtu.edu.cn

***Abstract*—Iterative hard thresholding algorithm (IHT) is a novel and efficient method to solve signal and image reconstruction in compressed sensing, but it is sensitive to the initial point and converges to a local optimal solution. Therefore, to overcome its shortcoming, in this paper a hybrid hard thresholding algorithm (HHT) is derived by introducing the simulated annealing algorithm (SA) into the IHT. And a series of experiments are provided on signal and image reconstruction to assess performance of the algorithm. The experiments and applications show that the proposed algorithm uses less sampling to construct the signal and image and is more stable, as compared with IHT.**

Keywords- simulated annealing; iterative hard thresholding; compressed sensing; signal reconstruction

I. INTRODUCTION

In 2006, E. Candes and D. Donoho officially proposed the concept of Compressed Sensing [1][2], whose core thought was to amalgamate compression with sampling. Compressed sensing mainly contains three aspects: sparse representation of a signal, designing on the measurements matrix and research on reconstruction algorithms. In this paper what we concentrate on is the signal and image reconstruction algorithm.

In general there have been two kinds of reconstruction algorithm. One is greedy algorithm, such as Matching Pursuit (MP) [3], Orthogonal Matching Pursuit (OMP) [4], Regularized Orthogonal Matching Pursuit (ROMP) [5] and Compressive Sampling Matching Pursuit (CoSaMP) [6], etc. However, only under very strict conditions can these methods converge to the optimal solution [7][8]. The other is relaxed strategy, including basis pursuit de-noising method (BP) [9], the interior-point method [10], gradient projection method [11] and so on. Though those methods offer better performance in many cases, they have high computational complexity.

Recently, iterative hard thresholding algorithm (IHT) [12][13] was suggested to solve the problems of signal and image reconstruction. It is a simple, fast convergent procedure, and can also solve large scale problems for its low computational cost of each iteration. Unfortunately, it is a local convergent algorithm. If combined it with a global convergent algorithm, we can obtain a robust, global convergent and effective hybrid algorithm, which will be a exciting thing. But some problems arise: Which global convergent algorithm should be combined with? And how? Whether the hybrid algorithm is efficient as we expected? Those problems will be discussed in this paper and arranged specifically as follows:

In section II, we elaborate the hybrid algorithm, and some related experimental results would be showed in section III. Finally we summarize this paper and make some plan in section VI.

II. THE HYBRID HARD THRESHOLDING ALGORITHM

The problems of signal reconstruction can be expressed concisely as: A measurement matrix $A \in R^{M \times N} (M \ll N)$ is given together with an observation vector $b \in R^M$,

$$b = Ax + \varepsilon \tag{1}$$

Where x is a sparse signal and ε is a possible observation noise (white gaussian noise $\varepsilon \sim N(0, \sigma^2)$ in general). The reconstruction problem is how to recover the original sparse signal x via A and b.

If original signal x is known as M-sparse, that is, the number of nonzero elements of x is not more than M. Then, this problem can be written as [12]:

$$\begin{aligned} \min \quad & f(x) \equiv \|b - Ax\|_2^2 \\ \text{s.t.} \quad & \|x\|_0 \leq M \end{aligned} \tag{2}$$

According to [13], we call the model (2) M-sparse problem. The main step of IHT to solve this problem is:

$$x^{n+1} = H_M(x^n + A^T(b - Ax^n)), \tag{3}$$

where $H_M(x)$ is the non-linear operator that retains the largest (in magnitude) M elements of x and sets others to zero. To carry out (3) what we only need are initial iterative point x^0 and stopping criterion, e.g., iterative times n.

A. The Hybrid Algorithm

Simulated Annealing algorithm (SA) [14]-[17] is a global convergent one. It can jump out of the local optimal "trap" by adopting metropolis criterion. The reasons why we choose SA to combine with IHT in this paper are as follows:

Firstly, they are easy to combine. SA is simple and very easy to conduct too, while IHT has a fixed format to select parameters for M-sparse problem, and has higher conve-

978-1-61284-486-2/11/$26.00 ©2011 IEEE

rgent speed. For instance, considering low computational complexity of IHT, we embed IHT in the framework of SA, SA provides initial iterative points for IHT, while IHT helps SA to do fast local search. For every point generated by SA process, we make it converge to the nearest local optimum by using IHT. The evaluation value for the original point is the value of the object function at the local optimum. Then, whether the local optimum is accepted or not is decided by the metropolis criterion.

Secondly, SA is able to deal with evaluation function f, which does not need to be differentiable or continuous etc, even what we only need are some values of f in its domain. So the request of function f is quite little, which helps us to prove global convergence of hybrid algorithm easily.

Finally, SA is easy to parallelize calculation, and after combining with IHT, it still has the potential to solve large scale problems.

Next we obtain a hybrid algorithm named hybrid hard thresholding Algorithm (HHT), which is described as :

Step 1. *Initialization, choose the initial vector* x_0 *(*M*-sparse). Given error* ϵ_0*, initial temperature* T*, terminal temperature* T_f*, temperature attenuate function* α*, maximum iterative times* $K\max$*, iterative times* n *and error* ϵ *(May be equal to* ϵ_0*) in stopping criterion of IHT. Let* $k=0$*;*

Step 2. *Select* x_1 *from the neighborhood of* x_0 *randomly, calculate the local minimum* $y=H_{M,\ \epsilon}(x_1)$ *correspond-ding to* x_1*, then calculate the change of objective function* $\Delta f=f(y)-f(x_0)$*,* $k=k+1$*. Where* $y=H_{M,n,\epsilon}(x_1)$ *is the* M*-sparse vector iterated by IHT with initial point* x_1 *after* n *iterations or* $f(y)\le\epsilon$*;*

Step 3. *if* $f(y)<\epsilon_0$*, stop, output vector* y*; else if* $\Delta f<0$*, Let* $x_0=y$*, else produce* $\eta\sim U(0,1)$*, if* $e^{-\Delta f/T}>\eta$*, let* $x_0=y$*, goto Step 4;*

Step 4. *If equilibrium is reached (iterative times* $k\ge Kmax$*), let* $k=0$*, goto Step 5, else goto Step 2;*

Step 5. *Reduce* T*,* $T=T\alpha$*, if* $T<T_f$*, stop, output vector* y*, else goto Step 2.*

B. The Convergence Analysis of HHT

In fact, HHT hasn't changed SA's structure, we will illustrate that HHT maintains the statistical promise of SA below. Define $\hat{f}$:

$$\hat{f}(x)=f(H_{M,n,\epsilon}(x)).$$

Keeping the same parameters, the only difference between employing HHT for finding the optimal point of f and using SA on $\hat{f}$ is neighborhood. In HHT, $x_0\xrightarrow{disturbance}x_1\xrightarrow{IHT}y$,if x_0 and y can achieve each other, then HHT is global convergent by citing proof of global convergence of SA [17].

Definition 1 *Suppose* x_0 *(*$x_0\in R^N$*) is a* M*-sparse column vector. We claim* x *equals* x_0 *in the error* ϵ_0*, if their non-zero elements are at the same positions and* $\|x-x_0\|<\epsilon_0$*. The set consists of all vectors which equal* x_0 *in the error* ϵ_0 *is named* x_0 *equal set in error* ϵ_0*, express it as* E_{x_0,ϵ_0}*.*

Because signal reconstruction makes sense in certain precision, if the error ($\|\hat{x}-x\|$) between reconstructive signal $\hat{x}$ and the initial signal x is less than a given precision ϵ_0 ($\epsilon_0=1e-n, n\in N$), or has the same order of magnitude with ϵ_0, then we regard $\hat{x}$ as a accurate reconstructed signal. If we choose a vector $y\in E_{\hat{x},}$ arbitrarily, we have

$$\|x-y\|\le\|x-\hat{x}\|+\|\hat{x}-y\|<2\epsilon_0.$$

Then, y accurately reconstructs x too. So the equal set can be treated as one vector in an order of magnitude. Suppose $\bar{x}$ is in the neighborhood (augmented) of $\hat{x}$, if the neighborhood of $\bar{x}$ contains one element $y\in E_{\hat{x},\epsilon_0}$ too, then in our opinion $\bar{x}$ and $\hat{x}$ can achieve each other.

Definition 2 (Basin) *Given the error* ϵ_0*,* x *is* M*-sparse vector,* $B_{x,}$ *is a set about* x,ϵ_0*, if* $\forall y\in B_{x,\epsilon_0}$*, we have* $H_{M,n,}$ E_{x,ϵ_0}*, else not. If so,* B_{x,ϵ_0} *is called basin about* x *in the error* ϵ_0*.*

Definition 3 (Neighborhood) Suppose x_0 ($x_0\in R^N$) is a M-sparse column vector. The neighborhood of x_0 is

$$N_{x_0,\mu}=\{x\mid x=H_{M,n,\ \epsilon}(y_0), y_0=x_0+\mu randn(N,1)\}$$

Where μ *is a given constant.* ${}_{\epsilon}(y)\in$

From Definition 3, $x_0\xrightarrow{disturbance}y_0\xrightarrow{IHT}x$, what we should state is x can achieve x_0 in the same way, that is, $\exists\tilde{x}\in\{x:x+\mu randn(N,1)\}\cap B_{x_0,\epsilon_0}$. According to the distribution of normal distribution we know

$$P(\{x:x+\mu randn(N,1)\}\cap B_{x,\epsilon_0}\ne\emptyset)>0.$$

Hence x_0 and x can achieve each other, so from the above discussion HHT is a global convergent algorithm. The experiments completed in next section show HHT is a stable, effective and global convergent algorithm.

III. Numerical experiments

In this section, some experiments were taken to compare the performance of HHT in signal and image reconstruction, and to confirm global convergence of HHT, to test and verify its ability of large scale image reconstruction. On the premise of accurate reconstruction, the lower sampling number is, the better algorithm reconstructive performance is.

Numerical simulated environment is: CPU frequency 2.4G, 2G memory of personal computer, Matlab 7.8.

A. The Parameters' Selection

Dealing with model (2), the parameters in HHT can be selected as follows [15]:

- Initial temperature T. In order to make SA quasi balance in the very beginning, the initial acceptance ratio should approximate to 1, that is, $e^{-\Delta f/T} \approx 1$. In experiments taken the initial acceptance ratio as 0.95, we calculate Δf to obtain T via $e^{-\Delta f/T} = 0.95$, take average of them after repeating many times, here $T = 20$.
- Terminal temperature T_f. T_f should be small enough so that the acceptive probability is sufficient small, e.g., $e^{-\Delta f/T_f} = 1e-10$. Using above method we take $T_f = 0.01$.
- Temperature attenuate function α and the length of Markov chains L (Maximum iterative times $K\max$). L should make markov chain quasi balance on each temperature as α is given. If α is close to 1, two temperatures in succession will approximate mutually. Therefore, if the quasi balance in temperature T_k is satisfied, then, just a small amount of transformations are enough to reach quasi balance on the next temperature T_{k+1}, in this case, we choose a smaller L. Similarly, if α is smaller, L should be larger. Due to IHT costs little time in low dimension, to improve the quality of the solution, we choose $\alpha = 0.5, L = 100$ in signal reconstruction. However, for image reconstruction problem, IHT costs much time and most of the time wastes on calculating local minimal points. Therefore, n and ϵ in Step 2 of HHT should be decreased and increased respectively. At this rate, lesser time is spent on computing, yet the quality of the solution may be not good enough. To solve this problem, a higher precision IHT is taken in the wake of obtaining the minimal point of each temperature or the point satisfied the predetermined lower precision. At last if the precision ϵ_0 fails to achieve, HHT keeps going on, else stop. So L is much smaller, so is α. We choose $\alpha = 0.9. L = 20$.
- Random disturbance μ. In order to enable algorithm to jump out of a local minimal point, current point should be caused large enough disturbance. Here $\mu = 0.5$.
- n and ϵ in IHT. In signal reconstruction they take values as $1e7$, $1e-16$ respectively [18], while in image reconstruction n is 500 and ϵ is $1e-2$.

B. Signal Reconstruction

Here A in (2) is a gaussian random matrix, we obtain it by using the method proposed in [1]. And x is random initial sparse signal ($x \in R^N, N = 512$, 130 - sparse). We compare the sampling ability of IHT and HHT respecttively in first step. In second step noisy signal reconstruction is considered.

1) Noiseless Signal Reconstruction: Given the precision $\epsilon_0 = 1e-7$, the first experiments we did were to reconstruct initial signal x via noiseless observation vector b and A. The experiments test two algorithms' ability of accurate reconstructing x in different sampling number K. "Error" stands for $\| x-\hat{x}\|_2$, where $\hat{x}$ is the constructive signal. "Time" is the computational time and is computed by second. The initial point is zero vector.

From TABLE I, we can see when K is 330 or 300, both algorithms can accurately reconstruct signal, but time consumed by HHT is shorter than IHT. when K is in range of $280 \sim 250$, HHT still can accurately reconstruct signal while IHT can not, and the additional time is acceptable. We can conclude from TABLE I that HHT is superior to IHT in the effect of noiseless signal reconstruction.

In addition, we also did the experiments for signal reconstruction about HHT with random initial point. For each K, operate HHT ten times with different initial points which follow $N(0,1)$ normal distribution, then, take the average of Errors and times, and report them in TABLE II, which offers the evidence that HHT does not rely on the initial point. That is, we may say HHT is a global convergent algorithm.

2) Noisy Signal Reconstruction: In this subsection we add the white Gaussian noise $\varepsilon \sim N(0,\sigma^2)(\sigma = 0.1)$ in observation vector b to compare antinoise ability of both algorithms. What we employ comparison criteria is oracle mean square error method. The detail of its procedure is to

TABLE I. NOISELESS SIGNAL RECONSTRUCTION WITH ZERO INITIAL POINT

K	*Method*	*Error*	*Time(s)*
330	IHT	3.31e-7	0.18
	HHT	3.44e-7	0.09
300	IHT	3.87e-7	0.14

K	*Method*	*Error*	*Time(s)*
	HHT	3.52e-7	0.09
280	IHT	7.14	0.24
	HHT	**4.19e-7**	0.7
270	IHT	4.041	0.27
	HHT	**4.37e-7**	0.71
250	IHT	5.60	0.38
	HHT	**5.58e-7**	5.61

TABLE II. NOISELESS SIGNAL RECONSTRUCTION WITH RANDOM INITIAL POINT

K	*Method*	*Error*	*Time(s)*
270	HHT	4.76e-7	2.253
260	HHT	5.04e-7	5.955

compute oracle mean square error(Or) firstly, that is,

$$Or = tr((A^T A)^{-1})\sigma^2,$$

where $tr(\Phi)$ means trace of matrix Φ. Then, calculate the error ϵ_0 between reconstructive signal and initial signal, let $\rho = \epsilon_0 / Or$, For different algorithms, the more ρ approximates to 1, the better antinoise ability of the algorithm is. Here we appoint that if $\rho > 1.5$, the signal can't be reconstructed successfully in our opinion.

We now take a look at TABLE III, which records the reconstructed results in different sampling numbers with the initial point both are zero vector. Those data suggest that HHT is more outstanding in antinoise ability. For K between 262 and 280 instance, antinoise performance of HHT is better than IHT obviously. And with the decrease of the sampling number, though signal is more and more difficulty to denoise, HHT's antinoise performance seems stable and its disadvantage of computing time is gradual shrink too.

TABLE III. NOISY SIGNAL RECONSTRUCTION

K	*Or*	*Method*	*Error*	ρ	*Time(s)*
330	2.50	IHT	2.88	1.15	0.21
		HHT	2.37	0.95	3.26
300	2.97	IHT	4.22	1.42	0.234
		HHT	3.21	1.08	2.34
280	3.18	IHT	10.53	3.31	0.203
		HHT	4.08	**1.28**	1.71
270	3.51	IHT	9.18	2.62	0.203
		HHT	3.74	**1.07**	1.41
266	3.62	IHT	9.77	2.70	0.219
		HHT	4.67	**1.29**	0.55
262	3.76	IHT	11.61	2.58	0.203
		HHT	4.5	**1.20**	0.83

TABLE IV shows experimental results of reconstructed signal in different noise level. Fix sampling number $K = 330$, sparsity $T = 130$, we take σ as 0.1, 0.11, 0.12, 0.13 respectively. Experimental results show that with the increase of the noise, HHT's antinoise performance is still better than IHT's.

TABLE IV. NOISY SIGNAL RECONSTRUCTION

σ	*Or*	*Method*	*Error*	ρ	*Time(s)*
0.10	2.50	IHT	2.88	1.15	0.21
		HHT	2.37	**0.95**	3.26
0.11	3.05	IHT	4.11	1.35	0.42
		HHT	3.94	**1.29**	2.12
0.12	3.57	IHT	3.96	1.11	0.36
		HHT	3.37	**0.95**	3.96
0.13	4.15	IHT	3.48	0.84	0.31
		HHT	3.82	**0.92**	1.61

C. *Image Reconstruction*

Image reconstruction is widely applied in imaging technology such as radar imaging, medical imaging and so on. Take example for magnetic resonance imaging (MRI), an MRI scanner in effect takes slices from the two dimensional fourier domain of the image. In order to reduce scan time and the exposure of the patient to electromagnetic radiation, we'd better to take fewer measurements. Therefore, on condition that image can be constructed accurately, the smaller sampling number K is, the better reconstruction algorithm is. By using Standard Shepp-Logan phantom [18] as the image, we would like to examine whether HHT is still better than IHT in noiseless image construction.

Given the precision $\epsilon_0 = 1e-6$, we did the experiments with initial point was zero vector in three different resolutions, part of results were chosen and listed in TABLE V.

TABLE V shows the results HHT can accurately reconstruct image of each resolution in less sampling number and less time. And with the increase of image size , its superiority is more and more obvious. For the large scale image reconstruction, HHT can accurately reconstruct quickly too. We can see that in TABLE VI.

TABLE V. LOW DIMENSION IMAGE RECONSTRUCTION

Resolution	*K*	*Method*	*Time(s)*	*Error*
32×32	24	IHT	8.78	0.10
		HHT	8.45	**1.04e-6**
64×64	27	IHT	69.51	1.12
		HHT	36.56	**2.22e-6**
128×128	33	IHT	258.07	1.35
		HHT	88.45	**4.75e-6**

TABLE VI. HIGH DIMENSION IMAGE RECONSTRUCTION

Resolution	*K*	*Time(s)*	*Error*
256×256	50	75.99	8.18e-6
512×512	80	136.87	1.45e-5

IV. CONCLUSION AND FUTURE DIRECTIONS

Results on a series of experiments suggest that HHT provides a powerful reconstruction algorithm. It is more effective and efficient in signal and image construction than IHT, and is a stable, global convergent algorithm. As a potential improvement to HHT, SA and ITH parameters can be made more adaptively, the local optimizer can be invoked with a probability on current temperature, which can save some time in searching local minimal point. Similar to hard threshold, there are soft and mixed thresholds. So we can combine SA with them in next step.

ACKNOWLEDGMENT

This author's work is supported partly by theNSFC project (10231060) and Chinese Postdoctoral Fund.

REFERENCES

[1] D. Donoho, Compressed sensing, IEEE Transactions on Information Theory, Vol 52, pp. 1289-1306, 2006.

[2] R.~Baraniuk, Compressive sensing, IEEE Signal Processing Magazine, Vol 24(4), pp. 118-121, 2007.

[3] S. Mallat and Z. Zhang, Matching pursuits with time frequency dictionaries, IEEE Transactions on Signal Processing, Vol 41, pp. 3397-3415, 1993.

[4] J. Tropp and A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, Transactions on Information Theory, Vol 53(12), pp. 4655-4666, 2007.

[5] D. Needell and R. Vershynin, Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit, IEEE Journal of Selected Topics in Signal Processing, Vol 4(2), pp. 310-316, 2007.

[6] D. Needell and J. A. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, Applied and Computational Harmonic Analysis, Vol 26(3), pp. 301-321 ,2009.

[7] J. A. Tropp, Greed is good: algorithmic results for sparse approximation, IEEE Transactions on Information Theory, vol 50(10), pp. 2231-2242, 2004.

[8] R. Gribonval and P. Vandergheynst, On the exponential convergence of matching pursuits in quasi-incoherent dictionaries, IEEE Transactions on Information Theory, vol 52(1), pp. 255-261, 2006.

[9] S. S. Chen, D. L. Donoho and M. A. Saunders, Atomic decomposition by basis pursuit, SIAM Journal of Scientific Computing, vol 20(1), pp. 33–61, 1998.

[10] S. J. Kim, K. Koh, M. Lustig, S. Boyd, D. Gorinevsky. A method for large-scale l1-regularized least squares， IEEE Journal on Selected Topics in Signal Processing, Vol 4, pp. 606-617, 2007.

[11] M. A. T. Figueiredo, R. D. Nowak and S. J. Wright, Gradient projection for sparse reconstruction :Application to compressed sensing and other inverse problems, Journal of Selected Topics in Signal Processing: Special Issue on Convex Optimization Methods for Signal Processing Vol 1(4), pp. 586-598, 2007.

[12] T. Blumensath and M. Davies, Iterative hard thresholding for compressed sensing, Applied and Computational Harmonic Analysis, 2009.

[13] T. Blumensath and M. Davies, Iterative thresholding for sparse approximations, Journal of Fourier Analysis and Applications, Vol 10, pp. 629-654, 2008.

[14] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, Optimization by Simulated Annealing, Science, New Series, Vol 220, pp. 671-680, 1983.

[15] Lishan Kang, Yun Xie, Shiyong You, Zuhua Luo, Non-numerical parallel algorithms-simulated annealing algorithm, Science Press, 1994.

[16] Dingwei Wang, Junwei Wang, Hongfeng Wang, Intelligent optimization methods, Higher Education Press, 2007.

[17] Jinshan Jiang, Chunxiong He, Shaohua Pan, Optimization algorithm, South China University of Technology Press, 2007.

[18] T. Blumensath and M. E. Davies, Matlab code: "hard_l0_Mterm.m", http://www.personal.soton.ac.uk/tb1m08/sparsify/sparsify.html

[19] E. Cand□s, J. Romberg and T. Tao, Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information, IEEE Trans on Information Theory , Vol 52 (2), pp. 489-509, 2006.