AN ALPS VIEW OF SPARSE RECOVERY

*Volkan Cevher**

Laboratory for Information and Inference Systems

ABSTRACT

We provide two compressive sensing (CS) recovery algorithms based on iterative hard-thresholding. The algorithms, collectively dubbed as algebraic pursuits (ALPS), exploit the restricted isometry properties of the CS measurement matrix within the algebra of Nesterov's optimal gradient methods. We theoretically characterize the approximation guarantees of ALPS for signals that are sparse on ortho-bases as well as on tight-frames. Simulation results demonstrate a great potential for ALPS in terms of phase-transition, noise robustness, and CS reconstruction.

1. INTRODUCTION

An application *du jour* in signal processing is compressive sensing (CS), which revolves around the classical underdetermined linear regression problem in learning, statistics and computer science:

$$u = \Phi x^* + n, \tag{1}$$

where $u \in \mathbb{R}^N$ is the vector of compressive samples, $\Phi \in \mathbb{R}^{M\times N}$ is the measurement matrix, and $n \in \mathbb{R}^M$ is the noise. Dimensionality reduction loses information since Φ has a non-trivial null-space; hence, CS exploits sparse representations to arbitrate the true solution among infinitely many vectors that can produce the same u.

By sparse representations, we mean one one of the following cases depending on the context. $x \in \mathbb{R}^N$ has a *synthesis*-sparse representation as $x = E\alpha$ in $E \in \mathbb{R}^{N\times N'}$ ($N' \leq N$), when $K \ll N'$ coefficients of α can well-approximate the signal x. $x \in \mathbb{R}^N$ has an *analysis*-sparse representation as $\alpha = Dx$ in $D \in \mathbb{R}^{N'\times N}$ ($N' \geq N$), when $K \ll N'$ coefficients of α can well-approximate the signal as $x = E\alpha_{|K}$, where E is the left inverse of D. Example representations include wavelets for synthesis, and overcomplete Gabor dictionary for analysis. Such synthesis and analysis formulations in fact go well beyond CS, touching on every problem that has been dealt with using sparse and redundant representation modeling. In the sequel, we will assume that we have an orthonormal basis for synthesis representations, or a tight-frame for analysis representations; hence, $E = D^T$.

In this paper, we are concerned with the following problem:

$$x^* = \text{argmin}_{x:\|Dx\|_0=K} f(x), \;\; f(x) = \|u - \Phi x\|_2^2; \tag{2}$$

where $\|x\|_0$ counts the number of non-zeros of x. This problem is related to the subset selection in statistical modeling, minimum weight solutions in error corrective coding, and the minimum set cover in computer science. Unfortunately, (2) is *NP*-hard.

To circumvent this intractability, we assume that the measurement matrix provides *stable embedding* (SE) for all $x_j \in \Sigma_{K_j}$ (i.e., Σ_{K_j} is the union of all subspaces spanned by all subsets of K_j columns of E) with constants $\mu_{K'}$ and $L_{K'}$ ($K' = K_1 + K_2$):

$$\frac{\mu_{K'}}{2}\|x_1 - x_2\|_2^2 \leq \|\Phi(x_1 - x_2)\|_2^2 \leq \frac{L_{K'}}{2}\|x_1 - x_2\|_2^2. \tag{3}$$

When E is an ortho-basis, random matrices Φ satisfy the SE in (3) with $M = \mathcal{O}(K\log(N'/K))$. This is the well-known K-restricted isometry property (K-RIP) in CS [1]. When E is an overcomplete dictionary, recent results show that random matrices Φ also satisfy the SE in (3) with $M \gtrsim K\log(N'/K)$. This property is known as the dictionary-RIP (D-RIP) [2, 3].

To solve (2), we propose two algorithms, collectively dubbed as algebraic pursuits (ALPS). The algorithms are called as Lipschitz iterative hard-thresholding (LIHT) and fast Lipschitz iterative hard-thresholding (FLIHT). Loosely speaking, these algorithms are hard-thresholding interpretations of the popular ISTA and FISTA algorithms [4] that are based on the algebra used in Nesterov's optimal gradient and smoothing techniques [5, 6]. Under certain restrictions on (L, μ) we discuss in the main text, the i-th iterate x_i of our algorithms satisfies:

$$\|x_i - x^*\|_2 \leq C(i) + c\|n\|_2, \tag{4}$$

where $C(i) = \mathcal{O}(\rho^i)$ for some $|\rho| < 1$; and, c is an algorithm dependent, absolute constant.

Notation. We use the ℓ_2-norm $\|\cdot\|_2$ as $\|\cdot\|$ throughout, unless otherwise stated. By objective function, we specifically mean $f(x)$ as defined in (2). We use $\nabla f(x)$ to denote the usual gradient of the objective $f(x)$ with respect to x. The bracket notation $\langle x, y\rangle = x^T y$ refers to the usual inner product in ℓ_2.

Organization. In Section 2, we set up key properties of objective function that the later sections build upon. Section 3 provides the algorithms and characterize their approximation guarantees using the SE assumption. Section 4 illustrates the compressive sensing performance of ALPS. Section 5 wraps up the paper, followed by an appendix of technical derivations for the main statements.

2. OPTIMIZATION OBJECTIVE: RELEVANT PROPERTIES AND INEQUALITIES

Properties. In this section, we highlight key properties and inequalities for the objective function in a series of lemmata, which are instrumental in obtaining the convergence and recovery guarantees of ALPS. First, we need to define the Bregman distance based on our objective: $B(y, x) = f(y) - f(x) - \langle\nabla f(x), y - x\rangle$, which will play a key role in the derivation of the guarantees.

Lemma 1. *The Bregman distance B satisfies*

$$\begin{array}{llcll} (1) & B(x_2, x_1) & = & \|\Phi(x_2 - x_1)\|^2 & \forall x_j \in \mathbb{R}^N, \\ (2) & B(x_2, x_1) & \leq & \frac{L}{2}\|x_2 - x_1\|^2 & L = 2\|\Phi\|, \forall x_j \in \mathbb{R}^N, \\ (3) & B(x_2, x_1) & \leq & \frac{L_{K'}}{2}\|x_2 - x_1\|^2 & \forall x_j \in \Sigma_{K_j}, \\ (4) & B(x_2, x_1) & \geq & \frac{\mu_{K'}}{2}\|x_2 - x_1\|^2 & \forall x_j \in \Sigma_{K_j}; \end{array} \tag{5}$$

*VC is with Ecole Polytechnique Federale de Lausanne, with a joint appointment at the Idiap Research Institute. Email: volkan.cevher@epfl.ch.
This work was supported in part by part by the European Commission under MIRG-268398 and DARPA KeCoM program #11-DARPA-1055. VC also would like to acknowledge Rice University for his Faculty Fellowship.

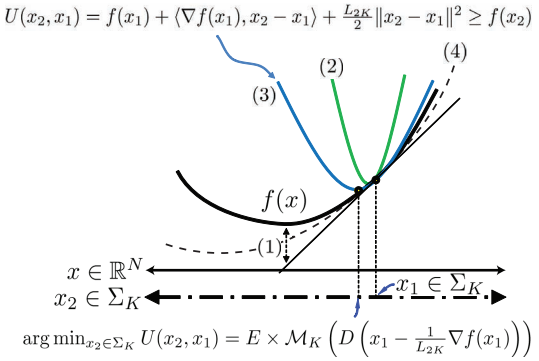


Fig. 1. The objective function $f(x) = \|u - \Phi x\|^2$ exhibits special geometric structure because Φ has stable embedding; it is possible to bound the objective with tractable quadratic functions.

where $K' = K_1 + K_2$. These equations directly follow from simple linear algebra and the SE assumption.

In general, Lemma 1 (2) and (3) provide the Lipschitz continuous gradient constant L and its sparsity-restricted version $L_{K'}$ of the objective function. Lemma 1 (4) is then related to its sparsity-restricted strong convexity constant $\mu_{K'}$. These parameters are fairly easy to calculate using concentration-of-measures for various matrix ensembles, when E is an orho-basis [7]: For example, when Φ is generated iid Gaussian and is column normalized to unit norm, we have $L_{K'} \leq 2(1 + \sqrt{K'/M} + \sqrt{2t/M})^2$, and $\mu_{K'} \geq 2(1 - \sqrt{K'/M} - \sqrt{2t/M})^2$ with probability $1 - \mathrm{e}^{-t}$.

Inequalities. We define the following upper-bound $U(x_2, x_1)$ to $f(x)$ at x_1 based on the Bregman distance in Lemma 1(3):

$$U(x_2,x_1) = f(x_1) + \langle \nabla f(x_1), x_2 - x_1 \rangle + \frac{L_{K'}}{2}\|x_2 - x_1\|^2, \quad (6)$$

where $x_1 \in \Sigma_{K_1}$ and $x_2 \in \Sigma_{K_2}$ and $K' = K_1 + K_2$ (c.f., Fig. 1 for visualization where $K_1 = K_2 = K$). Note that $U(x_2, x_1) \geq f(x_2)$. The special structure of this upper-bound will enable us to propose a majorization-minimization approach later on: instead of solving (2) directly, we will solve a sequence of simpler problems.

To obtain the minimum of the upper-bound (6), we exploit that E is an ortho-basis or a tight frame: $\arg\min_{x_2 : x_2 \in \Sigma_K} U(x_2, x_1)$

$$\begin{aligned}
&\doteq \arg \min_{x_2 : \|Dx_2\|_0 = K} \left\langle E^T \nabla f(x_1), D(x_2 - x_1) \right\rangle + \frac{L_{K'}}{2}\|x_2 - x_1\|^2 \\
&\doteq E \times \arg \min_{\alpha : \|\alpha\|_0 = K} \left\langle E^T \nabla f(x_1), \alpha - Dx_1 \right\rangle + \frac{L_{K'}}{2}\|\alpha - Dx_1\|^2 \\
&\doteq E \times \arg \min_{\alpha : \|\alpha\|_0 = K} \left\| \alpha - \left(Dx_1 - \frac{1}{L_{K'}} E^T \nabla f(x_1) \right) \right\|,
\end{aligned} \quad (7)$$

where we have used $\|x_2 - x_1\| = \|D(x_2 - x_1)\|$. The global optimum of the seemingly difficult, non-convex problem in (7) can be efficiently obtained by hard-tresholding: $\bar{x}_1 :=$

$$\arg \min_{x_2 \in \Sigma_K} U(x_2,x_1) = E \times \mathcal{M}_K \left(D \left(x_1 - \frac{1}{L_{K'}} \nabla f(x_1) \right) \right), \quad (8)$$

where $\mathcal{M}_K$ sorts the coefficients in decreasing magnitude, and zeros out all but the largest K coefficients.

We are now ready to state the key ingredient that leads to the approximation guarantees of LIHT and FLIHT:

Lemma 2. *The vector $\bar{x}_1 \in \Sigma_K$ in (8) satisfies the following*

$$\|u - \Phi \bar{x}_1\| \leq \sqrt{\frac{L_{K'} - \mu_{K'}}{2}} \|x^* - x_1\| + \|n\|, \quad (9)$$

where $K' = K + K_1$. Moreover, it holds that

$$\|x^* - \bar{x}_1\| \leq \sqrt{\frac{L_{K'} - \mu_{K'}}{\mu_{2K}}} \|x^* - x_1\| + \frac{2\|n\|}{\sqrt{\mu_{2K}/2}}. \quad (10)$$

A proof of Lemma 2 can be found in the appendix.

3. ALGEBRAIC PURSUITS

Lipschitz Iterative Hard Tresholding (LIHT) Iterative hard thresholding (IHT) is an popular method in CS with the following recursion [8] ($\tilde{\Phi} = \Phi E$): $\alpha_{i+1} = \mathcal{M}_K \left(\alpha_i + \tilde{\Phi}^T (u - \tilde{\Phi} \alpha_i) \right)$. IHT is quite easy to implement and presents a feasible alternative to convex optimization approaches for large-scale problems. When the measurement matrices have K-RIP, IHT's convergence rate is characterized to be linear up to a certain noise level [8].

In the light of the quadratic upperbound (6), LIHT starts with some $x_i \in \Sigma_K$, selects the argument of the minimum of $U(\cdot, x_i)$ as its next anchor point, and then iterates with the recursion below

$$x_{i+1} = E \times \mathcal{M}_K \left(D \left(x_i - \frac{1}{L_{2K}} \nabla f(x_i) \right) \right). \quad (11)$$

The algorithm has the word Lipschitz in its name due to the presence of the restricted Lipschitz gradient constant L_{2K} as the step size in front of the gradient term. Nesterov shows that this step size is basically optimal for the convergence speed of the gradient type of methods; any other step size is proven to have slower converge [5, 6]. Note that the IHT algorithm can be viewed as $\alpha_{i+1} = \mathcal{M}_K (\alpha_i - 0.5 \nabla f(\alpha_i))$, which takes a different step size than LIHT. When D is an ortho-basis, LIHT algorithm is the same as the GraDes algorithm [9].

LIHT has the following approximation guarantee, whose proof can be found in the appendix:

Theorem 1. *Define $\rho = \sqrt{\frac{L_{2K} - \mu_{2K}}{\mu_{2K}}}$ and $L_{2K} < 2\mu_{2K}$ so that $\rho < 1$. Then, the iterations of LIHT are contractive, and the i-th iteration satisfies*

$$\|x^* - x_i\| \leq \rho^i \|x - x_0\| + \frac{2\sqrt{2}(1 - \rho)^{-1}}{\sqrt{\mu_{2K}}} \|n\|. \quad (12)$$

Fast Lipschitz Iterative Hard Tresholding (FLIHT) While LIHT represents significant promise in flexibility and scalability for solving CS problems, it is possible to do better: Nesterov in 1983 introduced a first-order scheme whose convergence rate that matches the theoretical limit for the class of Lipschitz continuous and strongly convex functions [5, 6, 10]. Standing on the shoulders of the same framework, we define the FLIHT recursion ($a_{i+1} = 0.5 \left(1 + \sqrt{1 + 4a_i^2} \right)$, $a_1 = 1$, and $\theta_i = \frac{a_i - 1}{a_{i+1}}$):

$$\begin{aligned}
x_{i+1} &= E \times \mathcal{M}_K \left(D \left(y_i - \frac{1}{L_{3K}} \nabla f(y_i) \right) \right), \\
y_i &= x_i + \theta_i (x_i - x_{i-1}).
\end{aligned} \quad (13)$$

Note the step size within hard-thresholding is L_{3K} since FLIHT generates a sequence of $y_i \in \Sigma_{2K}$ while updating $x_i \in \Sigma_K$.

FLIHT has the following approximation guarantee, which is proved in the appendix:

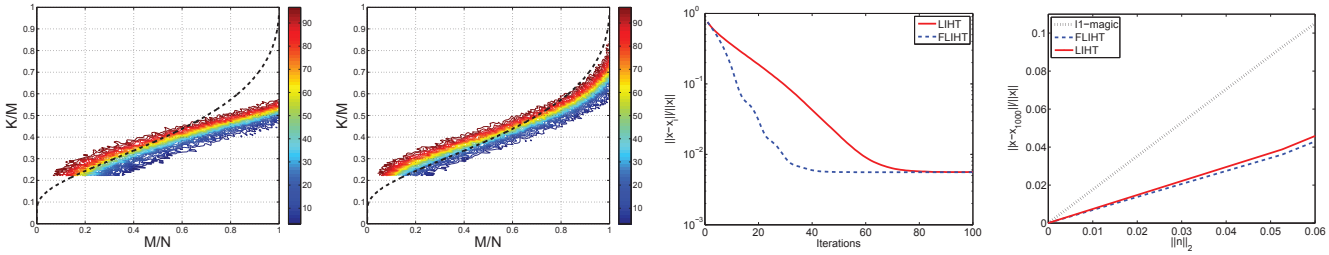


Fig. 2. Phase transition curves LIHT (first plot), FLIHT (second plot) are compared to Donoho-Tanner bound (dashed); Corresponding failure percentages are shown. Convergence rate of FLIHT is better than LIHT (third plot). Both algorithms exhibit better noise robustness (fourth plot) as compared to l1-magic due to the knowledge of K.

Theorem 2. *Define* $\rho = \sqrt{\frac{L_{3K}-\mu_{3K}}{\mu_{3K}}}$ *and* $\rho_\pm = \rho \pm \sqrt{\rho^2+\rho}$. *Let* $L_{3K} < 10/9\mu_{3K}$ *so that* $|\rho_\pm| < 1$. *Then, the iterations of FLIHT are contractive, and the* i*-th iteration satisfies*

$$\|x^* - x_i\| \leq \mathcal{O}(\rho_+^i) + \mathcal{O}(\rho_-^i) + \frac{2\sqrt{2}(1-3\rho)^{-1}}{\sqrt{\mu_{2K}}}\|n\|. \quad (14)$$

While FLIHT is rather restrictive on the SE constants, as compared to LIHT, its empirical performance does not seem to be affected from this restriction, as the next section illustrates.

4. EXPERIMENTS

Phase Transitions. Donoho and Tanner's combinatorial geometry based theory precisely quantifies the fundamental ℓ_1-sparsity and compression trade-off (K vs. M) [11]. The theory predicts the exact location in sparsity-undersampling domain where linear programming (ℓ_1-norm minimization) exhibits a phase transition in its performance. To see how ALPS compare to the ℓ_1 theoretical phase transition, we perform Monte Carlo simulations. We fix the signal dimension to $N = 1000$ and sweep across K and M values (120 and 200 sample points, respectively). For each (K, M)-pair, we repeat the following 100-times: (*i*) generate a random sparse vector with unit norm, (*ii*) generate compressive measurements (no noise) using iid Gaussian matrices, and (*iii*) recover the signals using ALPS. ALPS use the same number of iterations 1000. We then report the number of recoveries that obtain this accuracy or better.

Figure 2 summarizes the results, which are quite promising for ALPS. LIHT and FLIHT manages to follow the phase transition closely. We believe both algorithms require more iterations the stay on the theoretical curve as M increases.

Empirical convergence and stability performance. We test the empirical convergence rates and the stability of ALPS with the following set up. For $N = 1000$, we fix $K = 100$ and choose $M = 5K$ so that ALPS can recover 100-sparse vectors with high probability using 100 iterations according to our theory. During the convergence rate test, we add noise with $\|n\| = 10^{-2}$. For the stability test, we add iid Gaussian noise to the measurements with increasing variance. We repeat both experiments 100 times (c.f., Figure 2), and report averages. LIHT and FLIHT exhibit linear convergence first but slow down near the noise level; FLIHT clearly exhibits faster convergence. ALPS also show similar noise robustness, which is empirically better than l1-magic; this is probably because they exploit the correct signal sparsity during recovery.

Signal recovery using over-complete dictionaries. To demonstrate the great promise of the analysis sparsity model with ALPS, we recover a modulated narrow-band signal $N = 8192$ from $M = 80$ compressive measurements (c.f., Fig. 3), using FLIHT.[1] This is an unreasonably small amount of data corresponding to an under-sampling factor exceeding 100. The signal is approximately $K = 1000$ sparse in an overcomplete Gabor dictionary. FLIHT algorithm (initialized with 0) converges to a close approximation of the signal within 100 iterations, whereas LIHT exhibits slightly worse performance (not shown). l1-magic, using discrete cosine transform as sparsity basis, on the other hand (initialized with the true vector) cannot recover the signal and needs $M = 400$ for comparable performance.

5. CONCLUSIONS

Algebraic pursuits (ALPS) create a unifying connection between combinatorial optimization and the first-order optimization framework by Nesterov. The key enabler is the stable embedding assumption. We studied ALPS within analysis and synthesis sparsity models, and characterized their estimation guarantees. A weakness of ALPS is that they require a restricted-Lipschitz constant to use as a step-size. It is possible to overcome this difficulty and incorporate a step-size selection procedure into ALPS [13]. Finally, our proofs show that the algorithms can recover model-sparse signals [1] by replacing the hard-thresholding operation with the combinatorial model-projection.

Appendix: Technical derivations

Lemma 2. We first note that $\bar{x}_1$ minimizes the upper-bound U. Therefore, its objective value satisfies

$$f(\bar{x}_1) \leq \min_{x:x\in\Sigma_K} U(x, x_1). \quad (15)$$

Therefore, we also have $f(\bar{x}_1) \leq U(x^*, x_1)$ since $x^* \in \Sigma_K$. We then invoke Lemma 1(4):

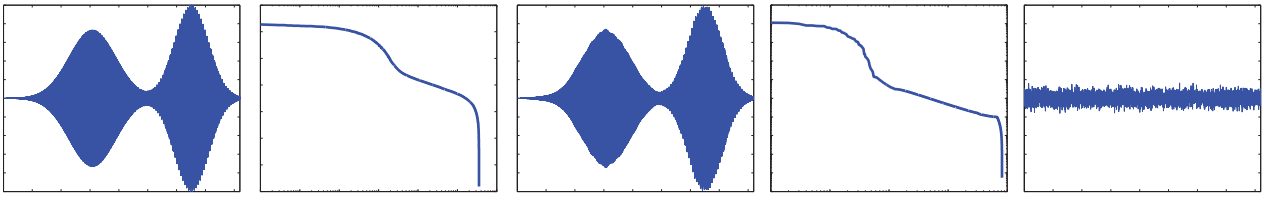
$$\langle \nabla f(x_1), x^* - x_1\rangle \leq f(x^*) - f(x_1) - \frac{\mu_{K'}}{2}\|x^* - x_1\|^2. \quad (16)$$

Using the definition of (6) and (16), we obtain

$$f(\bar{x}_1) \leq f(x^*) + \frac{L_{K'} - \mu_{K'}}{2}\|x^* - x_1\|^2. \quad (17)$$

Note that if $a_1^2 \leq a_2^2 + a_3^2$ for some positive a_j, then we have $a_1 \leq a_2 + a_3$. Using this fact, we can reach (9) by also noting that $f(x^*) \leq \|n\|^2$, as defined in (2).

[1] VC thanks Michael B. Wakin for providing the code for the signal generation and the over-complete Gabor dictionary.



(a) Target (b) Sorted Gabor coeff. (c) Analysis recon. (d) Sorted DCT coeff. (c) Synthesis recon.

Fig. 3. *(a) A modulated, narrow-band test signal is shown (N=8192). (b) Gabor analysis coefficients (N'=43 $\times$ N) of the signal are compressible. (c) FLIHT has $\|x - x_i\| = 0.0584$ error in CS reconstruction with the Gabor dictionary using 80 measurements (Φ is Bernoulli as in [12]). (d) Discrete cosine transform (DCT) coefficients (synthesis) of the signal can be well-approximated as K=50-sparse. (e) CS recovery (l1-magic) using DCT cannot recover the signal at the same measurement rate, and needs $M = 400$ measurements for comparable performance. FLIHT takes less than a few seconds on a laptop since the Gabor dictionary has an efficient implementation.*

To obtain (10), we use the SE and the triangle inequality on the left hand side of (9):

$$\|x^* - \bar{x}_1\| \leq \frac{\|\Phi x^* - \Phi\bar{x}_1 + n - n\|}{\sqrt{\mu_{K'}/2}} \leq \frac{\|u - \Phi\bar{x}_1\| + \|n\|}{\sqrt{\mu_{K'}/2}}, \tag{18}$$

which provides the desired result. □

Theorem 1. The result in Lemma 2 provides the iteration invariant of the LIHT algorithm. A straightforward induction leads to the main statement of Theorem 1. □

Theorem 2. We invoke Lemma 2 to first relate x_{i+1} and y_i:

$$\|x^* - x_{i+1}\| \leq \rho\|x^* - y_i\| + \frac{2\|n\|}{\sqrt{\mu_{2K}/2}}. \tag{19}$$

There are two changes in the definition of ρ, as compared to Theorem 1. The first change is in the numerator, since we have $K' = 3K$ as y_i is $2K$-sparse in general (c.f., Lemma 2). The second change is in the denominator where we use $\mu_{3K} \leq \mu_{2K}$ in order to obtain the condition on ρ so that the iterations are contractive.

Starting from $y_i = x_i + \theta_i(x_i - x_{i-1})$, we first subtract x^* from both sides and use the triangle inequality to obtain:

$$\begin{aligned} \|x^* - y_i\| &\leq (1 + \theta_i)\|x^* - x_i\| + \theta_i\|x^* - x_{i-1}\|, \\ &\leq 2\|x^* - x_i\| + \|x^* - x_{i-1}\|, \end{aligned} \tag{20}$$

where we use the fact that $\theta_i \leq 1$. Substituting (20) into (19), we stumble upon the iteration invariant of FLIHT:

$$\|x^* - x_{i+1}\| \leq 2\rho\|x^* - x_i\| + \rho\|x^* - x_{i-1}\| + \frac{2\|n\|}{\sqrt{\mu_{2K}/2}}. \tag{21}$$

We focus on the roots of the characteristics function of the difference equation, which can be found by solving $r^2 - 2\rho r - \rho = 0$. The roots of this characteristics function is given by $\rho_\pm = \rho \pm \sqrt{\rho^2 + \rho}$. In order for the difference equation to be contractive, we need to have $|\rho_\pm| < 1$, which leads to $\rho < 1/3$ and the condition we have on the SE constants in the Theorem.

To obtain the final upper-bound, we can assume a worst case convergence of the algorithm at some iteration j, where we have

$$\|x^* - x_j\| \leq 2\rho\|x^* - x_j\| + \rho\|x^* - x_j\| + \frac{2\|n\|}{\sqrt{\mu_{2K}/2}}, \tag{22}$$

which leads to $(1 - 3\rho)\|x^* - x_j\| \leq \frac{2\|n\|}{\sqrt{\mu_{2K}/2}}$, and the final result.

Note that as there are only a finitely many number of supports that the algorithm can visit, the algorithm is eventually periodic. As the iterations are contractive, this may in fact lead to convergence [14]. □

6. REFERENCES

[1] R. G. Baraniuk, V. Cevher, and M. B. Wakin, "Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective," *Proc. of the IEEE*, 2010.

[2] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Transactions on Information Theory*, 2008.

[3] E. J. Candes, Y. C. Eldar, and D. Needell, "Compressed sensing with coherent and redundant dictionaries," 2010.

[4] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, 2009.

[5] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical Programming*, vol. 103, 2005.

[6] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, Springer, 2004.

[7] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Info. Theory*, vol. 52, 2006.

[8] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, 2009.

[9] R. Garg and R. Khandekar, "Gradient Descent with Sparsification: An iterative algorithm for sparse recovery with restricted isometry property," in *ICML*, 2009.

[10] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$," in *Soviet Mathematics Doklady*, 1983, vol. 27.

[11] D. L. Donoho and J. Tanner, "Precise undersampling theorems," *Proc. of the IEEE*, vol. 98(6), 2010.

[12] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, 2008.

[13] Volkan Cevher, "On Accelerated Hard Thresholding Methods for Sparse Approximation," Tech. Rep., 2011.

[14] S. Foucart, "Hard thresholding pursuit: An algorithm for compressive sensing," preprint, 2010.