Hard Thresholding with Norm Constraints

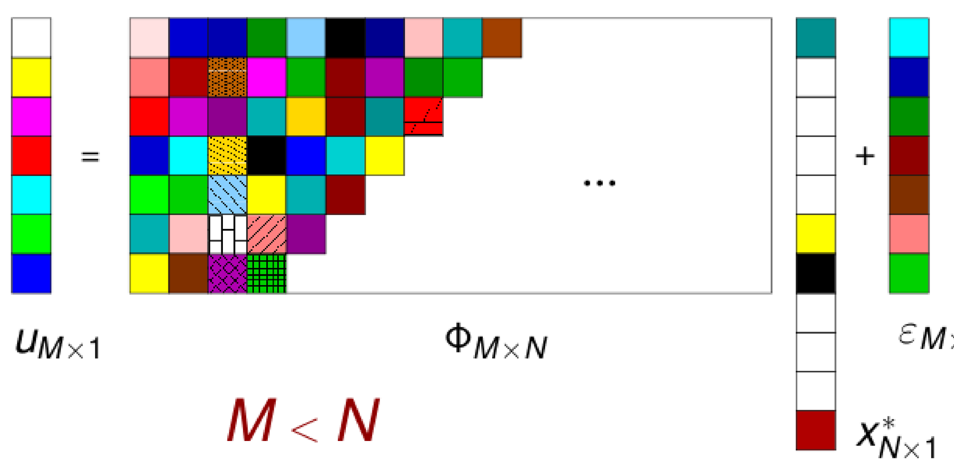
Anastasios Kyrillidis Gilles Puy Volkan Cevher

Laboratory for Information and Inference Systems, EPFL

{anastasios.kyrillidis, gilles.puy, volkan.cevher}@epfl.ch

Problem statement

- Underdetermined linear regression:



- **Goal**: find signal x^* that generated the set of measurements u via the sampling matrix Φ.
- **Challenge**: Non-trivial nullspace of Φ.
- **Signal prior**: Sparsity, i.e., only $K \ll N$ coordinates are nonzero. Sparsity models:
  1. Simple sparsity: $x \in \Sigma_K$.
  2. Structured (model-based) sparsity: $x \in \Sigma_{\mathcal{M}_K}$.
- **Paper overview**:
  1. **Combinatorial and norm constraints** in sparse recovery.
  2. Attractive theoretical guarantees.
  3. Combination of hard thresholding with norm constraints outperforms the state-of-the-art approaches.

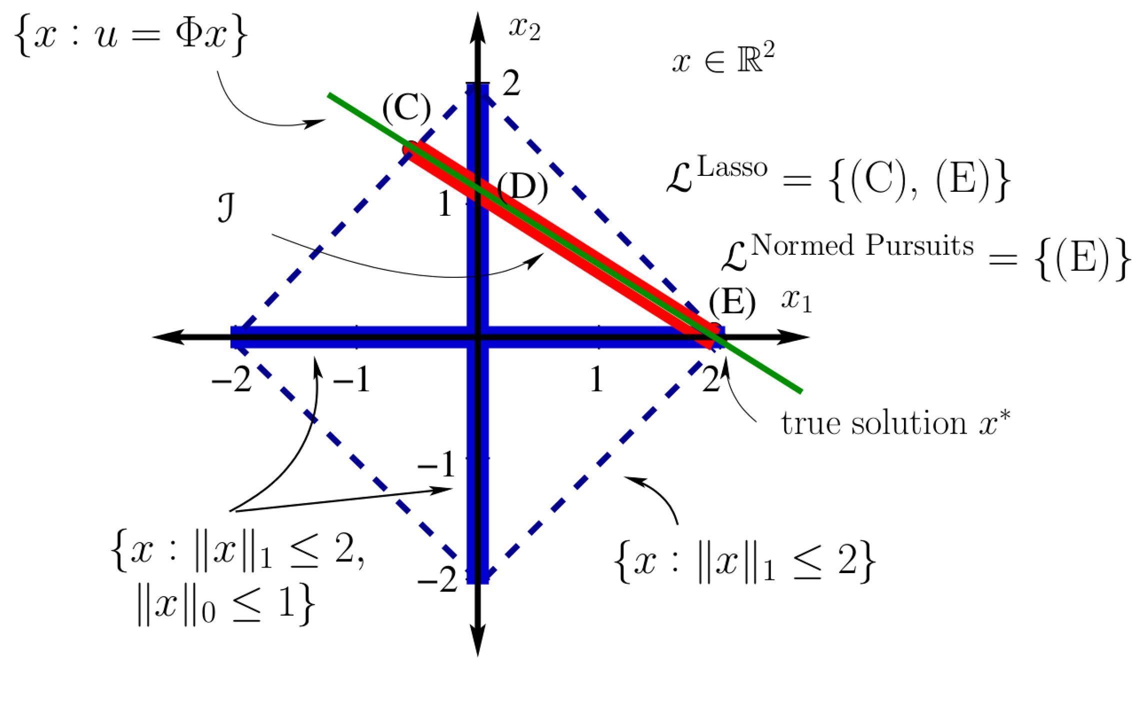
A Clash of algorithms

	Geometric
Encoding	atomic norm/convex relaxation
Example	$\min_{x:\|x\|_* \leq \lambda} \|u - \Phi x\|_2^2$, $* = 1, 2, \infty$, TV
Algorithms	BP, BPDN, Lasso, TV-Pursuit, ...
	Combinatorial
Encoding	non-convex union of subspaces
Example	$\min_{x:\|x\|_0 \leq K} \|u - \Phi x\|_2^2$
Algorithms	OMP, IHT, ALPS[1], CoSaMP, SP, HTP, ...

$$\hat{x} = \underset{\|x\|_* \leq \lambda, \|x\|_0 \leq K}{\operatorname{argmin}} \|u - \Phi x\|_2^2$$

[1] http://lions.epfl.ch/ALPS

Motivation



$$\mathcal{J} = \{\hat{x} : u = \Phi\hat{x} \text{ and } \|\hat{x}\|_1 \leq 2\}.$$

Forcing basic solutions in optimization:

$$\mathcal{L}^{\text{Lasso}} = \mathcal{J} \cap \{\hat{x} : \|\hat{x}\|_1 = 2\},$$

$$\mathcal{L}^{\text{Normed Pursuits}} = \mathcal{J} \cap \{\hat{x} : \|\hat{x}\|_1 = 2, \|\hat{x}\|_0 = 1\}$$

Combinatorial selection

- **Combinatorial sparsity models (CSMs)**:

$\mathcal{M}_K = \{\mathcal{S}_m : \forall m, \mathcal{S}_m \subseteq \{1, \dots, N\}, |\mathcal{S}_m| \leq K\}$.

- **Non-convex projection onto CSMs**:

$\mathcal{P}_{\mathcal{M}_K}(x) = \arg\min \left\{ \|y - x\|_2^2 : \text{supp}(y) \in \mathcal{M}_K \right\}$.

- **Lemma** *[Modularity of projections onto CSMs]*:

$$\text{supp}\left(\mathcal{P}_{\mathcal{M}_K}(x)\right) = \underset{\mathcal{S}:\mathcal{S}\in\mathcal{M}_K}{\arg\max} \sum_{i\in\mathcal{S}} |[x]_i|^2$$

- CSM projections via linear integer programs:

$$\text{supp}\Big(\underset{z:[z]_i \in \{0,1\}}{\arg\min} \{w^T z : \text{supp}(z) \in \mathcal{M}_K\}\Big)$$

where $[w]_i = -|[x]_i|^2$.

- Example CSMs:

i) intersection of the uniform matroid with any matroid,

ii) linear constraints: $Az \leq b$ for special A's ...

Regression with convex and non-convex constraints

Algorithm 1: Normed Pursuits

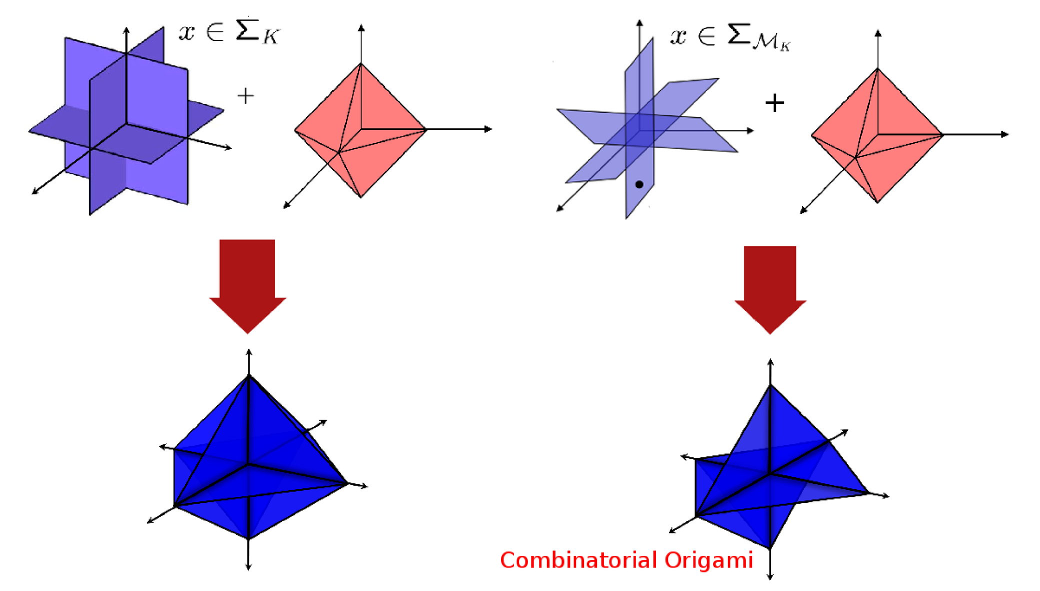
1. **Input:** u, Φ, λ, $\mathcal{P}_{\mathcal{M}_K}$, $\|\cdot\|_*$, Tolerance, MaxIter
2. **Initialize:** $x_0 \leftarrow 0$, $\mathcal{X}_0 \leftarrow \{\emptyset\}$, $i \leftarrow 0$

repeat

3. Active set expansion
4. Greedy descent with norm constraint
5. Combinatorial selection
6. De-bias with norm constraint

until *Stopping criteria or MaxIterations.*;

Example with the ℓ_1-norm constraint



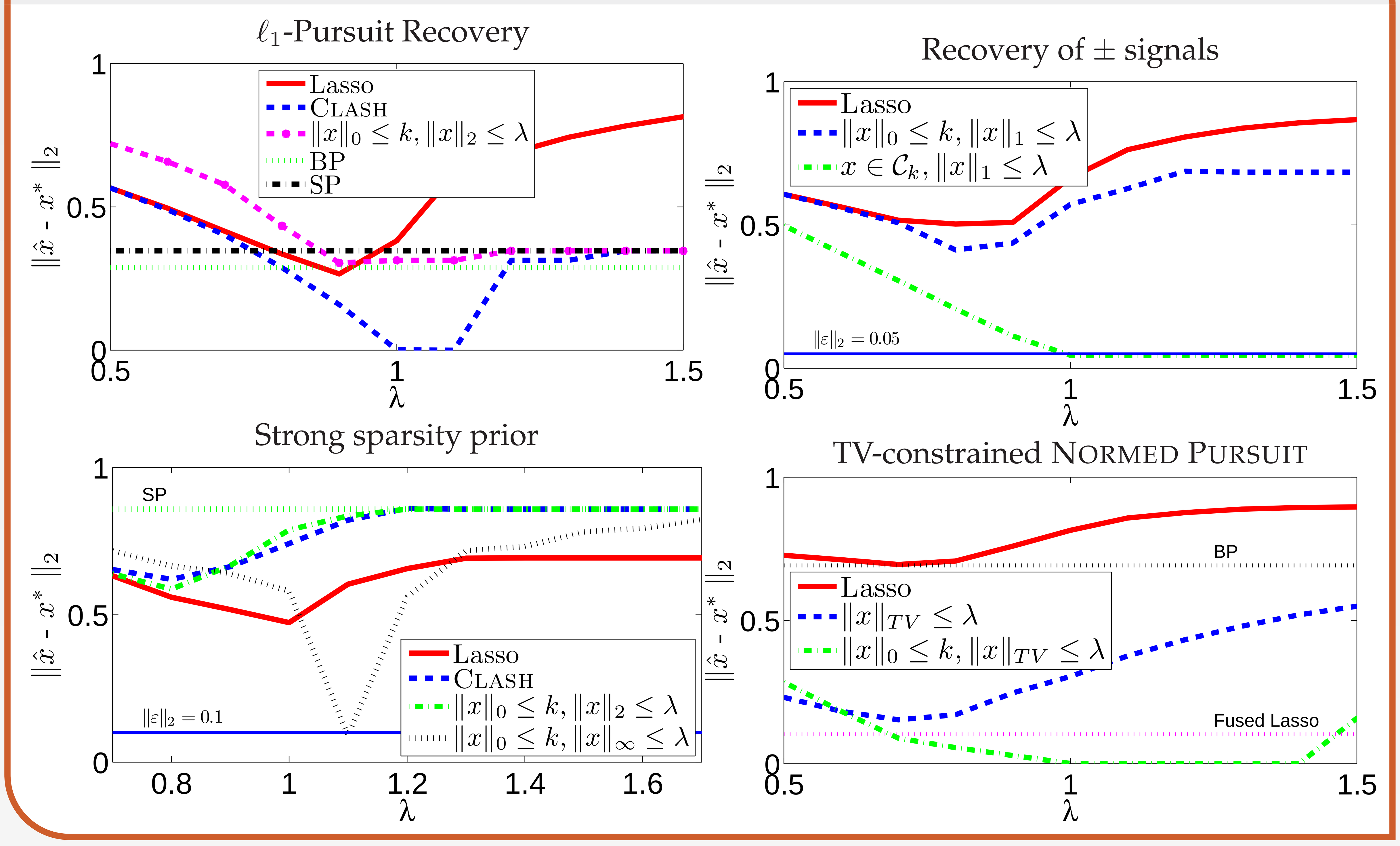
Convergence rate and approximation guarantees

Theorem: Let Φ satisfy the RIP with $\delta_{cK} \in (0,1)$: $(1 - \delta_{cK})\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_{cK})\|x\|_2^2$, $\forall x \in \Sigma_{\mathcal{M}_{cK}}$. Then, the proposed set of algorithms satisfies the following worst-case guarantee:

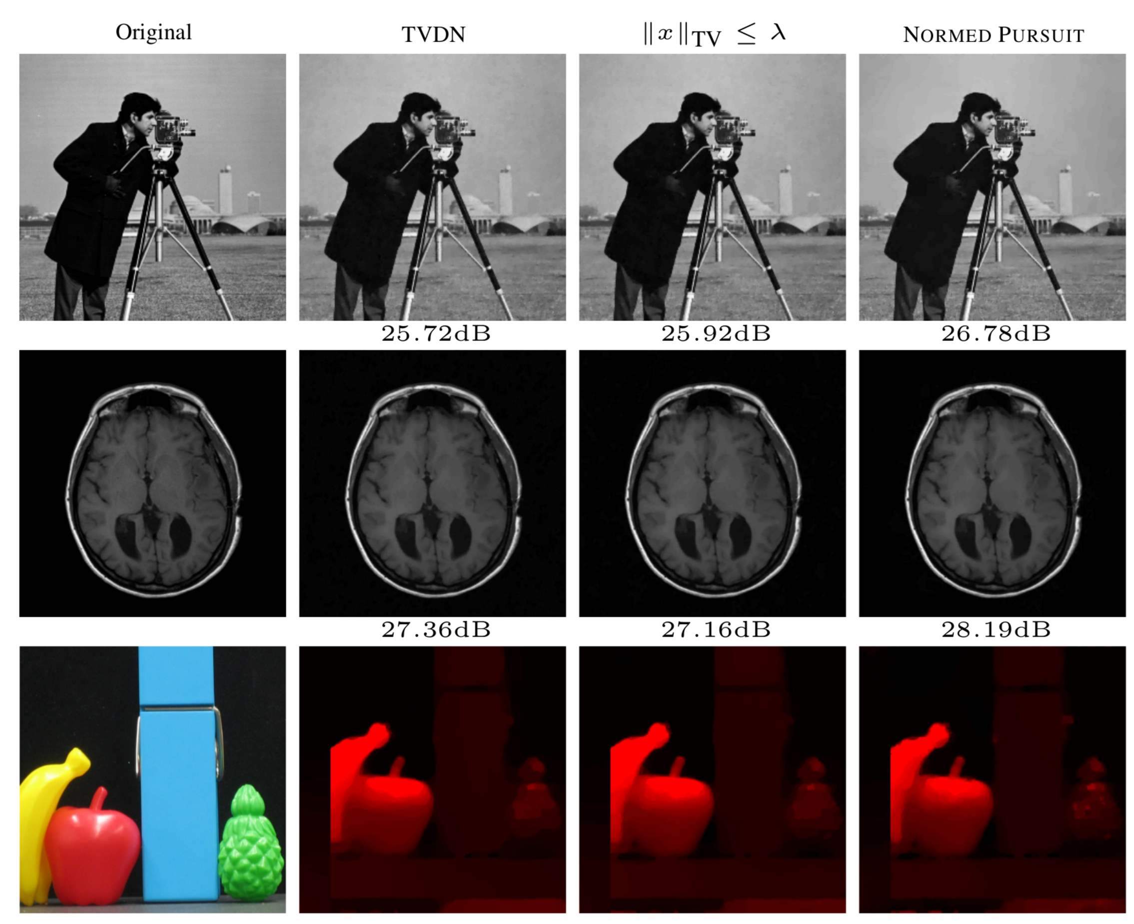
$$\|x_{i+1} - x^*\|_2 \leq \rho\|x_i - x^*\|_2 + c_1(\delta_{2K}, \delta_{3K})\|\varepsilon\|_2,$$

where $\rho = \frac{\delta_{3K} + \delta_{2K}}{\sqrt{1 - \delta_{2K}^2}} \sqrt{\frac{1 + 3\delta_{3K}^2}{1 - \delta_{3K}^2}}$ and $c_1(\delta_{2K}, \delta_{3K})$ is a small constant. Moreover, the iterations are contractive if $\delta_{3s} < 0.3658$ and $c_1(\delta_{2K}, \delta_{3K}) < 8.62$ for $\delta_{3s} < 0.3658$.

Synthetic results



Real data results



(TVDN): $\underset{x\in\mathbb{R}^N}{\text{minimize}} \ \|x\|_{\text{TV}}$ s.t. $\|u - \Phi x\|_2 \leq \sigma$

$(\|x\|_{\text{TV}} \leq \lambda)$: $\underset{x\in\mathbb{R}^N}{\text{minimize}} \ \|u - \Phi x\|_2$ s.t $\|x\|_{\text{TV}} \leq \lambda$