COMPRESSIVE SENSING MEETS GAME THEORY

Sina Jafarpour, Robert E. Schapire

Princeton University
Department of Computer Science
$\{sina, schapire\}@cs.princeton.edu$

Volkan Cevher

Ecole Polytechnique Federale de Lausanne
Idiap Research Institute
$volkan.cevher@epfl.ch$

ABSTRACT

We introduce the Multiplicative Update Selector and Estimator (MUSE) algorithm for sparse approximation in underdetermined linear regression problems. Given $\boldsymbol{f} = \Phi\boldsymbol{\alpha}^* + \boldsymbol{\mu}$, the MUSE provably and efficiently finds a k-sparse vector $\hat{\boldsymbol{\alpha}}$ such that $\|\Phi\hat{\boldsymbol{\alpha}} - f\|_\infty \leq \|\boldsymbol{\mu}\|_\infty + O\left(\frac{1}{\sqrt{k}}\right)$, for *any* k-sparse vector $\boldsymbol{\alpha}^*$, *any* measurement matrix Φ, and *any* noise vector μ. We cast the sparse approximation problem as a zero-sum game over a properly chosen new space; this reformulation provides salient computational advantages in recovery. When the measurement matrix Φ provides stable embedding to sparse vectors (the so-called restricted isometry property in compressive sensing), the MUSE also features guarantees on $\|\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}\|_2$. Simulation results demonstrate the scalability and performance of the MUSE in solving sparse approximation problems based on the Dantzig Selector.

***Index Terms*—** Compressed Sensing, Game Theory, Dantzig Selector, Multiplicative Weights Algorithm.

1. INTRODUCTION

Sparse approximation is a fundamental problem in many signal processing applications; examples include compressive sensing, probabilistic estimation, and model selection [1, 2, 3]. By sparse approximation, we mean the following: given a matrix $\Phi \in \mathbb{R}^{m\times n}$ $(M < N)$, a vector $\boldsymbol{f} \in \mathbb{R}^m$, find a vector $\hat{\boldsymbol{\alpha}}$ satisfying $\Phi\hat{\boldsymbol{\alpha}} \approx \boldsymbol{f}$, whenever it exists, such that $\hat{\boldsymbol{\alpha}}$ has at most $k \ll n$-nonzero entries.

In this paper, we focus on the sparse approximation problems, where $\Phi\hat{\boldsymbol{\alpha}} \approx \boldsymbol{f}$ is quantified in the ℓ_∞ norm as $\|\Phi\hat{\boldsymbol{\alpha}} - \boldsymbol{f}\|_\infty$. We prove that for every k-sparse $\boldsymbol{\alpha}^*$ and noise vector $\boldsymbol{\mu}$ that satisfy $\boldsymbol{f} = \Phi\boldsymbol{\alpha}^* + \boldsymbol{\mu}$, one can efficiently find a k-sparse vector $\hat{\boldsymbol{\alpha}}$ with $\|\Phi(\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}})\|_\infty \leq \|\boldsymbol{\mu}\|_\infty + O\left(\frac{1}{\sqrt{k}}\right)$. This guarantee is especially strong in high-dimensional settings of the problem, where k/n tends to a constant. To the best of our knowledge, this is the first ℓ_∞-based sparse approximation framework that provably works for *every* k-sparse $\boldsymbol{\alpha}^*$, *every* matrix Φ as well as *every* noise vector $\boldsymbol{\mu}$. Our algorithm to find the promised $\hat{\boldsymbol{\alpha}}$ with the desiderata is dubbed the Multiplicative Update Selector and Estimator (MUSE).

To demonstrate our approach, we study the Dantzig Selector (DS) problem [4] in compressive sensing (CS). The DS exploits ℓ_1-norm minimization to find sparse solutions $\hat{\boldsymbol{\alpha}}$ subject to the constraint of $\|\Phi^\top(\Phi\hat{\boldsymbol{\alpha}} - \boldsymbol{f})\|_\infty \leq \epsilon$. To obtain the DS solution, one can leverage linear programming, which has $O(m^2n^{1.5})$ computational complexity using the interior point method. In sharp contrast, we show that if the sensing matrix satisfies the restricted isometry property, then the MUSE algorithm can approximate the Dantzig Selector solution efficiently in $O(k\mathcal{M})$. While $\mathcal{M}$ is $O(mn)$ in general, it can be reduced to $O(n\log n)$ for many structured matrices, e.g., partial Fourier ensembles via the fast Fourier transform.

In our game-theoretic reformulation of the DS, we assume the problem is normalized so that $\|\boldsymbol{\alpha}^*\|_1 \leq 1$. This allows us to view the DS problem as a matrix-game. Instead of smoothing the matrix-game objective uniformly in the spirit of Nesterov's gradient approaches [5], we approximate it by a modular objective, which features salient computational advantages. For instance, the most costly operation per iteration of our algorithm is the sole application of $\Phi^\top$ (Φ is used only once). We establish the theoretical convergence rate of the algorithm: $O\big(1/\epsilon^2\big)$ iterations are needed to obtain an ϵ-approximation error. Nevertheless, the algorithm empirically exhibits $O(1/\epsilon)$ convergence, matching the best known rates based on smoothing that can be obtained by computationally competitive first order methods [5].

2. PRELIMINARIES

For every integer n, we denote $[n] \doteq \{1, \cdots, n\}$. Throughout this paper, we let k be an integer smaller than n. For each $i \in [n]$, let $\boldsymbol{e_i}$ denote the i-th canonical vector with one at its i-th entry, and zero everywhere else.

For each $\epsilon \in (0, 1)$, an $m \times n$ matrix Φ satisfies the (k, ϵ) Restricted Isometry Property, referred to as (k, ϵ)-RIP, if the following is satisfied for every k-sparse vector $\boldsymbol{x}$:

$$(1-\epsilon)\|\boldsymbol{x}\|^2 \leq \|\Phi\boldsymbol{x}\|^2 \leq (1+\epsilon)\|\boldsymbol{x}\|^2.$$

The simplex Δ^n is defined as the set of vectors in $\mathbb{R}^n$ with

This work was supported in part by the European Commission under Grant MIRG-268398 and DARPA KeCoM program #11-DARPA-1055. VC also would like to acknowledge Rice University for his Faculty Fellowship.

positive entries and unit ℓ_1 norm, and Δ_k^n represents all vectors in Δ^n which are also k-sparse.

The ℓ_∞ norm of an $m \times n$ matrix Φ is defined as

$$\|\Phi\|_\infty \doteq \max_{i\in[n]} \max_{j\in[m]} |\Phi_{ij}|.$$

3. SPARSE APPROXIMATION IN THE ℓ_∞-NORM

Let Φ be an $m \times n$ matrix, $\boldsymbol{\alpha}^*$ be a k-sparse vector in $\mathbb{R}^n$, and $\boldsymbol{\mu}$ be any vector in $\mathbb{R}^m$. Denote $\boldsymbol{f} \doteq \Phi\boldsymbol{\alpha}^* + \boldsymbol{\mu}$. Sparse approximation in the ℓ_∞-norm is then the task of finding the optimal solution of the problem

$$\min_{\boldsymbol{\alpha}:k-\text{sparse}} \|\Phi\boldsymbol{\alpha} - \boldsymbol{f}\|_\infty. \tag{1}$$

To solve the sparse approximation problem, we first reformulate the problem as a min-max game, and then adopt a multiplicative update algorithm to *approximately* estimate the game solution.

First we show that without loss of generality we can assume that $\boldsymbol{\alpha}^*$ is a sparse vector in Δ^{2n}. It is shown by Berinde et al. [6] that by incorporating $O(k\log n)$ extra linear measurements using hash functions, one can always estimate an upper-bound for $\|\boldsymbol{\alpha}^*\|_1$.[1] As a result, by dividing the measurement vector $\boldsymbol{f}$ by the provided upper-bound, we can always assume that $\|\boldsymbol{\alpha}^*\|_1 \leq 1$.

To convert the domain of the sparse approximation problem onto the positive simplex, we let $\Psi \doteq [\Phi, -\Phi]$, and also let $\boldsymbol{x}^* \in \mathbb{R}^{2n}$ be a vector whose entries are given by

$$\mathbf{x}_i^* = \begin{cases} \alpha_i^* & \text{if } \alpha_i^* > 0 \text{ and } i \leq n \\ -\alpha_i^* & \text{if } \alpha_i^* < 0 \text{ and } i > n \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

With these transformations, it is clear that every linear combination of the columns of Φ can be represented as a *positive* linear combination of the columns of Ψ. Therefore, if $\boldsymbol{f} = \Phi\boldsymbol{\alpha}^* + \boldsymbol{\mu}$, then $\boldsymbol{f} = \Psi\boldsymbol{x}^* + \boldsymbol{\mu}$ and vice versa. We then define

$$\boldsymbol{A} \doteq \begin{bmatrix} \Psi \\ -\Psi \end{bmatrix} = \begin{bmatrix} \Phi & -\Phi \\ -\Phi & \Phi \end{bmatrix}, \text{ and } \boldsymbol{y} \doteq \begin{bmatrix} \boldsymbol{f} \\ -\boldsymbol{f} \end{bmatrix}.$$

Hence, we can rewrite the sparse approximation problem as

$$\|\Psi\boldsymbol{x} - \boldsymbol{f}\|_\infty = \max_{j\in[m]} |(\Psi\boldsymbol{x} - \boldsymbol{f})_j| =$$
$$\max_{j\in[2m]} \boldsymbol{e_j}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) = \max_{\mathbf{P}\in\Delta^{2m}} \mathbf{P}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}).$$

The last equality follows from the fact that the maximum of a linear program occurs at a boundary point of the simplex Δ^{2m}. In the rest of this paper, for every $\mathbf{P} \in \Delta^{2m}$, and every $\boldsymbol{x} \in \Delta^{2n}$, we define

$$\mathcal{L}(\mathbf{P}, \boldsymbol{x}) \doteq \mathbf{P}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}), \tag{3}$$

[1] This upper bound is at most $2\|\boldsymbol{\alpha}^*\|_1$. We emphasize that the our results become more accurate if an even tighter upper-bound for $\|\boldsymbol{\alpha}^*\|_1$ is known *a priori*.

Algorithm 1 The Multiplicative Update Selector and Estimator (MUSE) Algorithm

Inputs: $\boldsymbol{y}$, $\boldsymbol{A}$, and parameters T, and $\eta > 0$.
Output: A T-sparse approximation $\hat{\boldsymbol{x}}$ for the vector $\boldsymbol{x}^*$.

Set $\mathbf{P}^1 = \frac{1}{2m}[\mathbf{1}]_{1\times 2m}$.
for $t = 1, \cdots, T$ **do**
  Find $\boldsymbol{x}^t \doteq \arg_{\boldsymbol{x}} \min \mathcal{L}(\mathbf{P}^t, \boldsymbol{x})$.
  For each $i \in [2m]$, update $\mathrm{P}_i^{t+1} = \mathrm{P}_i^t \, e^{\frac{\eta\mathcal{L}(e_i, \boldsymbol{x}^t)}{2\mathcal{L}_{\max}}}$.
  Let $Z^{t+1} = \sum_{i=1}^m \mathrm{P}_i^t e^{\frac{\eta\mathcal{L}(e_i, \boldsymbol{x}^t)}{2\mathcal{L}_{\max}}}$.
  For each $i \in [2m]$, let $\mathrm{P}_i^{t+1} = \frac{\mathrm{P}_i^{t+1}}{Z^{t+1}}$.
end for
Output $\hat{\boldsymbol{x}} \doteq \frac{1}{T}\sum_{t=1}^T \boldsymbol{x}^t$.

and $\mathcal{L}_{\max} \doteq \max_{\mathbf{P},\boldsymbol{x}} |\mathcal{L}(\mathbf{P}, \boldsymbol{x})| = \|\Phi\|_\infty + \|\boldsymbol{f}\|_\infty$.

Consequently, the sparse approximation problem in the ℓ_∞ norm is equivalent to the problem of finding the min-max optimal solution of $\mathcal{L}$:

$$\min_{\substack{\|\boldsymbol{\alpha}\|_1\leq 1 \\ \boldsymbol{\alpha}:k-\text{sparse}}} \|\Phi\boldsymbol{\alpha} - \boldsymbol{f}\|_\infty = \min_{\substack{\boldsymbol{x}\in\Delta^{2n} \\ \boldsymbol{x}:k-\text{sparse}}} \max_{\mathbf{P}\in\Delta^{2m}} \mathbf{P}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}).$$

Unfortunately, since we are restricted to k-sparse vectors, the search space is non-convex, and therefore finding this game-solution is intractable. Nevertheless, in Section 4 we introduce the MUSE Algorithm, which provides a *sparse approximation* to the min-max optimal solution.

4. THE MUSE ALGORITHM

The Multiplicative Update Selector and Estimator (MUSE) is a repurposing of the Multiplicative Weights Algorithm (MWA) [7]. MWA, as proposed by Freund and Schapire for learning to play repeated games, relies on Littlestone and Warmuth's Weighted Majority Algorithm [8]. A pseudo-code of the MUSE is given in **Algorithm 1**.

We show that running the MUSE for $T = k$ iteration is sufficient to obtain a k-sparse approximation to $\hat{\boldsymbol{x}}$. We first define the following bilinear function with range $[0, 1]$:

$$\mathcal{L}'(\mathbf{P}, \boldsymbol{x}) \doteq \frac{1}{2} - \frac{\mathcal{L}(\mathbf{P}, \boldsymbol{x})}{2\mathcal{L}_{\max}}. \tag{4}$$

The following lemma is a consequence of Theorem 1 in [7], and bounds the regret loss of the Multiplicative Weights strategy in zero-sum games:

Lemma 4.1. *Let T be any positive integer, and define $\eta = \ln\left(1 + \sqrt{\frac{2\ln(2m)}{T}}\right)$. Suppose $\langle(\boldsymbol{P}^1, \boldsymbol{x}^1), \cdots, (\boldsymbol{P}^T, \boldsymbol{x}^T)\rangle$ is the sequence of pairs generated by the MUSE Algorithm after T iterations. Then $\frac{1}{T}\sum_{t=1}^T \mathcal{L}'(\boldsymbol{P}^t, \boldsymbol{x}^t)$ is at most*

$$\frac{1}{T}\min_{\boldsymbol{P}\in\Delta^{2m}} \sum_{t=1}^T \mathcal{L}'(\boldsymbol{P}, \boldsymbol{x}^t) + (1+\sqrt{2})\sqrt{\frac{\ln 2m}{T}}.$$

To highlight the impact of Lemma 4.1, we note that at each iteration t, the solution of $\boldsymbol{x}^t \doteq \arg_{\boldsymbol{x}} \min \mathcal{L}(\mathbf{P}^t, \boldsymbol{x})$ is not necessarily unique; however, the bound of Lemma 4.1 is valid for every such solution. On the other hand, any such solution can be represented as a linear combination of pure (1-sparse) solutions which also minimize $\mathcal{L}(\mathbf{P}^t, \boldsymbol{x})$. Observe that each minimizer $\boldsymbol{x}^t$ is also a minimizer of $\left(\boldsymbol{A}^\top \mathbf{P}^t\right)^\top \boldsymbol{x}^t$. Therefore, at each iteration, we can enforce the algorithm to output a 1-sparse solution, corresponding to the index of the minimum entry of $\boldsymbol{A}^\top \mathbf{P}^t$. As a result, the vector $\hat{\boldsymbol{x}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}^t$ is at most T-sparse.

To transform this to an estimate $\hat{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}^* \in \mathbb{R}^n$, we recall that the first n elements of $\boldsymbol{x}^*$ correspond to the positive entries of $\boldsymbol{\alpha}^*$, and the second n elements of $\boldsymbol{x}^*$ correspond to the negative entries of $\boldsymbol{\alpha}^*$ (Equation (2)). Therefore, the vector $\hat{\boldsymbol{\alpha}}$ can be estimated from $\boldsymbol{\alpha}^*$ by setting

$$\hat{\boldsymbol{\alpha}}_i = \hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_{i+n} \text{ for every } i \in [n]. \tag{5}$$

Here, we use Lemma 4.1 to show that the MUSE Algorithm after T iterations finds a T-sparse vector $\hat{\boldsymbol{\alpha}}$ with bounded ℓ_∞ loss in the measurement domain.

Theorem 4.2. *Let δ be any number in $(0, 1]$, and let $\hat{\boldsymbol{x}}$ be the output of the MUSE Algorithm after $T = \frac{k}{\delta}$ iterations. Let $\hat{\boldsymbol{\alpha}}$ be as in Equation (5). Then $\hat{\boldsymbol{\alpha}}$ is a $\frac{k}{\delta}$-sparse vector with*

$$\|\Phi\hat{\boldsymbol{\alpha}} - \boldsymbol{f}\|_\infty \le \|\boldsymbol{\mu}\|_\infty + (1+\sqrt{2})\big(2\|\Phi\|_\infty + \|\boldsymbol{\mu}\|_\infty\big)\sqrt{\frac{\delta \ln(2m)}{k}}. \tag{6}$$

Proof. Observe that

$$\begin{aligned}
\min_{\boldsymbol{x}} \max_{\mathbf{P}} \mathcal{L}(\mathbf{P}, \boldsymbol{x}) &=^a \max_{\mathbf{P}} \min_{\boldsymbol{x}} \mathcal{L}(\mathbf{P}, \boldsymbol{x}) \ge^b \min_{\boldsymbol{x}} \mathcal{L}\left(\hat{\mathbf{P}}, \boldsymbol{x}\right) \\
&\ge^c \frac{1}{T}\sum_{t=1}^{T} \min_{\boldsymbol{x}} \mathcal{L}(\mathbf{P}^t, \boldsymbol{x}) =^d \frac{1}{T}\sum_{t=1}^{T} \mathcal{L}(\mathbf{P}^t, \boldsymbol{x}^t) \\
&\ge^e \max_{\mathbf{P}} \mathcal{L}\left(\mathbf{P}, \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}^t\right) - (1+\sqrt{2})\,\mathcal{L}_{\max}\sqrt{\frac{\delta \ln 2m}{k}}.
\end{aligned} \tag{7}$$

Equality (a) is the min-max theorem. Inequality (b) follows from the definition of max. Inequality (c) is a consequence of the linearity of $\mathcal{L}$ and concavity of min. Equality (d) is valid by the definition of $\boldsymbol{x}^t$, and Inequality (e) follows from Lemma 4.1 and linearity of $\mathcal{L}'$. As a result,

$$\max_{\mathbf{P}} \mathcal{L}(\mathbf{P}, \hat{\boldsymbol{x}}) \le \min_{\boldsymbol{x}} \max_{\mathbf{P}} \mathcal{L}(\mathbf{P}, \boldsymbol{x}) + (1+\sqrt{2})\,\mathcal{L}_{\max}\sqrt{\frac{\delta \ln 2m}{k}}. \tag{8}$$

Next, we use the triangle inequality to bound $\mathcal{L}_{\max}$:

$$\|\boldsymbol{f}\|_\infty \le \|\Phi\boldsymbol{\alpha}^*\|_\infty + \|\boldsymbol{\mu}\|_\infty \le \|\boldsymbol{\alpha}^*\|_1\|\Phi\|_\infty + \|\boldsymbol{\mu}\|_\infty.$$

Finally, it follows from the definition of $\boldsymbol{A}$, $\boldsymbol{y}$, and $\mathcal{L}$ that $\max_{\mathbf{P}} \mathcal{L}(\mathbf{P}, \hat{\boldsymbol{x}}) = \|\Phi\hat{\boldsymbol{\alpha}} - \boldsymbol{f}\|_\infty$, and

$$\min_{\boldsymbol{x}} \max_{\mathbf{P}} \mathcal{L}(\mathbf{P}, \boldsymbol{x}) = \min_{\boldsymbol{\alpha}: \|\alpha\|_1 \le 1} \|\Phi\boldsymbol{\alpha} - \boldsymbol{f}\|_\infty \le \|\boldsymbol{\mu}\|_\infty.$$

□

5. CONNECTIONS TO DANTZIG SELECTOR

In this section, we show that under standard compressed sensing assumptions, one can also obtain sparse approximation guarantees in the so-called signal domain. Throughout this section let $\kappa \doteq \sqrt{\frac{2\log(2n^2)}{m}}$, and let $\boldsymbol{B}$ be an $m \times n$ iid $\left\{\frac{-1}{\sqrt{m}}, \frac{1}{\sqrt{m}}\right\}$ Bernoulli matrix.[2] It has been shown by Candès et.al. [1], that as long as $m = \Omega\big(k \log\left(\frac{n}{k}\right)\big)$, with overwhelming probability $\boldsymbol{B}$ satisfies the $(k, 0.5)$-RIP.

Now let $\boldsymbol{\alpha}^*$ be as before, and let $\boldsymbol{b} = \boldsymbol{B}\boldsymbol{\alpha}^* + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is a vector in $\mathbb{R}^m$. Let $\Phi = \boldsymbol{B}^\top\boldsymbol{B}$, $\boldsymbol{\mu} = \boldsymbol{B}^\top\boldsymbol{\varepsilon}$, and $\boldsymbol{f} = \boldsymbol{B}^\top\boldsymbol{b} = \Phi\boldsymbol{\alpha}^* + \boldsymbol{\mu}$.

Since $\boldsymbol{B}$ satisfies the RIP, finding the exact solution of Equation (1) leads to a k-sparse vector close to $\boldsymbol{\alpha}^*$. The Dantzig Selector [4] approximates $\boldsymbol{\alpha}^*$ by finding the *exact* solution of a *relaxed* convex program. In contrast, we *approximate* the solution of (1) using the MUSE Algorithm. We show that with overwhelming probability, the solution $\hat{\boldsymbol{\alpha}}$ of the MUSE Algorithm is close to $\boldsymbol{\alpha}^*$.

Theorem 5.1. *Let δ be any number in $(0, 1]$, and assume that the Bernoulli sensing matrix $\boldsymbol{B}$ is $\left(\left(\frac{1}{\delta}+1\right)k, 0.5\right)$-RIP. Let $\hat{\boldsymbol{\alpha}}$ be the output of the MUSE Algorithm with inputs $\boldsymbol{B}^\top\boldsymbol{B}$, $\boldsymbol{B}^\top\boldsymbol{b}$, $T = \frac{k}{\delta}$, and η of Lemma 4.1. Then with probability $1 - n^{-1}$, $\hat{\boldsymbol{\alpha}}$ is a $\frac{k}{\delta}$-sparse vector with*

$$\|\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}\|_2^2 \le \left(8 + 10\sqrt{\frac{\delta\ln(2n)}{k}}\right)\kappa\|\boldsymbol{\varepsilon}\|_2 + 20\sqrt{\frac{\delta\ln(2n)}{k}}. \tag{9}$$

Proof. Since every column of $\boldsymbol{B}$ has unit ℓ_2 norm, $\|\boldsymbol{B}^\top\boldsymbol{B}\|_\infty \le 1$. Moreover, by applying Hoeffding's inequality to every fixed column of $\boldsymbol{B}$, and then taking the union bound over all n columns (see also [9]) we can show that with probability at least $1 - n^{-1}$, $\|\boldsymbol{B}^\top\boldsymbol{\varepsilon}\|_\infty \le \kappa\|\boldsymbol{\varepsilon}\|_2$.

Therefore, it follows from Theorem 4.2, and the triangle inequality[3] that $\hat{\boldsymbol{\alpha}}$ is $\frac{k}{\delta}$-sparse, and $\|\boldsymbol{B}^\top\boldsymbol{B}(\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}})\|_\infty$ is upper-bounded by

$$\left(2 + 2.5\sqrt{\frac{\delta\ln(2n)}{k}}\right)\kappa\|\boldsymbol{\varepsilon}\|_2 + 5\sqrt{\frac{\delta\ln(2n)}{k}}.$$

We also have

$$\begin{aligned}
\|\boldsymbol{B}^\top(\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}})\|_2^2 &\le \|(\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}})\|_1\|\boldsymbol{B}^\top\boldsymbol{B}(\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}})\|_\infty \\
&\le 2\|\boldsymbol{B}^\top\boldsymbol{B}(\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}})\|_\infty.
\end{aligned}$$

[2] We only provide the results for Bernoulli matrices; however, the results are more general and can be applied to any dictionary satisfying the RIP.

[3] Here we approximated $1 + \sqrt{2}$ by 2.5.

The first inequality is Holder's inequality, and the second inequality follows from the fact that both $\|\boldsymbol{\alpha}^*\|_1 \leq 1$, and $\|\hat{\boldsymbol{\alpha}}\|_1 \leq 1$. Finally, observe that since $\boldsymbol{\alpha}^*$ is k-sparse, and $\hat{\boldsymbol{\alpha}}$ is $\frac{k}{\delta}$ sparse, $\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}$ is $k\left(\frac{1}{\delta}+1\right)$-sparse. The result then follows from the RIP property of $\boldsymbol{B}$:

$$\|\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}\|_2^2 \leq 2\|\boldsymbol{B}(\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}})\|_2^2 \leq 4\|\boldsymbol{B}^\top \boldsymbol{B}(\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}})\|_\infty. \quad (10)$$

□

6. EXPERIMENTAL RESULTS

In this section, we provide experimental results to demonstrate the performance of the MUSE Algorithm. We fixed $n = 1000$, $k = 150$ and $m = 500$, and repeated the following experiment 100 times.[4] We generated a sparse vector with random support, random sign, and unit ℓ_1 norm, generated compressive measurements in the presence of white noise, and then recovered the signals using the MUSE. The noise vector consists of m iid $\mathcal{N}(0, \sigma^2)$ elements, where σ ranges from 10^{-5} to 1.

Figure 1(a) plots the dependency between the measurement domain error $\|\boldsymbol{B}^\top \boldsymbol{B}\,\hat{\boldsymbol{\alpha}} - \boldsymbol{B}^\top \boldsymbol{b}\|_\infty$ and the number of iterations of the algorithm. Here we let the algorithm iterate for $10,000$ iterations using the value of η provided in Lemma 4.1. Figure 1(a) shows that the measurement domain loss consistently decreases as the algorithm continues iterating; moreover, the convergence value highly depends on σ, and the rate of convergence is approximately $\frac{1}{T}$ (as opposed to slower rate $\frac{1}{\sqrt{T}}$ expected from theory).

Figure 1(b) illustrates the signal-domain ℓ_2-error ($\|\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}\|_2/\|\boldsymbol{\alpha}^*\|_2$) of the algorithm. Interestingly, the data-domain error also consistently decreases as the algorithm iterates, even after $10,000$ iterations. Note that this does not mean the algorithm provides a dense estimate; on the contrary, the updates on the estimate tend to concentrate on the true signal support. For instance, the final solution for $\sigma = 10^{-5}$ case is approximately 184-sparse after $10,000$ iterations.

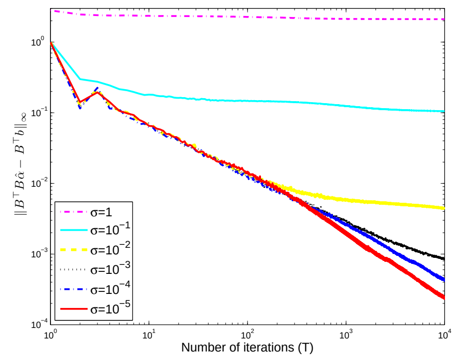
7. CONCLUSIONS

We proposed a scalable multiplicative-update algorithm to solve the sparse-approximation problem by reformulating the problem as a min-max game. We proved that the algorithm requires $O(1/\epsilon^2)$ iterations to obtain ϵ additive approximation error. However, the algorithm empirically needs $O(1/\epsilon)$ iterations. Future work will focus on closing the gap between the theoretical and empirical convergence rates, enforcing hard sparsity constraints, and on adapting the algorithm to the other convex relaxation problems.

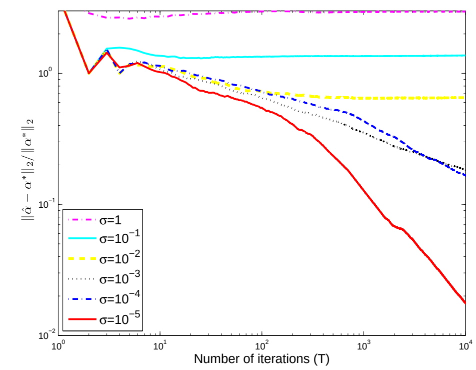
8. REFERENCES

[1] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[4] We observed consistent results for other values of n,k, and m.



(a) The measurement domain convergence rate.



(b) The data domain convergence rate.

Fig. 1. The dependency between the approximation error and the number of iterations of the MUSE Algorithm.

[2] M. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using l1-constrained quadratic programming (lasso)," *IEEE Trans. Inf. Theor.*, vol. 55, no. 5, pp. 2183–2202, 2009.

[3] 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.

[4] E. Candès and T. Tao, "The Dantzig Selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, 2007.

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

[6] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, "Combining geometry and combinatorics: a unified approach to sparse signal recovery," in *Proc. 46th Allerton Conf.*, 2008, pp. 798–805.

[7] Y. Freund and R. Schapire, "Adaptive game playing using multiplicative weights," *Games. Econ. Behav.*, vol. 29, pp. 79–103, 1999.

[8] N. Littlestone and M. Warmuth, "The weighted majority algorithm," in *Proc. 30th IEEE Symp. Found. Comput. Sci.*, 1989, pp. 256–261.

[9] W. Bajwa, R. Calderbank, and S. Jafarpour, "Why Gabor Frames? Two Fundamental Measures of Coherence and Their Role in Model Selection," *IEEE J. Commun. Networking*, vol. 12, no. 4, pp. 289–307, 2010.