Iterative Regularization with k-Support Norm: an Important Complement to Sparse Recovery

William de Vazelhes[1], Bhaskar Mukhoty[1], Xiaotong Yuan[2], Bin Gu[1,3]

[1] MBZUAI, Abu Dhabi, UAE
[2] Nanjing University, Suzhou, China
[3] Jilin University, Changchun, China
{wdevazelhes,bhaskar.mukhoty,xtyuan1980,jsgubin}@gmail.com

AAAI, February 2024

Introduction

Compressed sensing problem: observe $\boldsymbol{X}$ (input) and $\boldsymbol{y}^\delta$ (output) and reconstruct $\boldsymbol{w}^*$, assumed to be k-sparse, and with noise $\boldsymbol{\epsilon}$ such that $\|\boldsymbol{\epsilon}\|_2 \leq \delta$:

$$\boldsymbol{y}^\delta = \boldsymbol{X}\boldsymbol{w}^* + \boldsymbol{\epsilon}$$

Contribution: We propose an **algorithm with new conditions for recovery**, complementing usual existing ones based on ℓ_1 norm.

Notations

- For $S \subseteq [d]$, $\bar{S} := [d] \setminus S$
- $\boldsymbol{M}^{\dagger}$: Moore-Penrose pseudo-inverse [3]
- $\|\boldsymbol{M}\|$: nuclear norm
- $\boldsymbol{M}_S$ column-restriction of $\boldsymbol{M}$ to support $S \subseteq [d]$, i.e. the $n \times |S|$ matrix composed of the $|S|$ columns of $\boldsymbol{M}$ of indices in S
- $\mathrm{supp}(\boldsymbol{w})$: support of $\boldsymbol{w}$ (coordinates of the non-zero components of $\boldsymbol{w}$)
- $\boldsymbol{w}_S \in \mathbb{R}^k$ restriction of $\boldsymbol{w}_S$ to a support S of size k, i.e. the sub-vector of size k formed by extracting only the components w_i with $i \in S$
- $\mathrm{sgn}(\boldsymbol{w})$ vector of signs of $\boldsymbol{w}$

Conditions for recovery

Method	Condition on $\boldsymbol{X}$
IHT [2]	Restricted Isometry Property (RIP)
Lasso [8]	$\max_{\ell \in \bar{S}} \lvert\langle \boldsymbol{X}_S^\dagger \boldsymbol{x}_\ell, \mathrm{sgn}(\boldsymbol{w}_S^*)\rangle\rvert < 1^{(2)}$
ElasticNet [11]	-
KSN pen. [1]	-
OMP [9]	RIP
SRDI [7]	$\begin{cases} \exists \gamma \in (0,1] : \ \boldsymbol{X}_S^\top \boldsymbol{X}_S \geq n\gamma \boldsymbol{I}_{d,d} \\ \exists \eta \in (0,1) : \ \lVert \boldsymbol{X}_{\bar{S}} \boldsymbol{X}_S^\dagger \rVert_\infty \leq 1 - \eta \end{cases}$
IROSR [10]	RIP
IRCR [6]	$\max_{\ell \in \bar{S}} \lvert\langle \boldsymbol{X}_S^\dagger \boldsymbol{x}_\ell, \mathrm{sgn}(\boldsymbol{w}_S^*)\rangle\rvert < 1^{(2)}$
IRKSN (ours)	$\max_{\ell \in \bar{S}} \lvert\langle \boldsymbol{X}_S^\dagger \boldsymbol{x}_\ell, \boldsymbol{w}_S^*\rangle\rvert < \min_{j \in S} \lvert\langle \boldsymbol{X}_S^\dagger \boldsymbol{x}_j, \boldsymbol{w}_S^*\rangle\rvert$

Our contribution

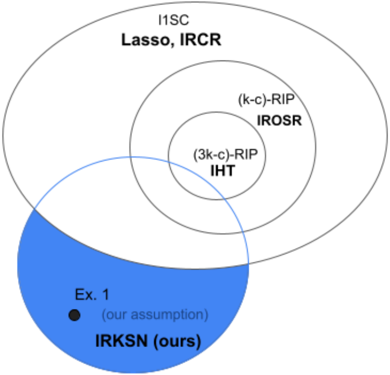


Figure: Conditions for recovery. In some cases (in blue), only IRKSN (our algorithm) can provably ensure sparse recovery.

Iterative Regularization

Iterative regularization (see e.g. IRCR [6]), solves the following problem with early stopping:

$$\min_{\boldsymbol{w}} R(\boldsymbol{w})$$
$$\text{s.t.} \quad \boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}^{\delta}$$

IRCR [6] uses $R(\boldsymbol{w}) = \|\boldsymbol{w}\|_1$. We propose to use instead a regularizer based on the k-support norm:

$$R(\boldsymbol{w}) = F(\boldsymbol{w}) + \frac{\alpha}{2}\|\boldsymbol{w}\|_2{}^2$$

where

$$F(\boldsymbol{w}) = \frac{1-\alpha}{2}(\|\boldsymbol{w}\|_k^{sp})^2$$

with $\|\cdot\|_k^{sp}$ is the k-support norm. We solve it via a primal-dual algorithm from [4].

Note on the k-support norm (KSN)

- KSN ball is tightest convex relaxation of ℓ_0 and ℓ_2 ball:

$$\{\boldsymbol{x} : \|\boldsymbol{x}\|_k^{sp} \leq D\} = \text{conv}(\{\boldsymbol{x} : \|\boldsymbol{x}\|_0 \leq k\} \cap \{\boldsymbol{x} : \|\boldsymbol{x}\|_2 \leq D\})$$

- The proximal operator for the squared KSN is known [5].

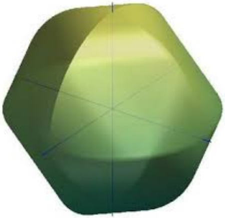


Figure: k-support norm ball (source: [1])

Algorithm: IRKSN

Algorithm 1: IRKSN

Initialization: $\hat{\boldsymbol{v}}_0 = \hat{\boldsymbol{z}}_{-1} = \hat{\boldsymbol{z}}_0 \in \mathbb{R}^d, \gamma = \alpha \|\boldsymbol{X}\|^{-2}, \boldsymbol{x}_0 = 1$

for $t = 0, ..., T$ **do**

$\hat{\boldsymbol{w}}_t \leftarrow \text{prox}_{\alpha^{-1}F}\left(-\alpha^{-1}\boldsymbol{X}^T\hat{\boldsymbol{z}}_{\boldsymbol{t}}\right)$

$\hat{\boldsymbol{r}}_t \leftarrow \text{prox}_{\alpha^{-1}F}\left(-\alpha^{-1}\boldsymbol{X}^T\hat{\boldsymbol{v}}_t\right)$

$\hat{\boldsymbol{z}}_t \leftarrow \hat{\boldsymbol{v}}_t + \gamma\left(\boldsymbol{X}\hat{\boldsymbol{r}}_{\boldsymbol{t}} - \boldsymbol{y}^\delta\right)$

$\theta_{t+1} \leftarrow \left(1 + \sqrt{1 + 4\theta_t^2}\right)/2$

$\hat{\boldsymbol{v}}_{t+1} = \hat{\boldsymbol{z}}_t + \frac{\theta_t - 1}{\theta_{t+1}}\left(\hat{\boldsymbol{z}}_t - \hat{\boldsymbol{z}}_{t-1}\right)$

end

Sufficient conditions for recovery: comparison with ℓ_1 norm

Assumption (Conditions for recovery with ℓ_1 norm-based algorithms)

Let $\boldsymbol{w}^*$ be supported on a support $S \subset [d]$. $\boldsymbol{w}^*$ is such that:

1. $\boldsymbol{X}\boldsymbol{w}^* = \boldsymbol{y}$
2. $\boldsymbol{X}_S$ *is injective*
3. $\max_{\ell \in \bar{S}} |\langle \boldsymbol{X}_S^\dagger \boldsymbol{x}_\ell, \mathrm{sgn}(\boldsymbol{w}_S^*) \rangle| < 1$

Assumption (Conditions for recovery with IRKSN)

- $\boldsymbol{w}^*$ *k-sparse, supp*$(\boldsymbol{w}^*) = S \subset [d]$, $\boldsymbol{X}\boldsymbol{w}^* = \boldsymbol{y}$
- $\boldsymbol{w}_S^* = \arg\min_{\boldsymbol{z} \in \mathbb{R}^k : \boldsymbol{X}_S \boldsymbol{z} = \boldsymbol{y}} \|\boldsymbol{z}\|_2$
- $\max_{\ell \in \bar{S}} |\langle \boldsymbol{X}_S^\dagger \boldsymbol{x}_\ell, \boldsymbol{w}_S^* \rangle| < \min_{j \in S} |\langle \boldsymbol{X}_S^\dagger \boldsymbol{x}_j, \boldsymbol{w}_S^* \rangle|$
- *Does not need $\boldsymbol{X}_S$ to be injective !*

Conditions for recovery, case where $\boldsymbol{X}_S$ is injective

If $\boldsymbol{X}_S$ is injective and $\boldsymbol{X}\boldsymbol{w}^* = \boldsymbol{y}$, the conditions become:

Assumption (Conditions for recovery with ℓ_1 norm-based algorithms)

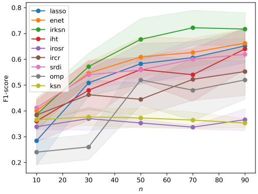
$$(A) : \max_{\ell \in \bar{S}} |\langle \boldsymbol{X}_S^{\dagger} \boldsymbol{x}_\ell, \operatorname{sgn}(\boldsymbol{w}_S^*) \rangle| < 1$$

Assumption (Conditions for recovery with IRKSN)

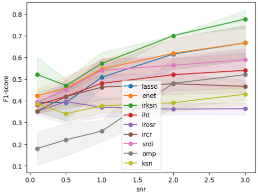
$$(B) : \max_{\ell \in \bar{S}} |\langle \boldsymbol{X}_S^{\dagger} \boldsymbol{x}_\ell, \frac{\boldsymbol{w}_S^*}{\min_{j \in S} |\boldsymbol{w}_S^*|} \rangle| < 1$$

It is possible to find examples of design matrix $\boldsymbol{X}$ and vector $\boldsymbol{w}^*$ which verify (B) but not (A): IRKSN is ensured to recover $\boldsymbol{w}^*$ there, contrary to ℓ_1 norm-based algorithms.

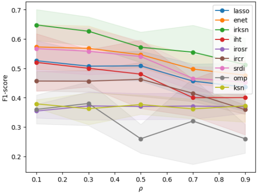
Experiments: Synthetic design matrix $\boldsymbol{X}$



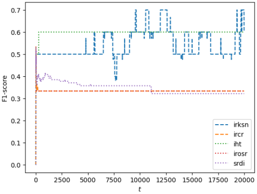
(a) F1-score vs. n



(b) F1-score vs. snr



(c) F1-score vs. ρ



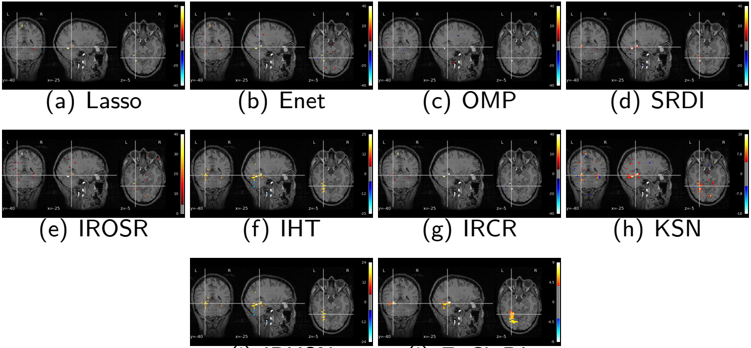
(d) F1-score vs. t

Figure: F1-score of support recovery for a correlated design matrix [6] ρ: correlation, snr: signal/noise ratio, n: num. samples.

Experiments: fMRI decoding

	Lasso	ElasticNet	OMP	IHT	KSN	IRKSN	IRCR	IROSR	SRDI
face'/'house'	.425	.349	.938	.2441	.247	**.2440**	.341	.381	.314
'house'/'shoe'	.528	.500	.938	.2968	.299	**.2965**	.407	.502	.357

Table: Model estimation $\|\boldsymbol{w} - \boldsymbol{w}^*\|$ ($\boldsymbol{w}^*$: obtained by EnCluDL).



(a) Lasso (b) Enet (c) OMP (d) SRDI
(e) IROSR (f) IHT (g) IRCR (h) KSN
(i) IRKSN (j) EnCluDL

Figure: Reconstructed functional region (EnCluDL $\sim$ ground-truth)

References I

Andreas Argyriou, Rina Foygel, and Nathan Srebro.
Sparse prediction with the k-support norm.
Advances in Neural Information Processing Systems, 25, 2012.

Thomas Blumensath and Mike E Davies.
Iterative hard thresholding for compressed sensing.
Applied and computational harmonic analysis, 27(3):265–274, 2009.

Gene H Golub and Charles F Van Loan.
Matrix computations.
JHU press, 2013.

Simon Matet, Lorenzo Rosasco, Silvia Villa, and Bang Long Vu.
Don't relax: early stopping for convex regularization.
arXiv preprint arXiv:1707.05422, 2017.

Andrew McDonald, Massimiliano Pontil, and Dimitris Stamos.
Fitting spectral decay with the k-support norm.
In *Artificial Intelligence and Statistics*, pages 1061–1069. PMLR, 2016.

Cesare Molinari, Mathurin Massias, Lorenzo Rosasco, and Silvia Villa.
Iterative regularization for convex regularizers.
In *International conference on artificial intelligence and statistics*, pages 1684–1692. PMLR, 2021.

Stanley Osher, Feng Ruan, Jiechao Xiong, Yuan Yao, and Wotao Yin.
Sparse recovery via differential inclusions.
Applied and Computational Harmonic Analysis, 41(2):436–469, 2016.

Robert Tibshirani.
Regression shrinkage and selection via the lasso.
Journal of the Royal Statistical Society: Series B (Methodological), 58(1):267–288, 1996.

Joel A Tropp and Anna C Gilbert.
Signal recovery from random measurements via orthogonal matching pursuit.
IEEE Transactions on information theory, 53(12):4655–4666, 2007.

Tomas Vaskevicius, Varun Kanade, and Patrick Rebeschini.
Implicit regularization for optimal sparse recovery.
Advances in Neural Information Processing Systems, 32, 2019.

Hui Zou and Trevor Hastie.
Regularization and variable selection via the elastic net.
Journal of the royal statistical society: series B (statistical methodology), 67(2):301–320, 2005.