

# HARD THRESHOLDING WITH NORM CONSTRAINTS

Anastasios Kyrillidis, Gilles Puy, and Volkan Cevher

École Polytechnique Fédérale de Lausanne

## ABSTRACT

We introduce a new sparse recovery paradigm, called **NORMED PURSUITS**, where efficient algorithms from combinatorial and convex optimization interface for interpretable and model-based solutions. Synthetic and real data experiments illustrate that **NORMED PURSUITS** can significantly enhance the performance of both hard thresholding methods and convex solvers in sparse recovery.

## 1. INTRODUCTION

In many applications, we are interested in recovering a high-dimensional signal  $x^* \in \mathbb{R}^n$  from an underdetermined set of linear observations  $y \in \mathbb{R}^m$  ( $m < n$ ), generated via a known  $\Phi \in \mathbb{R}^{m \times n}$ :

$$y = \Phi x^* + \varepsilon. \quad (1)$$

In (1),  $\varepsilon \in \mathbb{R}^m$  denotes an additive noise term with  $\|\varepsilon\|_2 \leq \sigma$ .

Signal sparsity in a known representation is a useful prior to circumvent ill-posed nature of (1). Therefore, assuming  $x^*$  is itself  $k$ -sparse (i.e., it has at most  $k$  non-zero coefficients), the following problem emerges as a natural estimator of  $x^*$ :

$$\operatorname{argmin}_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{subject to} \quad \|y - \Phi x\|_2 \leq \sigma. \quad (2)$$

Here,  $\|x\|_0$  denotes the  $\ell_0$ -“norm” that counts the nonzero elements in  $x$ . Unfortunately, (2) is NP-hard. In contrast, convex approaches relax (2) by replacing the  $\|\cdot\|_0$  with the *sparsity-inducing*  $\|\cdot\|_1$ -norm. For example, the Basis Pursuit (BP) approach in [1] converts (2) into the following:

$$\operatorname{argmin}_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{subject to} \quad \|y - \Phi x\|_2 \leq \sigma. \quad (3)$$

Hard thresholding algorithms [2–6] provide a different perspective to (2) by considering an  $\ell_0$ -constrained optimization problem:

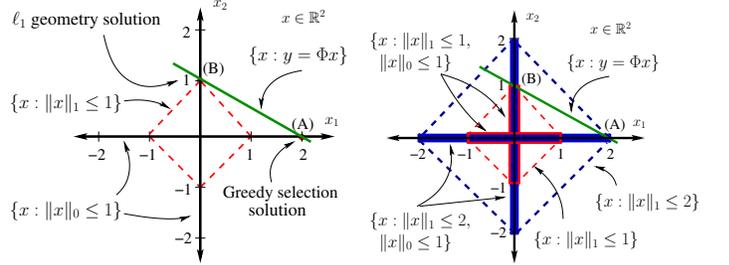
$$\operatorname{argmin}_{x \in \mathbb{R}^n} \|y - \Phi x\|_2^2 \quad \text{subject to} \quad \|x\|_0 \leq k, \quad (4)$$

where a putative  $k$ -sparse solution is iteratively refined using locally greedy decision rules. Similar to above, the Lasso algorithm [7] can be also thought as a relaxation of (4):

$$\operatorname{argmin}_{x \in \mathbb{R}^n} \|y - \Phi x\|_2^2 \quad \text{subject to} \quad \|x\|_1 \leq \lambda, \quad (5)$$

where  $\lambda > 0$  is a parameter that governs the sparsity of the solution.

While both  $\ell_0$  and  $\ell_1$  sparse recovery formulations above have similar theoretical guarantees, it is *incorrect* to view the convex  $\ell_1$ -norm as a convex relaxation of the  $\ell_0$ -set, which extends to infinity (c.f., Fig. 1). For instance,  $\ell_1$ -norm in (5) not only acts as a geometrical proxy to  $k$ -sparse signals, but also provides a *scale* that the hard thresholding methods in (4) cannot exploit. Unsurprisingly, the convex methods typically outperform the hard thresholding methods. Fig. 1(a) provides an illustrative toy example in  $\mathbb{R}^2$  ( $\sigma = 0$ ), which has two 1-sparse solutions (A) and (B). In this example, a hard thresholding method can greedily choose solution (A) on  $x_1$ -axis due to greedy selection satisfying  $y = \Phi x^*$ . In contrast, BP



(a) BP (B) and greedy (A) solutions. (b) Hard thresholding with norm constraints.

**Fig. 1.** Geometric interpretation of the selection process of (a) convex- and combinatorial-based methods and (b) the proposed framework for a simple test case  $y = \Phi x^*$  where  $\|x^*\|_0 = 1$ . The admissible set of greedy solutions with the norm constraint lie on the segments inside the boxes.

method would choose (B) as it has a smaller norm (or scale) than (A).

Interestingly, hard thresholding methods can exploit further combinatorial prior information on the support of the sparse signals to provably outperform the  $\ell_1$  methods in theory and in practice [8]. In this setting, a key ingredient is efficient combinatorial algorithms that can perform model-based projections. Hence, the extend of the model-based approach is not clearly understood. In parallel, a significant effort has gone into customizing convex norms that induces model-based sparsity in the literature [9–11].

**Contributions:** Currently, there are no attempts in directly incorporating norm constraints into the model-based recovery framework of [8]. To this end, we propose a unique optimization paradigm, dubbed **NORMED PURSUITS**, where both combinatorial (hard-thresholding) and norm constraints are active in sparse recovery. While combinatorial constraints enforce a union-of-subspaces model with sparsity parameter  $k$ , norm constraints further restrict the candidate set of concise solutions to have a given norm. Thus, we can capture both 1-sparse solutions in our toy example of Fig. 1 via, say, an  $\ell_1$ -norm constraint (i.e., (B) with any constraint  $\|x\|_1 \leq \lambda$ ,  $\lambda \in [1, 2)$ )—in case of solution ambiguity, combinatorial selection rules dictate the sparse solution.

**NORMED PURSUITS** accommodate a manifold of convex norm and combinatorial constraints that enhance the signal recovery performance; apart from the  $\ell_1$ -norm constraint, a non-exhaustive list of candidates include  $\ell_2$ ,  $\ell_\infty$  and total variation (TV) constraints.

We present theoretical approximation guarantees of **NORMED PURSUITS** and provide empirical evidence that they outperform the state-of-the-art approaches. As a special case of **NORMED PURSUITS**, we highlight the combinatorial selection and least absolute shrinkage (CLASH) operator [12], which uses  $\ell_1$ -norm constraints.

**Notation:**  $[x]_i$  denotes the  $i$ -th element of vector  $x$  and  $x_i$  represents the estimate of  $x$  in the  $i$ -th iteration of the algorithm. The index set of  $n$  dimensions is denoted as  $\mathcal{N} = \{1, 2, \dots, n\}$ . Given a set  $\mathcal{S} \in \mathcal{N}$  and a vector  $x \in \mathbb{R}^n$ ,  $x_{\mathcal{S}}$  represents a  $n$ -dimensional

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.

vector where  $[x_S]_S = [x]_S$  and  $[x_S]_{\mathcal{N} \setminus S} = 0$ . Given  $\mathcal{X} \subseteq \mathcal{N}$ ,  $\mathcal{X}^c$  denotes the set complement of  $\mathcal{X}$  such that  $\mathcal{X} \cup \mathcal{X}^c = \mathcal{N}$  and  $\mathcal{X} \cap \mathcal{X}^c = \{\emptyset\}$ . We denote the cardinality of  $S$  as  $|S|$ . The support set of a vector  $x$  is denoted by  $\text{supp}(x) = \{i : [x]_i \neq 0\}$ .  $\|\cdot\|_q$  denotes the  $\ell_q$ -norm where  $\|x\|_q = (\sum_{i=1}^n |[x]_i|^q)^{1/q}$ , for  $x \in \mathbb{R}^n$ . The gradient of  $f(x) := \|y - \Phi x\|_2^2$  is computed as  $\nabla f(x) = -2\Phi^T(y - \Phi x)$ .

**Preliminaries:** We assume  $x^* \in C_k$ , where  $C_k$  is a combinatorial sparsity model. Moreover, we assume  $\Phi$  satisfies the RIP-condition. These properties are defined below:

**Definition 1** (Combinatorial sparsity model (CSM)). *We define a combinatorial sparsity model  $C_k = \{S_j : \forall j, S_j \subseteq \mathcal{N}, |S_j| \leq k\}$  with sparsity parameter  $k$  as a collection of index subsets  $S_j$  of  $\mathcal{N}$ .*

**Definition 2** (Restricted Isometry Property (RIP) [13]). *A matrix  $\Phi$  satisfies the RIP with constant  $\delta_k \in (0, 1)$  if and only if:*

$$(1 - \delta_k)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_k)\|x\|_2^2, \quad \forall x \in \mathbb{R}^n \text{ s.t. } \text{supp}(x) \in C_k.$$

## 2. EUCLIDEAN PROJECTIONS ONTO CONSTRAINTS

### 2.1. Projections onto combinatorial sets

The Euclidean projection of a signal  $w \in \mathbb{R}^n$  on the subspace defined by  $C_k$  is provided by:

$$\mathcal{P}_{C_k}(w) = \arg \min_{x: x \in C_k} \|x - w\|_2. \quad (6)$$

Lemma 1 relates  $\mathcal{P}_{C_k}(\cdot)$  operator to discrete optimization:

**Lemma 1** (Modularity of Euclidean projections onto CSMs [12]). *The support of the Euclidean projection onto  $C_k$  in (6) can be obtained as a solution to the following discrete optimization problem:*

$$\text{supp}(\mathcal{P}_{C_k}(w)) = \arg \max_{S: S \in C_k} F(S; w), \quad (7)$$

where  $F(S; w) \triangleq \|w\|_2^2 - \|w_S\|_2^2 = \sum_{i \in S} |[w]_i|^2$ . Moreover, let  $\hat{S} \in C_k$  be the maximizer of the discrete problem in (7). Then, it holds that  $\mathcal{P}_{C_k}(w) = w_{\hat{S}}$ , which is hard thresholding.

Many CSMs are encoded via matroids and totally unimodular (TU) models where (7) can be solved efficiently.

### 2.2. Projections onto convex norms

Given  $w \in \mathbb{R}^n$ , the Euclidean projection onto a convex  $\ell_*$ -norm ball of radius at most  $\lambda$  defines the optimization problem:

$$\mathcal{P}_{\ell_*}(w, \lambda) = \arg \min_{x: \|x\|_* \leq \lambda} \|x - w\|_2. \quad (8)$$

In this work, we focus on the  $\ell_1$ ,  $\ell_2$ ,  $\ell_\infty$ , and TV norm constraints.

An efficient projection onto the  $\ell_1$ -norm ball requires only  $O(n)$  expected time complexity [14]; however, we use an  $O(n \log n)$  implementation in our framework, as described in [14]. It is trivial to calculate the projection onto the  $\ell_2$  and  $\ell_\infty$ -norm balls with radius  $\lambda$  in  $O(n)$  time complexity.

We adopt the discrete TV norm definition in [15]. With the minor modifications, we can obtain the projection onto the TV-norm ball with the constraint, where the signal support is restricted to  $S$ , by using the algorithms described in [15]. The complexity of these projections are on the order of the cost of applying the adjoint of  $\Phi$ .

---

## Algorithm 1 NORMED PURSUITS

---

**Input:**  $y, \Phi, \lambda, \mathcal{P}_{C_k}$ , Tolerance, MaxIterations

**Initialize:**  $x_0 \leftarrow 0, \mathcal{X}_0 \leftarrow \{\emptyset\}, i \leftarrow 0$

**repeat**

- 1:  $\hat{\mathcal{X}}_i \leftarrow \text{supp}(\mathcal{P}_{C_k}(\nabla_{\mathcal{X}_i^c} f(x_i))) \cup \mathcal{X}_i$
- 2:  $v_i \leftarrow \arg \min_{v: v \in \mathcal{V}_i} \|y - \Phi v\|_2^2$
- 3:  $\gamma_i \leftarrow \mathcal{P}_{C_k}(v_i)$  with  $\Gamma_i \leftarrow \text{supp}(\gamma_i)$
- 4:  $x_{i+1} \leftarrow \arg \min_{w: w \in \mathcal{W}_i} \|y - \Phi w\|_2^2$  with  $\mathcal{X}_{i+1} \leftarrow \text{supp}(x_{i+1})$   
 $i \leftarrow i + 1$

**until**  $\|x_i - x_{i-1}\|_2 \leq \text{Tolerance}\|x_i\|_2$  or MaxIterations.

*Nomenclature:*  $\mathcal{V}_i \triangleq \{v : \text{supp}(v) \in \hat{\mathcal{X}}_i, \|v\|_* \leq \lambda\}$ ,

$\mathcal{W}_i \triangleq \{w : \text{supp}(w) \in \Gamma_i, \|w\|_* \leq \lambda\}$ .

---

## 3. EXPLOITING BOTH CLASSES OF CONSTRAINTS

In this section, we define a broad class of sparse recovery algorithms, using different combinatorial models and norm constraints, which constitute the NORMED PURSUITS:

$$\hat{x}_{\text{NORMED PURSUITS}} = \arg \min \{f(x) : \|x\|_* \leq \lambda, \text{supp}(x) \in C_k\},$$

where  $\|\cdot\|_*$  is a convex norm, such as  $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$  and  $\|\cdot\|_{\text{TV}}$ . The CLASH algorithm in [12] is a special case with  $*$  = 1.

We provide a pseudo-code of a NORMED PURSUITS variant in Algorithm 1. The presented instance is based on the Subspace Pursuit algorithm [3], and it is trivial to come up with other variants by using the existing hard thresholding methods for computational trade-offs [2–6] (left to the reader). At each iteration, the algorithm goes through the following motions:

1) *Active set expansion (Step 1):* We identify the support where the projected gradient onto  $C_k$  can make most impact on the loading vector in the support complement of its current solution. We then merge this support with the support of the current solution.

2) *Descent with the norm constraint (Step 2):* We decrease the data error  $f(x)$  as much as possible on the active set with the  $\|\cdot\|_*$ -norm constraint by exploiting the convex projector of the norm.

3) *Combinatorial selection (Step 3):* We project the constrained solution onto  $C_k$  to arbitrate the active support set.

4) *De-bias with the norm constraint (Step 4):* We de-bias the result on the current support with the  $\|\cdot\|_*$ -norm constraint.

NORMED PURSUITS features the following guarantee:

**Theorem 1.** [Iteration invariant] *The  $i$ -th iterate  $x_i$  of NORMED PURSUITS in Algorithm 1 satisfies the following recursion*

$$\|x_{i+1} - x^*\|_2 \leq \rho \|x_i - x^*\|_2 + c_1(\delta_{2k}, \delta_{3k})\|\varepsilon\|_2,$$

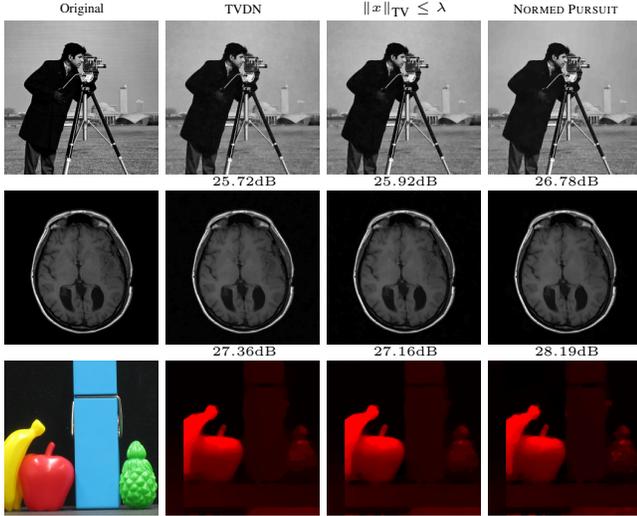
where  $c_1(\delta_{2k}, \delta_{3k}) \triangleq \frac{1}{\sqrt{1-\delta_{2k}^2}} \left( \sqrt{1+3\delta_{3k}^2} \left( \sqrt{\frac{2(1+\delta_{3k})}{1-\delta_{3k}^2}} + \sqrt{\frac{1+\delta_{2k}}{1-\delta_{3k}}} \right) + \sqrt{3(1+\delta_{2k})} \right) + \frac{\sqrt{1+\delta_k}}{1-\delta_{2k}}$  and  $\rho \triangleq \frac{\delta_{3k} + \delta_{2k}}{\sqrt{1-\delta_{2k}^2}} \sqrt{\frac{1+3\delta_{3k}^2}{1-\delta_{3k}^2}}$ . Moreover, when  $\delta_{3k} < 0.3658$ , the iterations are contractive (i.e.,  $\rho < 1$ ).

The proof of Theorem 1 is a direct extension of the CLASH proof in [12], and is omitted due to lack of space. Surprisingly, Theorem 1 shows that the isometry requirements of NORMED PURSUITS are competitive with those of mainstream hard thresholding methods, even with the additional norm constraints.

**Extensions:** A key strength of NORMED PURSUITS is the ability to explicitly enforce sparsity using efficient combinatorial pro-

Method	NORMED PURSUITS Formulation
Elastic net [9]	$\hat{x} = \arg \min \{f(x) : \ x\ _0 \leq k, \ x\ _2 \leq \lambda\}$
Fused Lasso [10]	$\hat{x} = \arg \min \{f(x) : \ x\ _0 \leq k, \ x\ _{\text{TV}} \leq \lambda\}$
Group Lasso [11]	$\hat{x} = \arg \min \{f(x) : x \in \mathcal{C}_k, \ x_G\ _0 \leq k\}$

**Fig. 2.** NORMED PURSUITS problem formulation of [9–11]. Here,  $\|x_G\|_0 \leq k$  represents sparsity constraints over groups.



**Fig. 3.** Results from real data.

jections,<sup>1</sup> while using the norm constraints to regularize the sparse coefficient values. The current sparse recovery literature offers a variety of convex optimization formulations that attempt to capture the strength of combinatorial models via *exclusively* norm information [9–11]. In Fig. 2, we present the corresponding NORMED PURSUITS formulations of [9–11].

## 4. EXPERIMENTS

### 4.1. Real Data

In this subsection, we evaluate the following formulation:

$$\hat{x}_{\text{TV-PURSUIT}} = \arg \min \{f(x) : \|x\|_{\text{TV}} \leq \lambda, \|\Psi x\|_0 \leq k\}, \quad (9)$$

where  $\Psi$  represents an orthonormal wavelet transform.

**Compressive sensing with TV-NORMED PURSUIT:** To study the compressive sensing recovery performance of NORMED PURSUITS, we use the classical cameraman and brain<sup>2</sup> images of  $n = 256 \times 256$  pixels. We compressively measure both images with the spread spectrum technique [16]. That is, the sensing matrix  $\Phi$  consists of a random pre-modulation followed by a random selection of  $m = 0.25n$  complex Fourier coefficients. The performance of NORMED PURSUIT (9) is compared with: *i*) the TV version of Basis Pursuit where the TV norm is substituted for the  $\ell_1$ -norm, *ii*) the TV-constrained version of Lasso (5). We choose the Daubechies-4 wavelet for  $\Psi$  and  $k = m$ . The parameter  $\lambda$  in (9) and (5) was chosen to obtain the best reconstruction for each method. Fig. 3 shows

<sup>1</sup>with combinatorial algorithms that go beyond simple selection heuristics towards provable solution quality as well as runtime/space bounds.

<sup>2</sup>BRAINIX database: <http://pubimage.hcuge.ch:8080/>.

the reconstruction obtained with the three methods. NORMED PURSUIT (9) *outperforms* all the other methods with an improvement of at least 0.8 dB on the signal-to-noise ratio.

**CASSI recovery with TV-NORMED PURSUIT:** We test the performance of our approach using the Coded Aperture Snapshot Spectral Imager (CASSI) data [17].<sup>3</sup> We reconstruct three-dimensional spatio-spectral data cube from measurements. While we obtain the full set of images, we only provide the result at the wavelength 549 nm due to lack of space. Full results are available at <http://lions.epfl.ch/TV-Normed-Pursuit>. We also illustrate results using the three algorithms mentioned above in Fig. 3. In general, the contours are better resolved with NORMED PURSUIT compared to the other methods studied. Moreover, in our subjective evaluation, the contrast is improved overall across all the wavelengths.

### 4.2. Synthetic Data

In this subsection,  $\Phi$  is populated with independent and identically distributed (iid) Gaussian random variables with zero mean and variance  $1/m$ . The nonzero coefficients of  $x^*$  are generated iid according to the standard normal distribution with  $\|x^*\|_2 = 1$ .

**Recovery with CLASH.** We generate random realizations of the model  $y = \Phi x^*$  for  $n = 200$ ,  $m = 64$  and  $k = 23$ , where  $k$  is known a-priori. We then sweep  $\lambda$  and examine the signal recovery performance of CLASH compared to the following methods: *i*) Lasso (5) using a Nesterov first order method, *ii*) Basis Pursuit [1] using the SPGL1 implementation [18], *iii*) Subspace Pursuit (SP) [3], and *iv*) NORMED PURSUIT with  $\ell_2$ -norm constraint. Note that, if  $\lambda$  is large, norm constraints have no impact in recovery and CLASH and *iv*) must admit identical performance to SP.

Figure 4(a) illustrates that the combination of hard thresholding with norm constraints can *improve* the signal recovery performance significantly over convex-only and hard thresholding-only methods. CLASH perfectly recovers the signal when  $\lambda \in [1, 1.2]$ . When  $\lambda < \|x^*\|_1$ , the performance degrades.

**$\pm 1$ -signal recovery with NORMED PURSUITS.** We instantiate (1) where  $\varepsilon$  satisfies  $\|\varepsilon\|_2 = 0.1$ , and  $n = 125$ ,  $m = 65$ ,  $k = 25$ . Here, the coefficients of  $x^*$  are randomly assigned to  $\pm 1$  values. To reconstruct  $x^*$  from  $y$ , we test NORMED PURSUITS with *i*)  $\ell_1$ -norm, *ii*)  $\ell_2$ -norm and, *iii*)  $\ell_\infty$ -norm.

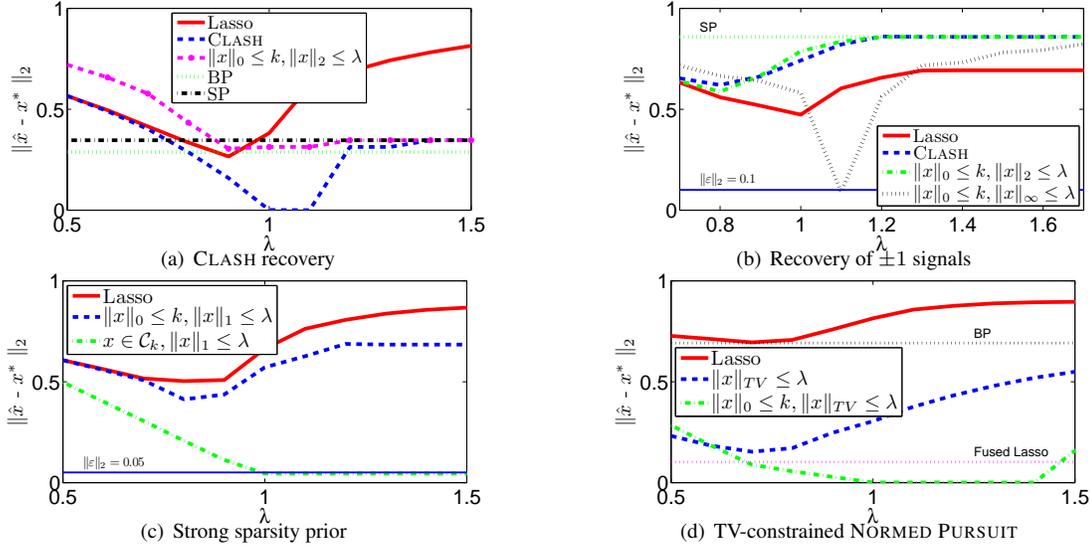
We illustrate the signal recovery results in Fig. 4(b). We notice that the reconstruction performance varies by using different norm constraints; in the case of signed signals, we deduce that  $\ell_\infty$  norm provides the best results as compared to  $\ell_1$  and  $\ell_2$ -norm constraints.

**Norm-constraints do not always help.** In the model (1), we assume  $x^* \in \mathcal{C}_k$  has  $k$  nonzero coefficients in  $C$ -contiguous blocks on an undirected, acyclic chain graph; the CSM  $\mathcal{C}_k$  is dubbed as the  $(k, C)$ -clustered sparsity model—c.f. [19] for details.

In Fig. 4(c), we observe that structure prior information provides great advantage in signal reconstruction compared to Lasso (5) formulation and the CLASH algorithm with simple sparsity model assumption. Here, the noise energy level satisfies  $\|\varepsilon\|_2 = 0.05$ , and  $n = 500$ ,  $m = 125$ , and  $k = 50$ . As  $\lambda$  increases, while CLASH performance improve, norm constraints have *no effect* in the model-based recovery procedure; the structure dominates the norm constraints, providing the enough information for robust recovery up to the noise level.

**$\|\cdot\|_{\text{TV}}$  provides a strong norm constraint.** Here,  $x^*$  follows the  $(k, C)$ -clustered sparsity model where the clustered nonzero elements have approximately flat values. We consider the noiseless case  $y = \Phi x^*$  for  $n = 500$ ,  $m = 100$  and  $k = 50$ .

<sup>3</sup><http://www.disp.duke.edu/projects/CASSI>



**Fig. 4.** For each  $\lambda$ , we run 100 Monte-Carlo iterations and pick the median value of signal data error  $\|\hat{x} - x^*\|_2$ .

We now compare (9), where  $\Psi$  is identity, with: *i*) the Lasso method, *ii*) Basis Pursuit in (3) using the SPGL1 implementation, *iii*) the TV-constrained version of Lasso (5) where  $\|x\|_1 \leq \lambda$  is replaced with  $\|x\|_{TV} \leq \lambda$  and, *iv*) Fused Lasso [10] with TV and  $\ell_1$ -norm constraints, where the *true regularization parameters are assumed known*. Figure 4(d) provides empirical evidence that hard thresholding with the TV-norm constraint *outperforms* the other algorithms in terms of signal reconstruction, where sparsity constraints assist norm-constrained optimization in the estimation performance.

## 5. CONCLUSIONS

We introduce a sparse signal recovery paradigm, dubbed NORMED PURSUITS, where efficient algorithms from combinatorial and convex optimization can efficiently interface for interpretable and model-based sparse solutions. NORMED PURSUITS has the ability to explicitly enforce sparsity using efficient combinatorial projections, while using the norm constraints to regularize the sparse coefficient values, whereby enhancing the performance of the class of hard thresholding methods. Alternatively, we can view NORMED PURSUITS as a sparsity guided active set method in convex optimization (note that the  $\ell_1$  methods themselves typically do not exploit the fact that they are looking for sparse solutions).

## 6. REFERENCES

- [1] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM Journal on Scientific Computing*, 20:33, 1998.
- [2] A. Kyrillidis and V. Cevher. Recipes for hard thresholding methods. *Technical Report*, 2011.
- [3] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Transactions on Information Theory*, vol. 5, 2009.
- [4] D. Needell and J. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):16–42, 2007.
- [5] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009.
- [6] S. Foucart. Hard thresholding pursuit: An algorithm for compressive sensing, preprint, 2011.
- [7] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1):267–288, 1996.
- [8] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, vol. 56, 2010.
- [9] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67:301–320, 2005.
- [10] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused Lasso. *Journal of the Royal Statistical Society: Series B*, 67(1):91–108, 2005.
- [11] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*, 68(1):49–67, 2006.
- [12] A. Kyrillidis and V. Cevher. Combinatorial selection and least absolute shrinkage via the CLASH algorithm. *Technical Report*, 2011.
- [13] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans on Info. Theory*, 2006.
- [14] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient projections onto the  $l_1$ -ball for learning in high dimensions. In *ICML*, 2008.
- [15] J.M. Fadili and G. Peyré. Total variation projection with first order schemes. *IEEE Transactions on Image Processing*, vol. 20, 2011.
- [16] G. Puy, P. Vandergheynst, R. Gribonval, and Y. Wiaux. Spread spectrum for universal compressive sampling. In *SPARS*, 2011.
- [17] A. Wagadarikar, N. P. Pitsianis, X. Sun, and D. J. Brady. Spectral image estimation for coded aperture snapshot spectral imagers. *Proceedings of SPIE*, 7076, 2008.
- [18] E. van den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 2008.
- [19] V. Cevher, P. Indyk, C. Hegde, and R.G. Baraniuk. Recovery of clustered sparse signals from compressive measurements. *SAMPTA*, 2009.