

A STOCHASTIC ALTERNATING MINIMIZATION METHOD FOR SPARSE PHASE RETRIEVAL

CAI Jian-feng¹, JIAO Yu-ling², LU Xi-liang³, YOU Jun-tao^{1,4}

(1. Department of Mathematics, The Hong Kong University of Science and Technology Kowloon,
Hong Kong 999077, China)

(2. School of Statistics and Mathematics, Zhongnan University of Economics and Law,
Wuhan 430073, China)

(3. School of Mathematics and Statistics; Hubei Key Laboratory of Computational Science,
Wuhan University, Wuhan 430072, China)

(4. Department of Mathematics, Southern University of Science and Technology,
Shenzhen 518055, China)

Abstract: Sparse phase retrieval plays an important role in many fields of applied science and thus attracts lots of attention. In this paper, we propose a stochastic alternating minimization method for sparse phase retrieval (StormSpar) algorithm which empirically is able to recover n -dimensional s -sparse signals from only $O(s \log n)$ measurements without a desired initial value required by many existing methods. In StormSpar, the hard-thresholding pursuit (HTP) algorithm is employed to solve the sparse constrained least-square sub-problems. The main competitive feature of StormSpar is that it converges globally requiring optimal order of number of samples with random initialization. Extensive numerical experiments are given to validate the proposed algorithm.

Keywords: phase retrieval; sparse signal; stochastic alternating minimization method; hard-thresholding pursuit

2010 MR Subject Classification: 90C99

Document code: A **Article ID:** 0255-7797(2020)04-0379-10

1 Introduction

Phase retrieval is to recover the phase information from its magnitude measurements, i.e.,

$$y_i = |\langle a_i, x \rangle| + \epsilon_i, \quad i = 1, 2, \dots, m, \quad (1.1)$$

where $x \in \mathbb{F}^n$ is the unknown vector, $a_i \in \mathbb{F}^n$ are given sampling vectors which are random Gaussian vector in this paper, y_i are the observed measurements, ϵ_i are the noise, and m is

* **Received date:** 2020-01-19

Accepted date: 2020-06-15

Foundation item: The first author is partially supported by Hong Kong Research Grant Council (HKRGC) grant GRF 16306317; the second author is partially supported by National Science Foundation of China (NSFC) (11871474 ; 61701547); the third author is partially supported by NSFC (11871385).

Biography: Cai Jianfeng(1978–), male, born at Lixian, Hunan, professor, major in computational harmonic analysis.

the number of measurements (or the sample size). The \mathbb{F}^n can be \mathbb{R}^n or \mathbb{C}^n , and we consider the real case $\mathbb{F}^n = \mathbb{R}^n$ in this work. The phase retrieval problem arises in many fields like X-ray crystallography [1], optics [2], microscopy [3] and others, see e.g. [4]. Due to the lack of phase information, the phase retrieval problem is a nonlinear and ill-posed problem.

When the measurements are overcomplete, i.e., $m > n$, there are many algorithms in the literature. Earlier approaches were mostly based on alternating projections, e.g. the work of Gerchberg and Saxton [5] and Fienup [4]. Recently, convex relaxation methods such as phase-lift [6] and phase-cut [7] were proposed. These methods transfer the phase retrieval problem into a semi-definite programming, which can be computationally expensive. Another convex approach named phase-max which does not lift the dimension of the signal was proposed in [8]. In the meanwhile, there are other works based on solving nonconvex optimization via first and second order methods including alternating minimization [9] (or Fienup methods), Wirtinger flow [6] Kaczmarz [10], Riemannian optimization [11]; Gauss-Newton [12, 13] etc. With a good initialization obtained via spectral methods, the above mentioned methods work with theoretical guarantees. Progresses were made by replacing the desired initialization with random initialized ones in alternating minimization [14, 15], gradient descent [16] and Kaczmarz method [17, 18] while keeping convergence guarantee with high probability. Also, recent analysis in [19, 20] showed that some nonconvex objective functions for phase retrieval have a nice landscape — there is no spurious local minima — with high probability. As a consequence, for these objective functions, any algorithms finding a local minima are guaranteed to give a successful phase retrieval.

For the large scale problem, the requirement $m > n$ becomes unpractical due to the huge measurement and computation cost. In many applications, the true signal x is known to be sparse. Then the sparse phase retrieval problem can be solved with a small number of samplings, thus possible to be applied to large scale problems. It was proved in [21] that $m = O(s \log n/s)$ measurement is sufficient to ensure successful recovery in theory with high probability when the model is Gaussian (i.e., the sampling vector a_i are i.i.d Gaussian and the target is real). But the exiting computational trackable algorithms require $O(s^2 \log n)$ number of measurements to reconstruct the sparse signal, for example, ℓ_1 regularized PhaseLift method [22], sparse AltMin [9], GESPAR [23], Thresholding/projected Wirtinger flow [24, 25], SPARTA [26] and so on. Two stage methods based on phase-lift and compressing has been introduced in [27, 28], which is able to do successful reconstruction with $O(s \log n)$ measurements for some specially designed sampling matrix which exclude the Gaussian model (1.1). When a good initialization is available, the sample complexity can be improved to $O(s \log n)$ [25, 29]. However, it requires $O(s^2 \log n)$ samples to get a desired sparse initialization in the existing literature. This gap naturally raises the following challenging question.

Can one recover the s -sparse target from the phaseless generic Gaussian model (1.1) with $O(s \log n)$ measurements via just using random initializations ?

In this paper, we propose a novel algorithm to solve the sparse phase retrieval problems

in the very limited number of measurements (numerical examples show that $m = O(s \log n)$ can be enough). The algorithm is a stochastic version of alternating minimizing method. The idea of alternating minimization method is: during each iteration, we first given an estimation of the phase information, then substitute the approximated phase into (1.1) with the sparse constraint and solve a standard compressed sensing problem to get an updated sparse signal. But since the alternating minimization method is a local method, it is very sensitive to the initialization. Without enough measurements, it is very difficult to compute a good initial guess. To overcome this difficulty, we change the sample matrix during each iteration via bootstrap technique, see Algorithm 1 for details. The numerical experiments show that the proposed algorithm needs only $O(s \log n)$ measurements to recover the true signal with high probability in Gaussian model, and it works for a random initial guess. The experiments also show that the proposed algorithm is able to recover signal in a wide range of sparsity.

The rest of this paper is organized as follows. In Section 2 we introduce the setting of problem and the details of the algorithm. Numerical experiments are given in Section 3.

2 Algorithm

First, we introduce some notations. For any $a, b \in \mathbb{R}^n$, we denote that $a \odot b = (a_1 b_1, a_2 b_2, \dots, a_n b_n)$, $\|x\|_0$ is the number of nonzero entries of x , and $\|x\|_2$ is the standard l_2 -norm, i.e., $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$. The floor function $\lfloor c \rfloor$ is the greatest integer which is less than or equal to c .

Recall from (1.1), we denote the sampling matrix and the measurement vector by $A = [a_1^t, \dots, a_m^t] \in \mathbb{R}^{m \times n}$ and $y = [y_1, \dots, y_m] \in \mathbb{R}^m$, respectively. Let $x \in \mathbb{R}^n$ be the unknown sparse signal to be recovered. In the noise free case, the problem can be written as to find x such that

$$y = |Ax| \quad \text{s.t. } \|x\|_0 \leq s.$$

In the noisy case, this can be written by the nonconvex minimization problem

$$\min_x \frac{1}{2} \|y - |Ax|\|_2^2 \quad \text{s.t. } \|x\|_0 \leq s. \quad (2.1)$$

Now we propose the stochastic alternating minimization method for sparse phase retrieval (StormSpar) as follows. It starts with a random initial guess x^0 . In the ℓ -th step of iteration ($\ell = 1, 2, \dots$), we first randomly choose some rows of the sampling matrix A to form a new matrix A^ℓ (which is a submatrix of A), and denoted by the corresponding rows of y to y^ℓ . Then we compute the phase information of $A^\ell x^{\ell-1}$, say $p^\ell = \text{sign}(A^\ell x^{\ell-1})$, and to solve the standard compressed sensing subproblem

$$\min_x \frac{1}{2} \|A^\ell x - \tilde{y}^\ell\|^2 \quad \text{s.t. } \|x\|_0 \leq s, \quad (2.2)$$

where $\tilde{y}^\ell = p^\ell \odot y^\ell$. Problem (2.2) can be solved by a lot of compressed sensing solver, and we will use the efficient Hard Thresholding Pursuit (HTP) [30] in our algorithm. For

completion, HTP is given in Algorithm 2. We summarize the StormSpar algorithm in the Algorithm 1.

3 Numerical Results and Discussions

3.1 Implementation Details

The true signal x is chosen as s -sparse with random support and the design matrix $A \in \mathbb{R}^{m \times n}$ is chosen to be random Gaussian matrix. The additive Gaussian noise following the form $\epsilon = \sigma * \text{randn}(n, 1)$, thus the noise level is determined by σ . The parameter γ is set to be $\min(\frac{s}{m} * \log \frac{n}{0.001}, 0.6)$, and $\delta = 0.01$.

The estimation error r between the estimator \hat{x} and the true signal x is defined as

$$r = \min\{\|\hat{x} + x\|_2, \|\hat{x} - x\|_2\} / \|x\|_2.$$

We say it is a successful recovery when the relative estimation error r satisfy that $r \leq 1e-2$ or the support is exactly recovered. The tests repeat independently for 100 times to compute a successful rate. ‘‘Aver Iter’’ in Tables 1 and 2 means the average number of iterations for 100 times of tests. All the computations were performed on an eight-core laptop with core i7 6700HQ@3.50 GHz and 8 GB RAM using MATLAB 2018a.

Algorithm 1 StormSpar

- 1: Input: Normalized $A \in \mathbb{R}^{m \times n}$, y , sparsity level s , $\gamma \in (0, 1)$, small constant δ , a random initial value x^0 .
 - 2: **for** $\ell = 1, 2, \dots$ **do**
 - 3: Randomly selected $\lfloor \gamma m \rfloor$ rows of A and y , denote the index as i^ℓ , to form $A^\ell = A(i^\ell, :)$ $y^\ell = y(i^\ell)$.
 - 4: Compute $p^\ell = \text{sign}(A^\ell x^{\ell-1})$, $\tilde{y}^\ell = p^\ell \odot y^\ell$.
 - 5: Get x^ℓ by solving $\min_{x, \|x\|_0 \leq s} \frac{1}{2} \|A^\ell x - \tilde{y}^\ell\|^2$ via Algorithm 2 (HTP).
 - 6: Check stop criteria $\|x^\ell - x^{\ell-1}\| \leq \delta$.
 - 7: **end for**
 - 8: Get the first s position of x^ℓ and refit on it as output.
-

Algorithm 2 HTP solving (2.2)

- 1: Input: Initialization: $k = 0, x^0 = 0$;
 - 2: **for** $k = 1, 2, \dots$ **do**
 - 3: $S^k \leftarrow \{\text{indices of } s \text{ largest entries of } x^{k-1} + \mu(A^\ell)^t(\tilde{y}^\ell - A^\ell x^{k-1})\}$;
 - 4: Solve $x^k \leftarrow \text{argmin}_{\text{supp}(x) \subset S^k} \|A^\ell x - \tilde{y}^\ell\|_2$.
 - 5: **end for**
-

3.2 Examples

Example 1 First we examine the effect of sample size m to the probability of successful recovery in Algorithm 1. The dimension of the signal x is $n = 1000$.

a) When we set sparsity to be $s = 10, 25, 50$, Figure 1 shows how the successful rate changes in terms of the sample size m . In this experiment, we fix a number $K = \lfloor (s(\log n + \log \frac{1}{0.01})) \rfloor$, which is 115, 287, 575 with respect to the sparsity 10, 25, 50. Then we compute the probability of success when m/K changes: for each s and each $m/K = 1, 1.25, \dots, 3$, we run our algorithm for 100 times. It shows when the sample size is in order $O(s \log n)$ in this setting, we can recover the signal with high possibility.

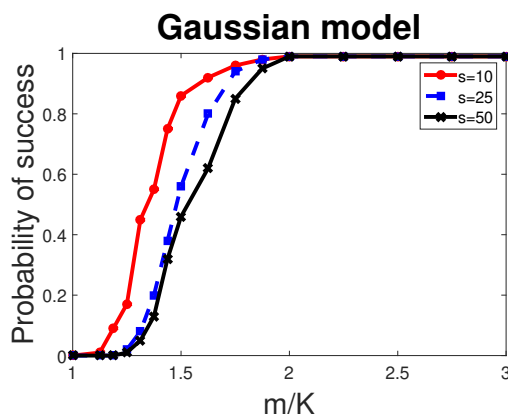


Figure 1 The probability of success in recovery v.s. sample size m/K for Gaussian model, $K = \lfloor (s(\log n + \log \frac{1}{0.01})) \rfloor$ which is 115, 287, 575 with respect to sparsity $s = 10, 25, 50$, signal dimension $n = 1000$, noise level $\sigma = 0.01$

b) We compare StormSpar to some existing algorithm, i.e., CoPRAM [31], Thresholded Wirtinger Flow (ThWF) [24] and SPArse truncated Amplitude flow (SPARTA) [26]. The sparsity is set to be 30 and the model is noise free. Figure 2 shows the successful rate comparison in terms of sample size, the results are obtained by averaging the results of 100 trials. We find it that StormSpar requires more iterations and more cpu time than these algorithms which requires initialization. But StormSpar achieves better accuracy with less sample complexity.

Example 2 Figure 3 shows that StormSpar is robust to noise. We set $n = 1000$, $s = 20$, and $m = \lfloor (2.5s(\log n + \log \frac{1}{0.01})) \rfloor (= 575)$. The noise we added is i.i.d. Gaussian, and the noise level is shown by signal-to-noise ratios (SNR), we plot the corresponding relative error of reconstruction in the Figure 3. The results are obtained by average of 100 times trial run.

Example 3 We compare StormSpar with a two-stage method Phaselift+BP proposed in [27], which has been shown to be more efficient than the standard SDP of [32]. The dimension of data is set to be $n = 1000$. The comparison are two-folder. First, for different sparsity level, we compare the minimum number of measurements required to give successful

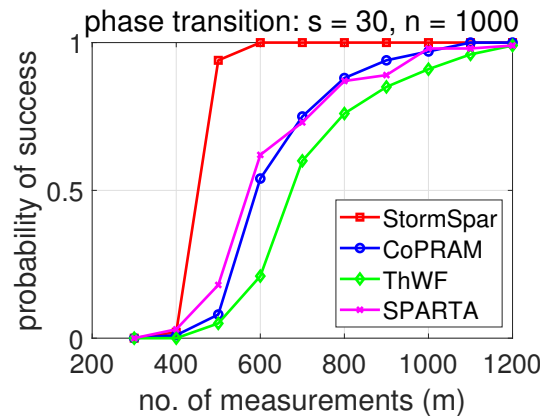


Figure 2 The probability of success in recovery for different algorithms in terms of changing sample size, dimension $n = 1000$, sparsity $s = 30$ and the model is noise free

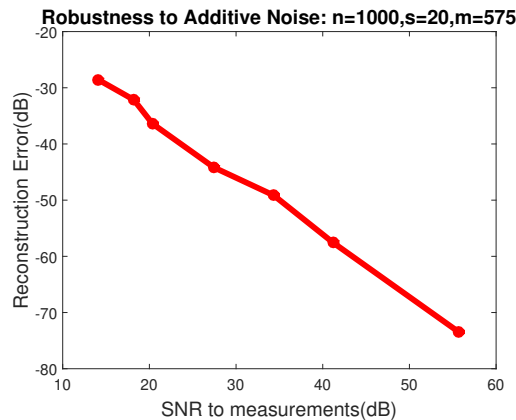


Figure 3 The reconstruction error v.s. SNR to measurements for Gaussian model, $m = \lfloor (2.5s(\log n + \log \frac{1}{0.01})) \rfloor = 575$ with sparsity $s = 20$, signal dimension $n = 1000$ and several noise level, i.e. SNR to measurements

recovery rate higher than 95% , the result can be found in Figure 4. Second the average computational time is given in Figure 5, where $m = \lfloor (2.5s(\log n + \log \frac{1}{0.01})) \rfloor$.

Example 4 Let $m = O(s \log n)$, we test for different sparse levels and different dimensions. In Table 1, we fix dimension $n = 2000$, and the sample size is chosen to be $m = \lfloor (2.5s(\log n + \log \frac{1}{0.01})) \rfloor$. The sparsity level changes from 5 to 100, we find the algorithm can successfully recover the sparse signal in most case, and the iteration number is very stable.

In Table 2, the sparsity level is fixed by $s = 10$, the sample size is $m = \lfloor (2.5s(\log n + \log \frac{1}{0.01})) \rfloor$ for dimension n from 100 to 10000. We find the algorithm can successfully recover the sparse signal in most cases, and the number of iteration dependent on the dimension n .

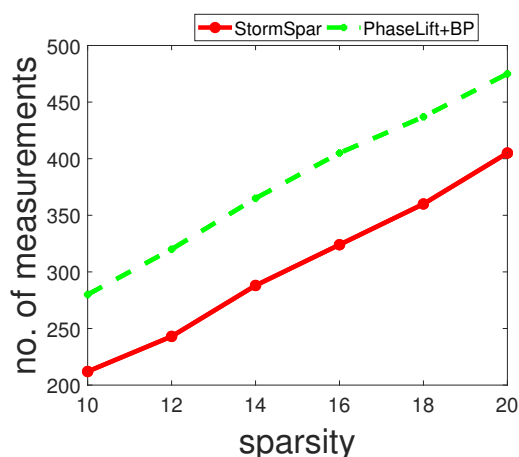


Figure 4 Comparison of Minimum number of measurements required for Gaussian model, signal dimension $n = 1000$ and free of noise

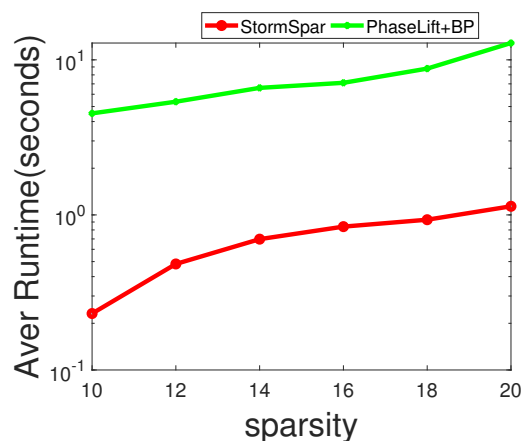


Figure 5 Comparison of efficiency for Gaussian model, signal dimension $n = 1000$ and free of noise

4 Conclusion

In this paper, we propose a novel algorithm (StormSpar) for the sparse phase retrieval. StormSpar start with a random initialization and employ a alternating minimization method for a changing objective function. The subproblem $\min_{x, \|x\|_0 \leq s} \frac{1}{2} \|A^\ell x - \tilde{y}^\ell\|^2$ is a standard compressed sensing problem, which can be solved by HTP method. Numerical examples show that the proposed algorithm requires only $O(s \log n)$ samples to recover the s -sparse signal with a random initial guess.

Table 1 Numerical results for sparsity test, with random sampling A of size $n \times m$, $n = 2000$, $m = \lfloor (2.5s(\log n + \log \frac{1}{0.01})) \rfloor$, s is the sparsity, with $\sigma = 0.01$, and Aver Iter= \lfloor average number of iterations for 100 times of test \rfloor

Dimension n	Sparsity s	Sample m	Successful Rate	Aver Iter
2000	5	152	98%	109
2000	10	305	99%	229
2000	15	457	99%	359
2000	20	610	98%	395
2000	25	762	100%	407
2000	30	915	99%	403
2000	35	1068	100%	482
2000	40	1220	100%	331
2000	45	1373	100%	305
2000	50	1525	100%	324
2000	75	2288	100%	289
2000	100	3051	100%	285

Table 2 Numerical results for different dimensions, with random sampling A of size $n \times m$, $m = \lfloor (2.5s(\log n + \log \frac{1}{0.01})) \rfloor$, s is the sparsity, with $\sigma = 0.01$, and Aver Iter= \lfloor average number of iterations for 100 times of test \rfloor

Dimension n	Sparsity s	Sample m	Successful Rate	Aver Iter
100	10	230	98%	38
200	10	249	99%	46
300	10	257	100%	56
400	10	264	100%	72
500	10	270	100%	93
750	10	280	100%	123
1000	10	287	100%	157
1500	10	297	99%	192
2000	10	305	99%	229
3000	10	315	99%	298
4000	10	322	98%	508
5000	10	328	97%	748
7500	10	338	95%	1142
10000	10	345	96%	1271

References

- [1] Harrison R W. Phase problem in crystallography[J]. *JOSA*, 1993, 10(5): 1046–1055.
- [2] Walther A. The question of phase retrieval in optics[J]. *Journal of Modern Optics*, 1963, 10(1): 41–49.
- [3] Miao J, Ishikawa T, Shen Q, Earnest T. Extending x-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes[J]. *Annu. Rev. Phys. Chem.*, 2008, 59: 387–410.
- [4] Fienup J R. Phase retrieval algorithms: a comparison[J]. *Applied Optics*, 1982, 21(15): 2758–2769.
- [5] Gerchberg R W. A practical algorithm for the determination of the phase from image and diffraction plane pictures[J]. *Optik*, 1972, 35: 237–246.
- [6] Candes E J, Eldar Y C, Strohmer T, Vershynski V. Phase retrieval via matrix completion[J]. *SIAM Review*, 2015, 57(2): 225–251.
- [7] Waldspurger I, d’Aspremont A, Mallat S. Phase recovery, maxcut and complex semidefinite programming[J]. *Mathematical Programming*, 2015, 149(1-2): 47–81.
- [8] Goldstein T, Studer C. Phasemax: convex phase retrieval via basis pursuit[J]. *IEEE Transactions on Information Theory*, 2018, 64(4): 2675–2689.
- [9] Sanghavi S, Netrapalli P, Jain P. Phase retrieval using alternating minimization[J]. *IEEE Transactions on Signal Processing: a Publication of the IEEE Signal Processing Society*, 2015, 63(18): 4814–4826.
- [10] Wei K. Solving systems of phaseless equations via kaczmarz methods: a proof of concept study[J]. *Inverse Problems*, 2015, 31(12): 125008.
- [11] Cai J F, Wei K. Solving systems of phaseless equations via Riemannian optimization with optimal sampling complexity[J]. *arXiv preprint arXiv:1809.02773*, 2018.
- [12] Gao B, Xu Z. Gauss-Newton method for phase retrieval[J]. *IEEE Transactions on Signal Processing*, 2017, 65(22): 5885–5896.
- [13] Ma Chao, Liu Xin, Wen Zaiwen. Globally convergent levenberg-marquardt method for phase retrieval[J]. *IEEE Transactions on Information Theory*, 2019, 65(4): 2343–2359.
- [14] Waldspurger I. Phase retrieval with random gaussian sensing vectors by alternating projections [J]. *IEEE Transactions on Information Theory*, 2018, 64(5): 3301–3312.
- [15] Zhang Teng. Phase retrieval using alternating minimization in a batch setting[J]. *Applied and Computational Harmonic Analysis*, 2020, 49(1): 279–295.
- [16] Chen Yuxin, Chi Yuejie, Fan Jianqiang, Ma Cong. Gradient descent with random initialization: fast global convergence for nonconvex phase retrieval[J]. *Mathematical Programming*, 2019, 176(1–2): 5–37.
- [17] Tan Y S, Vershynin R. Phase retrieval via randomized kaczmarz: theoretical guarantees[J]. *Information and Inference: a Journal of the IMA*, 2018, 8(1): 97–123.
- [18] Jeong H, GÜntÜrk C S. Convergence of the randomized kaczmarz method for phase retrieval[J]. *arXiv preprint arXiv:1706.10291*, 2017.
- [19] Sun J, Qu Q, Wright J. A geometrical analysis of phase retrieval[J]. *Foundations of Computational Mathematics*, 2018, 18(5): 1131–1198.
- [20] Li Z, Cai J F, Wei K. Towards the optimal construction of a loss function without spurious local minima for solving quadratic equations[J]. *arXiv preprint arXiv:1809.10520*, 2018.
- [21] Eldar Y C, Mendelson S. Phase retrieval: Stability and recovery guarantees[J]. *Applied and Computational Harmonic Analysis*, 2014, 36(3): 473–494.

- [22] Li X, Voroninski V. Sparse signal recovery from quadratic measurements via convex programming[J]. *SIAM Journal on Mathematical Analysis*, 2013, 45(5): 3019–3033.
- [23] Shechtman Y, Beck A, Eldar Y C. Gspar: Efficient phase retrieval of sparse signals[J]. *IEEE Transactions on Signal Processing*, 2014, 62(4): 928–938.
- [24] Cai T T, Li X, Ma Z. Optimal rates of convergence for noisy sparse phase retrieval via thresholded wirtinger flow[J]. *The Annals of Statistics*, 2016, 44(5): 2221–2251.
- [25] Soltanolkotabi M. Structured signal recovery from quadratic measurements: Breaking sample complexity barriers via nonconvex optimization[J]. *IEEE Transactions on Information Theory*, 2019, 65(4): 2374–2400.
- [26] Wang G, Zhang L, Giannakis G B, Akçakaya M, Chen J. Sparse phase retrieval via truncated amplitude flow[J]. *IEEE Transactions on Signal Processing*, 2018, 66(2): 479–491.
- [27] Iwen M, Viswanathan A, Wang Y. Robust sparse phase retrieval made easy[J]. *Applied and Computational Harmonic Analysis*, 2017, 42(1): 135–142.
- [28] Bahmani S, Romberg J. Efficient compressive phase retrieval with constrained sensing vectors[J]. *Advances in Neural Information Processing Systems*, 2015, 28: 523–531.
- [29] Hand P, Voroninski V. Compressed sensing from phaseless gaussian measurements via linear programming in the natural parameter space[J]. *arXiv preprint arXiv:1611.05985*, 2016.
- [30] Foucart S. Hard thresholding pursuit: an algorithm for compressive sensing[J]. *SIAM Journal on Numerical Analysis*, 2011, 49(6): 2543–2563.
- [31] Jagatap G, Hegde C. Sample-efficient algorithms for recovering structured signals from magnitude-only measurements[J]. *IEEE Transactions on Information Theory*, 2019, 65(7): 4434–4456.
- [32] Ohlsson H, Yang A Y, Dong R, Sastry S S. CPRL-an extension of compressive sensing to the phase retrieval problem[J]. *Advances in Neural Information Processing Systems*, 2012, 25(2): 1367–1375.

求解稀疏相位恢复问题的随机交替方向法

蔡剑锋¹, 焦雨领², 吕锡亮³, 游俊韬^{1,4}

(1. 香港科技大学数学系, 香港 999077)

(2. 中南财经政法大学统计与数学学院, 湖北 武汉 430073)

(3. 武汉大学数学与统计学院; 计算科学湖北省重点实验室, 湖北 武汉 430072)

(4. 南方科技大学数学系, 广东 深圳 518055)

摘要: 近年来稀疏相位恢复问题受到了越来越多的关注. 本文提出了一种随机交替方向法方法求解稀疏相位恢复问题, 该算法采用硬阈值追踪算法求解带稀疏约束的最小二乘子问题. 大量的数值实验表明, 该算法可以通过 $O(s \log n)$ 次测量(理论上最少测量值)稳定的恢复 n 维 s 稀疏向量, 并且在随机初值下可以获得全局收敛性.

关键词: 相位恢复; 稀疏信号; 随机交替方向法; 硬阈值追踪

MR(2010)主题分类号: 90C99 中图分类号: O242.1; O241.5