Multi-parameter regularization for solving inverse problems of unmixing type:

theoretical and practical aspects

Valeriya Naumova

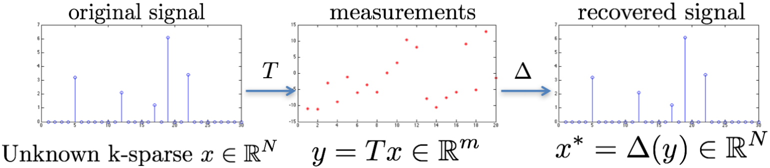
Simula Research Laboratory AS

Applied and computational analysis seminar,
University of Cambridge,

June 19, 2018

Introduction

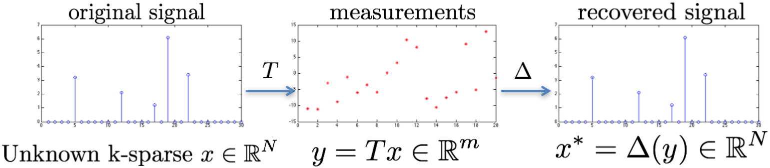
Compressed sensing



- known (linear) encoder $T \in \mathbb{R}^{m \times N}$ with RIP properties;
- unknown (non-linear) decoder Δ;
- Popular decoders:
  - $\Delta_0(y) = \text{argmin}_{Tz=y} \|z\|_0$; (non-convex, NP-hard)
  - $\Delta_1(y) = \text{argmin}_{Tz=y} \|z\|_1$. (convex)

Introduction

Compressed sensing

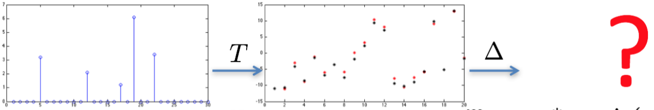


- known (linear) encoder $T \in \mathbb{R}^{m \times N}$ with RIP properties;
- unknown (non-linear) decoder Δ;
- Popular decoders:
  - $\Delta_0(y) = \text{argmin}_{Tz=y} \|z\|_0$; (non-convex, NP-hard)
  - $\Delta_1(y) = \text{argmin}_{Tz=y} \|z\|_1$. (convex)

Introduction

Compressed sensing with measurement noise

original signal | measurements | recovered signal



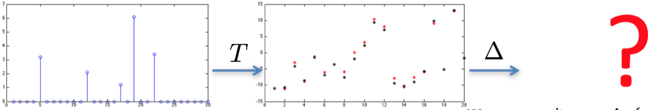
Unknown k-sparse $x \in \mathbb{R}^N$ | $y = Tx + \varepsilon \in \mathbb{R}^m$ | $x^* = \Delta(y) \in \mathbb{R}^N$

- Popular decoder:
  - $\Delta_{1,\delta}(y) = \operatorname{argmin}_{\|Tz-y\| \leqslant \delta} \|z\|_1$

Introduction

Compressed sensing with measurement noise

original signal | measurements | recovered signal



Unknown k-sparse $x \in \mathbb{R}^N$ | $y = Tx + \varepsilon \in \mathbb{R}^m$ | $x^* = \Delta(y) \in \mathbb{R}^N$

- Popular decoder:
  - $\Delta_{1,\delta}(y) = \mathrm{argmin}_{\|Tz-y\| \leqslant \delta} \|z\|_1$

    $\Longleftrightarrow$

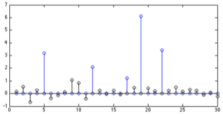
    $\Delta_{\ell_1} = \mathrm{argmin}\, \lambda\|z\|_1 + \|Tz - y\|_2^2$. (basis pursuit, ℓ_1—minimization, Lasso)
- Large amount of literature:

  Candes, Romberg, and Tao, IEEE Trans Inf Theory '06; Donoho, IEEE Trans Inf Theory, '06; Rauhut and Fourcart, Springer, '13.

Introduction

Compressed sensing with signal noise

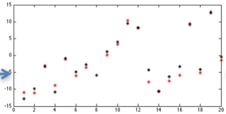
original signal



Unknown $x + n \in \mathbb{R}^N$

$\xrightarrow{T}$

measurements



$y = T(x+n) \in \mathbb{R}^m$

$\xrightarrow{\Delta}$

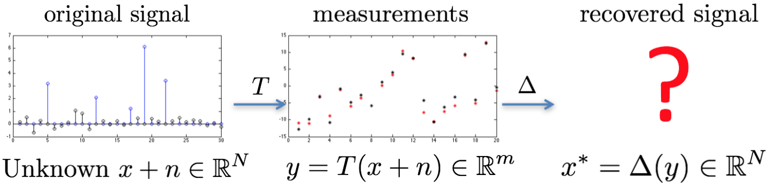
recovered signal

?

$x^* = \Delta(y) \in \mathbb{R}^N$

Introduction

Compressed sensing with signal noise

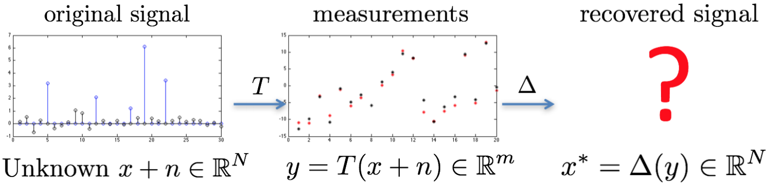


- General approach $y = T(x+n) = Tx + Tn$;
- Define $\varepsilon = Tn$ and consider it as noise on measurements;

Arias-Castro and Eldar, IEEE Signal Process Lett, ’11; Aeron, Saligrama, and Zhao, IEEE Trans Inf Theory, ’10.

Introduction

Compressed sensing with signal noise

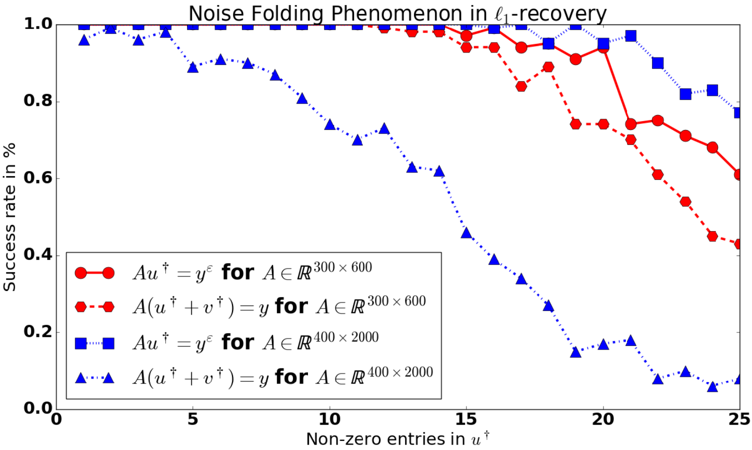


- General approach $y = T(x+n) = Tx + Tn$;
- Define $\varepsilon = Tn$ and consider it as noise on measurements;
- Noise-folding phenomenon, i.e., the variance of the noise on the original signal is amplified by a factor of $\frac{N}{m}$;
- Due to noise-folding phenomenon, e.g., the sole ℓ_1-regularization fails to accurately separate the signal from the noise.

  Arias-Castro and Eldar, IEEE Signal Process Lett, ’11; Aeron, Saligrama, and Zhao, IEEE Trans Inf Theory, ’10.

Introduction

Compressed sensing with signal / measurement noise



Inverse problems of the unmixing type

Unmixing problem: restore and separate two (or more) components $u^\dagger$ **and** $v^\dagger$ of the solution from an **observed datum** y where

$$y = T(u^\dagger + \underbrace{v^\dagger}_{\text{signal noise}}) + \underbrace{\varepsilon}_{\text{noise}},$$

Ill-posedness: an infinite number of solutions, operator T may have non-closed range.

Inverse problems of the unmixing type

Unmixing problem: restore and separate two (or more) components $u^\dagger$ **and** $v^\dagger$ of the solution from an **observed datum** y where

$$y = T(u^\dagger + \underbrace{v^\dagger}_{\text{signal noise}}) + \underbrace{\varepsilon}_{\text{noise}},$$

Ill-posedness: an infinite number of solutions, operator T may have non-closed range.

Assumption: $u^\dagger \in \ell_p$ and $v^\dagger \in \ell_2 = \ell_q \cap \ell_2$ for $0 \leqslant p < 2$ and $2 \leqslant q < \infty$.

Regularization functional:

$$J_{p,q}(u,v) := \underbrace{\|T(u+v) - y\|_2^2}_{\text{data fitting}} + \underbrace{R(u,v)}_{\text{regularization}} \to \min_{u,v}$$

- $R(u,v) = \lambda_1 \|u+v\|_{\ell_1}$

Inverse problems of the unmixing type

Unmixing problem: restore and separate two (or more) components $u^\dagger$ **and** $v^\dagger$ of the solution from an **observed datum** y where

$$y = T(u^\dagger + \underbrace{v^\dagger}_{\text{signal noise}}) + \underbrace{\varepsilon}_{\text{noise}},$$

Ill-posedness: an infinite number of solutions, operator T may have non-closed range.

Assumption: $u^\dagger \in \ell_p$ and $v^\dagger \in \ell_2 = \ell_q \cap \ell_2$ for $0 \leqslant p < 2$ and $2 \leqslant q < \infty$.

Regularization functional:

$$J_{p,q}(u,v) := \underbrace{\|T(u+v) - y\|_2^2}_{\text{data fitting}} + \underbrace{R(u,v)}_{\text{regularization}} \to \min_{u,v}$$

- $R(u,v) = \lambda_1 \|u+v\|_{\ell_1}$ (ℓ_1 fails to accurately separate the signal from the noise).

Inverse problems of the unmixing type

Unmixing problem: restore and separate two (or more) components $u^\dagger$ **and** $v^\dagger$ of the solution from an **observed datum** y where

$$y = T(u^\dagger + \underbrace{v^\dagger}_{\text{signal noise}}) + \underbrace{\varepsilon}_{\text{noise}},$$

Ill-posedness: an infinite number of solutions, operator T may have non-closed range.

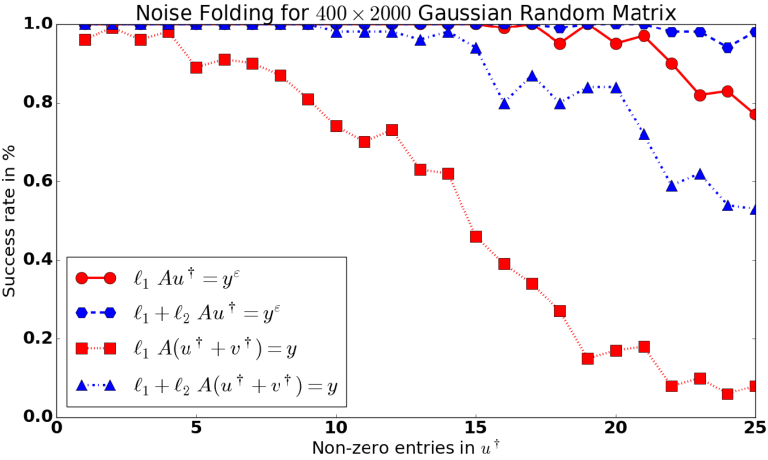
Assumption: $u^\dagger \in \ell_p$ and $v^\dagger \in \ell_2 = \ell_q \cap \ell_2$ for $0 \leqslant p < 2$ and $2 \leqslant q < \infty$.

Regularization functional:

$$J_{p,q}(u,v) := \underbrace{\|T(u+v) - y\|_2^2}_{\text{data fitting}} + \underbrace{R(u,v)}_{\text{regularization}} \to \min_{u,v}$$

- $R(u,v) = \lambda_1 \|u+v\|_{\ell_1}$ (ℓ_1 fails to accurately separate the signal from the noise).
- $R(u,v) = \lambda_1 \|u\|_{\ell_p}^p + \lambda_2 \|v\|_{\ell_q}^q$ (multi-penalty regularization)
  - $\lambda_1, \lambda_2 \in \mathbb{R}_+$ are regularization parameters;
  - $\lambda_1, \lambda_2, p, q$ are unknown.

Does multi-penalty regularization really work?



Problems on the way...

- **Iterative alternating algorithm** to perform minimization of $J_{p,q}$;
  Problem: non-convexity of the functional for $0 < p < 1$;
  Way-out: adaptation of several techniques on a single-parameter regularization with sparsity promoting terms.

Problems on the way...

- **Iterative alternating algorithm** to perform minimization of $J_{p,q}$;

  Problem: non-convexity of the functional for $0 < p < 1$;

  Way-out: adaptation of several techniques on a single-parameter regularization with sparsity promoting terms.

- **Theoretical guarantees on the support recovery by means of multi-penalty regularization**

  Problem: non-linearity with respect to the parameters;

  Way-out: generalization and extension of the results from single-penalty regularization.

Problems on the way...

- **Iterative alternating algorithm** to perform minimization of $J_{p,q}$;

  Problem: non-convexity of the functional for $0 < p < 1$;

  Way-out: adaptation of several techniques on a single-parameter regularization with sparsity promoting terms.

- **Theoretical guarantees on the support recovery by means of multi-penalty regularization**

  Problem: non-linearity with respect to the parameters;

  Way-out: generalization and extension of the results from single-penalty regularization.

- **Adaptive choice of the regularization parameters for optimal support recovery.**

  Problem: how to choose multiple parameters and which ones allow the best support reconstruction;

  Way-out: extension of Lasso-path for multi-penalty, statistical learning theory.

Multi-parameter regularization

Some contributions

- Image processing: Meyer ’02; Vese and Osher ’03, ’04; Daubechies and Teschke ’05; Bredies and Holler ’14; Holler and Kunisch ’14; De Los Reyes, Schönlieb, and Valkonen ’15; Calatroni, De Los Reyes, and Schönlieb ’16.
- Signal processing: Donoho et al. ’89, ’01, ’13.
- Geomathematics: Lu and Pereverzyev ’11.
- Regularization and learning theory: Lu and Pereverzyev ’11; VN, Pereverzyev ’13; Fornasier, VN, and Pereverzyev ’13; Sivananthan ’16.
  - Huber regularization: Huber ’64; Beck and Teboulle ’12; Zadorozhnyi et al., ’16.

Alternating minimization algorithm

Joint work with Steffen Peter and Massimo Fornasier, TU Munich

Peter and VN, Inverse Problems, '14.

Alternating minimization algorithm

Problem formation and state of the art

- We are interested in designing an algorithm for minimization of the non-convex non-smooth functional

$$J_{p,q}(u,v) := \|T(u+v) - y\|_2^2 + \lambda_1 \|u\|_{\ell_p}^p + \left(\lambda_2 \|v\|_{\ell_q}^q + \varepsilon \|v\|_{\ell_2}^2\right),$$

where $\lambda_1, \lambda_2, \varepsilon \in \mathbb{R}_+$, $0 \leqslant p < 2$, $2 \leqslant q < \infty$ are parameters of the problem.

Alternating minimization algorithm

Problem formation and state of the art

- We are interested in designing an algorithm for minimization of the non-convex non-smooth functional

$$J_{p,q}(u,v) := \|T(u+v)-y\|_2^2 + \lambda_1\|u\|_{\ell_p}^p + \left(\lambda_2\|v\|_{\ell_q}^q + \varepsilon\|v\|_{\ell_2}^2\right),$$

  where $\lambda_1, \lambda_2, \varepsilon \in \mathbb{R}_+$, $0 \leqslant p < 2$, $2 \leqslant q < \infty$ are parameters of the problem.
- Regularization with non-convex constraints for single-parameter regularization:
  - Bredies and Lorenz '09; Ramlau and Zarzer '12; Lu et al. '13; Ito and Kunisch '14.

Alternating minimization algorithm

Problem formation and state of the art

- We are interested in designing an algorithm for minimization of the non-convex non-smooth functional

$$J_{p,q}(u,v) := \|T(u+v)-y\|_2^2 + \lambda_1\|u\|_{\ell_p}^p + \left(\lambda_2\|v\|_{\ell_q}^q + \varepsilon\|v\|_{\ell_2}^2\right),$$

  where $\lambda_1, \lambda_2, \varepsilon \in \mathbb{R}_+$, $0 \leqslant p < 2$, $2 \leqslant q < \infty$ are parameters of the problem.
- Regularization with non-convex constraints for single-parameter regularization:
  - Bredies and Lorenz '09; Ramlau and Zarzer '12; Lu et al. '13; Ito and Kunisch '14.

The first work that provides an **explicit direct mechanism for minimization of the multi-penalty functional with non-convex and non-smooth terms,** and highlights its improved accuracy power with respect to more traditional one-parameter regularizations.

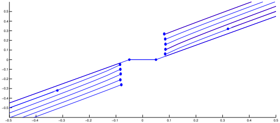
Alternating minimization algorithm

based on iterative thresholding

set $\varepsilon > 0$, $VTOL > 0$.
set $u^{(0)} = u^{(1,0)} = 0$ and $v^{(0)} = v^{(1,0)} = 0$.
repeat
  $u^{(n)} = u^{(n,L)} = u^{(n+1,0)}$
  for $l = 1$ to $L-1$ **do**
    $u_\lambda^{(n+1,l+1)} = H^p_{\lambda_1}(u_\lambda^{(n+1,l)} + [T^*(y - Tv^{(n,M)} - Tu^{(n+1,l)})]_\lambda)$.
  end for
  $v^{(n)} = v^{(n,M)} = v^{(n+1,0)}$
  for $l = 1$ to $M-1$ **do**
    $v_\lambda^{(n+1,l+1)} = S^q_{\lambda_2,\varepsilon}(v_\lambda^{(n+1,l)} + [T^*(y - Tu^{(n+1)} - Tv^{(n+1,l)})]_\lambda)$
  end for
until $\|u^{(n)} - u^{(n+1)}\|_{\ell_1} > VTOL$

$S^q_{\lambda_2,\varepsilon}, H^p_{\lambda_1}$ are the thresholding functions.

Alternating minimization algorithm



The thresholding function $H^p_{\lambda_1}$ for $p = 0, 0.15, 0.3, 0.45, 0.6, 0.9, 1$ and $\lambda_1 = 0.1$.

$$H^p_{\lambda_1}(x) = \begin{cases} 0, & |x| \leqslant \tau_{\lambda_1}, \\ (F^p_{\lambda_1})^{-1}(x), & |x| \geqslant \tau_{\lambda_1}, \end{cases}$$

where

$F^p_{\lambda_1}(t) = t + \frac{\lambda_1 p}{2} \operatorname{sgn}(t)|t|^{p-1}$ and

$\tau_{\lambda_1} = \frac{2-p}{2-2p}(\lambda_1(1-p))^{1/(2-p)}$.

Theoretical results

Theorem (Weak Convergence)

Assume $0 < p < 1$ *and* $2 \leqslant q < \infty$. *The algorithm produces sequences* $(u^{(n)}),\ (v^{(n)})$ *in* ℓ_2 *whose weak accumulation points are fixed points of the algorithm.*

Theoretical results

Theorem (Weak Convergence)

Assume $0 < p < 1$ and $2 \leqslant q < \infty$. The algorithm produces sequences $(u^{(n)}),\ (v^{(n)})$ in ℓ_2 whose weak accumulation points are fixed points of the algorithm.

Theorem (Minimizers)

Let T have the FBI property, i.e., T is injective whenever restricted to finitely many coefficients. Then we have the following inclusion

$$\mathcal{F}ix \subset \mathcal{L},$$

where $\mathcal{F}ix$ is the set of fixed points, $\mathcal{L}$ is the set of local minimizers of $J_{p,q}$.

Theoretical results

Theorem (Weak Convergence)

Assume $0 < p < 1$ *and* $2 \leqslant q < \infty$. *The algorithm produces sequences* $(u^{(n)})$, $(v^{(n)})$ *in* ℓ_2 *whose weak accumulation points are fixed points of the algorithm.*

Theorem (Minimizers)

Let T *have the FBI property, i.e.,* T *is injective whenever restricted to finitely many coefficients. Then we have the following inclusion*

$$\mathcal{F}ix \subset \mathcal{L},$$

where $\mathcal{F}ix$ *is the set of fixed points,* $\mathcal{L}$ *is the set of local minimizers of* $J_{p,q}$.

Theorem (Strong Convergence)

The algorithm produces sequences $(u^{(n)})$ *and* $(v^{(n)})$ *in* ℓ_2 *that converge strongly to the vectors* $u^*, v^* \in \mathcal{F}ix$ *respectively.*

Numerical experiments

The model problem

$$y = T(u^\dagger + v^\dagger),$$

where $T \in \mathbb{R}^{m \times N}$ is an i.i.d Gaussian matrix, $u^\dagger$ is a sparse vector and $v^\dagger$ is a noise vector.

Numerical experiments

The model problem

$$y = T(u^\dagger + v^\dagger),$$

where $T \in \mathbb{R}^{m \times N}$ is an i.i.d Gaussian matrix, $u^\dagger$ is a sparse vector and $v^\dagger$ is a noise vector.

Data:

- 20 problems, $m = 40, \ N = 100$;

Numerical experiments

The model problem

$$y = T(u^\dagger + v^\dagger),$$

where $T \in \mathbb{R}^{m \times N}$ is an i.i.d Gaussian matrix, $u^\dagger$ is a sparse vector and $v^\dagger$ is a noise vector.

Data:

- 20 problems, $m = 40,\ N = 100$;
- $u^\dagger$ and $v^\dagger$ are randomly generated;
- $\#\,\mathrm{supp}(u^\dagger) = 7$;
- $\|v^\dagger\|_2 = 0.7$;

Numerical experiments

The model problem

$$y = T(u^\dagger + v^\dagger),$$

where $T \in \mathbb{R}^{m\times N}$ is an i.i.d Gaussian matrix, $u^\dagger$ is a sparse vector and $v^\dagger$ is a noise vector.

Data:

- 20 problems, $m = 40,\ N = 100$;
- $u^\dagger$ and $v^\dagger$ are randomly generated;
- $\#\,\mathrm{supp}(u^\dagger) = 7$;
- $\|v^\dagger\|_2 = 0.7$;
- $p \in \{0, 0.3, 0.5, 0.8, 1\}$ and $q \in \{2, 4, 10, \infty\}$;
- $L = M = 20,\ u^{(0)} = v^{(0)} = 0$.

Comparison with the single-parameter counterpart

One-parameter regularization:

$$J_p(u) := \|Tu - y\|_2^2 + \lambda_1 \|u\|_p^p,$$

with $p \in [0, 1]$.

A local minimizer $u^*_{\lambda_1,p}$ can be computed by the iterations

$$u^{(n+1)}_{\lambda_1,p} = H^p_{\lambda_1}(u^{(n)}_\lambda + [T^*(y - Tu^{(n)})]_\lambda), \quad n \geqslant 0,$$

where $H^p_{\lambda_1}$ is the thresholding operator.

Comparison with the single-parameter counterpart

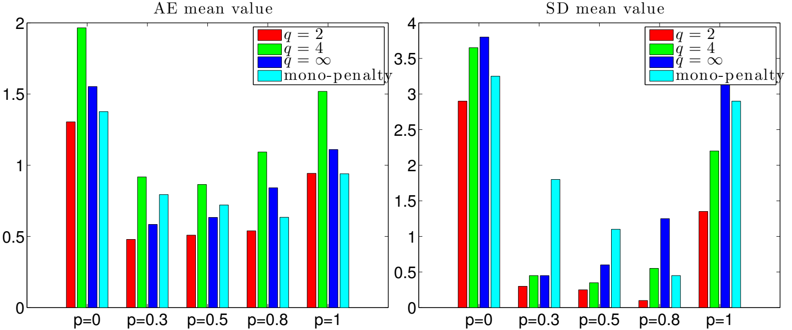


Figure: For each $p \in \{0, 0.3, 0.5, 0.8, 1\}$ the mean of the AE (left) and SD (right) for the solution $u^\dagger$ for 20 problems for different parameter values $q \in \{2, 4, \infty\}$ and $u^\dagger_{\lambda_1,p}$. For each of the 20 problems and each pair (p, q), the best individual parameter pair (λ_1, λ_2) was chosen for comparison

Comparison with the single-parameter counterpart

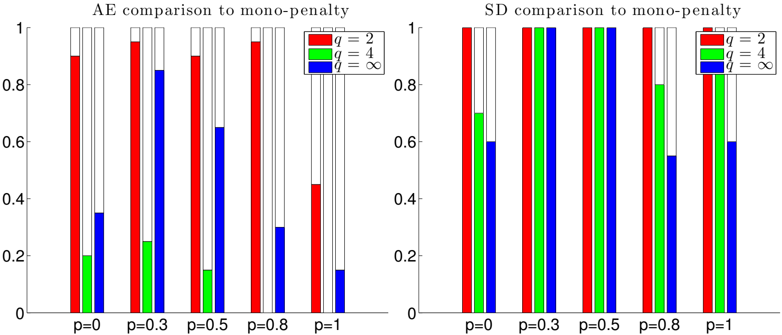


Figure: A coloured bar indicates the empirical probability of better performance by the multi-penalty approach in terms of AE (left) and SD (right) with respect to the mono-penalty approach

Conditions on optimal support recovery

Joint work with Markus Grasmair, NTNU

Grasmair and VN, Inverse Problems, ’16.

Conditions on optimal support recovery

- We are interested in deriving theoretical results for multi-penalty Tikhonov regularization of the form

$$J(u,v) = \frac{1}{2}\|T(u+v) - y^\delta\|_2^2 + \lambda_1\|u\|_{\ell_1} + \frac{\lambda_2}{2}\|v\|_{\ell_2}^2,$$

- Conditions on convergence of sparsity-promoting regularization:
  - Grasmair, Sherzer, and Haltmeier, '11; Bredies and Holler, '14; Lu et al., '13.

Conditions on optimal support recovery

- We are interested in deriving theoretical results for multi-penalty Tikhonov regularization of the form

$$J(u,v) = \frac{1}{2}\|T(u+v) - y^\delta\|_2^2 + \lambda_1 \|u\|_{\ell_1} + \frac{\lambda_2}{2}\|v\|_{\ell_2}^2,$$

- Conditions on convergence of sparsity-promoting regularization:
  - Grasmair, Sherzer, and Haltmeier, '11; Bredies and Holler, '14; Lu et al., '13.

The first work that provides a **theoretical analysis of the multi-penalty regularization with a non-smooth sparsity promoting regularization term**, and an explicit comparison with the single-parameter counterpart.

Conditions on optimal support recovery

We consider the multi-penalty Tikhonov regularization of the form

$$J(u,v) = \frac{1}{2}\|T(u+v) - y^{\delta}\|_2^2 + \lambda_1\|u\|_{\ell_1} + \frac{\lambda_2}{2}\|v\|_{\ell_2}^2, \qquad (1)$$

and signals belonging to the class $S_{c,d,k} = \cup_{\#I<k} S_{c,d,I}$ with

$$S_{c,d,I} = \{(u,v) \in \mathbb{R}^N \times \mathbb{R}^N : \mathrm{supp}(u) = I, \inf_{i \in I} u_i > c, \|v\|_\infty < d\}$$

Conditions on optimal support recovery

We consider the multi-penalty Tikhonov regularization of the form

$$J(u, v) = \frac{1}{2}\|T(u+v) - y^\delta\|_2^2 + \lambda_1\|u\|_{\ell_1} + \frac{\lambda_2}{2}\|v\|_{\ell_2}^2, \qquad (1)$$

and signals belonging to the class $S_{c,d,k} = \cup_{\#I<k} S_{c,d,I}$ with

$$S_{c,d,I} = \{(u, v) \in \mathbb{R}^N \times \mathbb{R}^N : \mathrm{supp}(u) = I, \inf_{i \in I} u_i > c, \|v\|_\infty < d\}$$

Definition

Let $\lambda_2 \in \mathbb{R}_+ \cup \{\infty\}$ be fixed. We call $S_{c,d,I}$ a set of exact support recovery if there exists $\lambda_1 > 0$ such that the solution $(u_{\lambda_1,\lambda_2}, v_{\lambda_1,\lambda_2})$ of (1) satisfies $\mathrm{supp}(u_{\lambda_1,\lambda_2}) = \mathrm{supp}(u^\dagger)$ whenever the given data y has the form $y = T(u^\dagger + v^\dagger)$ with $(u^\dagger, v^\dagger) \in S_{c,d,I}$.

The parameters $\lambda_1 > 0$ for which this property holds are called admissible for $S_{c,d,I}$.

Multi-penalty $\Rightarrow$ single-penalty regularization

We would like to address the following two fundamental questions:

- Could multi-penalty regularization allow for the exact recovery of the support of the true solution $u^\dagger$?
- How is ”theoretical performance” of the multi-penalty regularization compared to the mono-penalty one?

Multi-penalty $\Rightarrow$ single-penalty regularization

Lemma

The pair $(u_{\lambda_1,\lambda_2}, v_{\lambda_1,\lambda_2})$ *solves (1) if and only if*

$$v_{\lambda_1,\lambda_2} = (\lambda_2 \mathbb{I} + T^* T)^{-1}(T^* y - T^* T u_{\lambda_1,\lambda_2})$$

and u_{λ_1,λ_2} *solves the optimization problem*

$$\frac{1}{2}\|B_{\lambda_2} u - y_{\lambda_2}\|_2^2 + \lambda_1 \|u\|_1 \to \min$$

with

$$B_{\lambda_2} = \left(\mathbb{I} + \frac{TT^*}{\lambda_2}\right)^{-1/2} T \quad \text{and} \quad y_{\lambda_2} = \left(\mathbb{I} + \frac{TT^*}{\lambda_2}\right)^{-1/2} y.$$

Multi-penalty $\Rightarrow$ single-penalty regularization

Lemma

The pair $(u_{\lambda_1,\lambda_2}, v_{\lambda_1,\lambda_2})$ *solves (1) if and only if*

$$v_{\lambda_1,\lambda_2} = (\lambda_2 \mathbb{I} + T^* T)^{-1} (T^* y - T^* T u_{\lambda_1,\lambda_2})$$

and u_{λ_1,λ_2} *solves the optimization problem*

$$\frac{1}{2} \|B_{\lambda_2} u - y_{\lambda_2}\|_2^2 + \lambda_1 \|u\|_1 \to \min$$

with

$$B_{\lambda_2} = \left(\mathbb{I} + \frac{TT^*}{\lambda_2}\right)^{-1/2} T \quad \textit{and} \quad y_{\lambda_2} = \left(\mathbb{I} + \frac{TT^*}{\lambda_2}\right)^{-1/2} y.$$

Note: The theory of ℓ^1-regularization works for the multi-penalty setting: for fixed $\lambda_2 > 0$ under a source conditions, we get

$$\|u^\dagger - u_{\lambda_1,\lambda_2}\|_1 \leqslant C_{1,\lambda_2} \lambda_1 + C_{2,\lambda_2} \frac{\|y_{\lambda_2} - B_{\lambda_2} u^\dagger\|_2^2}{\lambda_1}$$

Optimality conditions for single-penalty

Lemma

The vector u_{λ_1} minimizes

$$T(u) := \frac{1}{2}\|Tu - y\|_2^2 + \lambda_1\|u\|_1$$

Optimality conditions for single-penalty

Lemma

The vector u_{λ_1} *minimizes*

$$T(u) := \frac{1}{2}\|Tu - y\|_2^2 + \lambda_1\|u\|_1 \iff T^*(Tu_{\lambda_1} - y) \in -\lambda_1 \operatorname{sgn}(u_{\lambda_1}).$$

Optimality conditions for single-penalty

Lemma

The vector u_{λ_1} *minimizes*

$$T(u) := \frac{1}{2}\|Tu - y\|_2^2 + \lambda_1\|u\|_1 \iff T^*(Tu_{\lambda_1} - y) \in -\lambda_1 \operatorname{sgn}(u_{\lambda_1}).$$

Lemma

We have $\operatorname{supp}(u_{\lambda_1}) = I \iff \exists\, w_{\lambda_1} \in (\mathbb{R}\setminus\{0\})^I$ *such that*

$$T_I^*(T_I w_{\lambda_1} - y) = -\lambda_1 \operatorname{sgn}(w_{\lambda_1}) \quad \text{and} \quad \|T_J^*(T_I w_{\lambda_1} - y)\|_\infty \leqslant \lambda_1,$$

where $J := \{i : u_i^\dagger = 0\}$ *and* T_I *is the restriction of* T *to the span of the support of* $u^\dagger$.

Optimality conditions for single-penalty

Proposition

Assume that T_I is injective and that

$$\| T_J^* T_I (T_I^* T_I)^{-1} \|_\infty < 1. \tag{2}$$

Then the set $S_{c,d,I}$ is a set of exact support recovery whenever

$$\frac{c}{d} > \frac{\| T_J^* (T_I (T_I^* T_I)^{-1} T_I^* - \mathbb{I}) T \|_\infty \| (T_I^* T_I)^{-1} \|_\infty}{1 - \| T_J^* T_I (T_I^* T_I)^{-1} \|_\infty} + \| (T_I^* T_I)^{-1} T_I^* T \|_\infty .$$

Moreover, every parameter $\lambda_1 > 0$ satisfying

$$\frac{d \| T_J^* (T_I (T_I^* T_I)^{-1} T_I^* - \mathbb{I}) T \|_\infty}{1 - \| T_J^* T_I (T_I^* T_I)^{-1} \|_\infty} \leqslant \lambda_1 < \frac{c - d \| (T_I^* T_I)^{-1} T_I^* T \|_\infty}{\| (T_I^* T_I)^{-1} \|_\infty}$$

is admissible on $S_{c,d,I}$.

Optimality conditions for multi-penalty regularization

Lemma

We have $\operatorname{supp}(u_{\lambda_1,\lambda_2}) = I \iff \exists w_{\lambda_1} \in (\mathbb{R}\setminus\{0\})^I$ *such that*

$$T^*_{\lambda_2,I}(T_I w_{\lambda_1,\lambda_2} - y) = -\lambda_1 \operatorname{sgn}(w_{\lambda_1,\lambda_2}) \quad \textit{and} \quad \| T^*_{\lambda_2,J}(T_I w_{\lambda_1} - y)\|_\infty \leqslant \lambda_1,$$

where $T_{\lambda_2} := \left(\mathbb{I} + \frac{TT^*}{\lambda_2}\right)^{-1} T.$

Optimality conditions for multi-penalty regularization

Lemma

We have $\operatorname{supp}(u_{\lambda_1,\lambda_2}) = I \iff \exists w_{\lambda_1} \in (\mathbb{R} \setminus \{0\})^I$ *such that*

$$T^*_{\lambda_2,I}(T_I w_{\lambda_1,\lambda_2} - y) = -\lambda_1 \operatorname{sgn}(w_{\lambda_1,\lambda_2}) \quad \textit{and} \quad \|T^*_{\lambda_2,J}(T_I w_{\lambda_1} - y)\|_\infty \leqslant \lambda_1,$$

where $T_{\lambda_2} := \left(\mathbb{I} + \frac{TT^*}{\lambda_2}\right)^{-1} T$.

Sketch of the proof.

Using the fact that

$$B^*_{\lambda_2} B_{\lambda_2} = T^*_{\lambda_2} T \quad \text{and} \quad B^*_{\lambda_2} y_{\lambda_2} = T^*_{\lambda_2} y$$

with $B_{\lambda_2} = \left(\mathbb{I} + \frac{TT^*}{\lambda_2}\right)^{-1/2} T$, and the lemma for single-penalty case. $\square$

Optimality conditions for multi-penalty regularization

Proposition

Assume that $0 < \lambda_2 < \infty$ *is such that*

$$\| T^*_{\lambda_2,J} T_I (T^*_{\lambda_2,I} T_I)^{-1} \|_\infty < 1. \tag{3}$$

Then the set $S_{c,d,I}$ *is a set of exact support recovery in the multi-penalty setting whenever*

$$\frac{c}{d} > \|(T^*_{\lambda_2,I} T_I)^{-1} T^*_{\lambda_2,I} T\|_\infty + \frac{\| T^*_{\lambda_2,J}(T_I(T^*_{\lambda_2,I} T_I)^{-1} T^*_{\lambda_2,I} - \mathbb{I}) T\|_\infty \|(T^*_{\lambda_2,I} T_I)^{-1}\|_\infty}{1 - \| T^*_{\lambda_2,J} T_I (T^*_{\lambda_2,I} T_I)^{-1}\|_\infty}.$$

Moreover, all pairs of parameters (λ_1, λ_2) *satisfying above and*

$$\frac{d\| T^*_{\lambda_2,J}(T_I(T^*_{\lambda_2,I} T_I)^{-1} T^*_{\lambda_2,I} - \mathbb{I}) T\|_\infty}{1 - \| T^*_{\lambda_2,J} T_I (T^*_{\lambda_2,I} T_I)^{-1}\|_\infty} \leqslant \lambda_1 < \frac{c - d\|(T^*_{\lambda_2,I} T_I)^{-1} T^*_{\lambda_2,I} T\|_\infty}{\|(T^*_{\lambda_2,I} T_I)^{-1}\|_\infty}$$

are admissible on $S_{c,d,I}$.

Optimality conditions

Setup:

- 20 i.i.d. Gaussian matrices, $m = 30$, $N = 60$ and $m = 40$, $N = 80$.
- $\#\,\mathrm{supp}(u^\dagger) = 2, 3, 4$.
- Check conditions $\| T_J^* T_I (T_I^* T_I)^{-1} \|_\infty < 1$ or $\| T_{\lambda_2,J}^* T_I (T_{\lambda_2,I}^* T_I)^{-1} \|_\infty < 1$.

$m = 30$, $N = 60$	Mono-penalty	Multi-penalty $\lambda_2 = 10$	Multi-penalty $\lambda_2 = 1$	Multi-penalty $\lambda_2 = 0.1$
Median	0.5425	0.3814	0.1214	0.0623
Mean	0.5559	0.3922	0.1225	0.0635
SD	0.05652	0.04142	0.01518	0.01083

$m = 40$, $N = 80$	Mono-penalty	Multi-penalty $\lambda_2 = 10$	Multi-penalty $\lambda_2 = 1$	Multi-penalty $\lambda_2 = 0.1$
Median	0.2696	0.1523	0.0396	0.0256
Mean	0.2746	0.1547	0.0413	0.0262
SD	0.03060	0.01848	0.00659	0.00447

Table: Percentage of 3-sparse subspaces for which (2) or (3) failed.

Optimality conditions

Setup:

- 20 i.i.d. Gaussian matrices, $m = 30$, $N = 60$ and $m = 40$, $N = 80$.
- $\#\,\mathrm{supp}(u^\dagger) = 2, 3, 4$.
- Check conditions $\|T_J^* T_I (T_I^* T_I)^{-1}\|_\infty < 1$ or $\|T_{\lambda_2,J}^* T_I (T_{\lambda_2,I}^* T_I)^{-1}\|_\infty < 1$.

$m = 30$ $N = 60$	Mono-penalty	Multi-penalty $\lambda_2 = 10$	 $\lambda_2 = 1$	 $\lambda_2 = 0.1$
Median	0.5425	0.3814	0.1214	0.0623
Mean	0.5559	0.3922	0.1225	0.0635
SD	0.05652	0.04142	0.01518	0.01083

$m = 40$ $N = 80$	Mono-penalty	Multi-penalty $\lambda_2 = 10$	 $\lambda_2 = 1$	 $\lambda_2 = 0.1$
Median	0.2696	0.1523	0.0396	0.0256
Mean	0.2746	0.1547	0.0413	0.0262
SD	0.03060	0.01848	0.00659	0.00447

Table: Percentage of 3-sparse subspaces for which (2) or (3) failed.

Adaptive parameter choice for multi-penalty regularization

Joint work with Markus Grasmair, NTNU, and Timo Klock, Simula Research Lab

Grasmair, Klock, and VN, submitted, ’17.

Adaptive parameter choice for multi-penalty regularization

- We are interested in designing a rule for adaptive choice of the regularization parameters for multi-penalty regularization of the form

$$J(u, v) = \frac{1}{2}\|T(u+v) - y^\delta\|_2^2 + \lambda_1\|u\|_{\ell_1} + \frac{\lambda_2}{2}\|v\|_{\ell_2}^2,$$

- Parameter choice for ℓ_1- regularization:
  - Rosset and Zhu, '07; Efron, Hastie, Johnstone, and Tibshirani, '04; Tibshirani, '13; Jua and Rohe, '15.

Adaptive parameter choice for multi-penalty regularization

- We are interested in designing a rule for adaptive choice of the regularization parameters for multi-penalty regularization of the form

$$J(u,v) = \frac{1}{2}\|T(u+v) - y^\delta\|_2^2 + \lambda_1\|u\|_{\ell_1} + \frac{\lambda_2}{2}\|v\|_{\ell_2}^2,$$

- Parameter choice for ℓ_1- regularization:
  - Rosset and Zhu, '07; Efron, Hastie, Johnstone, and Tibshirani, '04; Tibshirani, '13; Jua and Rohe, '15.

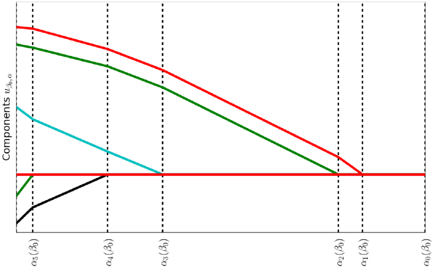
The first work that provides an **efficient algorithm for identification of possible parameter regions leading to structurally similar solutions** from multi-penalty regularization: solutions with the same support and sign pattern.

Lasso path for single-penalty regularization

$$u_{\lambda_1,\lambda_2^0} = \underset{u}{\operatorname{argmin}} \frac{1}{2}\|B_{\lambda_2^0}u - y_{\lambda_2^0}\|^2 + \lambda_1\|u\|_1.$$

- Lasso solutions are piece-wise linear and can be computed successively.

  $\Longrightarrow$ Only nodes $\lambda_1^i(\lambda_2^0)$ need to be computed.



Extension of the Lasso path

Multi-penalty framework

- The nodes are calculated by inductive verification of the KKT conditions:
  - Find $\lambda_1^{i+1}(\lambda_2^0)$ from $\lambda_1^i(\lambda_2^0)$, I, σ via iterative verification of KKT conditions, here I **is a support** and σ **is a sign pattern.**
  - Entries are in the support as long as

$$|B^T_{\lambda_2^0,j} \cdot \mathrm{Residual}(B_{\lambda_2^0}, y_{\lambda_2^0}, \lambda_1)| = \lambda_1.$$

Extension of the Lasso path

Multi-penalty framework

- The nodes are calculated by inductive verification of the KKT conditions:
  - Find $\lambda_1^{i+1}(\lambda_2^0)$ from $\lambda_1^i(\lambda_2^0)$, I, σ via iterative verification of KKT conditions, here I **is a support** and σ **is a sign pattern.**
  - Entries are in the support as long as

$$|B^T_{\lambda_2^0,j} \cdot \mathrm{Residual}(B_{\lambda_2^0}, y_{\lambda_2^0}, \lambda_1)| = \lambda_1.$$

- Candidates for new nodes:

$$\tilde{\lambda}_1^j(I,\sigma,\lambda_2^0) := \begin{cases} \dfrac{B^T_{\lambda_2^0,j}(\mathbb{I} - B_{\lambda_2^0,I}(B^T_{\lambda_2^0,I}B_{\lambda_2^0,I})^{-1}B^T_{\lambda_2^0,I})y_{\lambda_2^0}}{\pm 1 - B^T_{\lambda_2^0,j}B_{\lambda_2^0,I}(B^T_{\lambda_2^0,I}B_{\lambda_2^0,I})^{-1}\sigma} & \text{if } j \notin I \\[2ex] \dfrac{\left((B^T_{\lambda_2^0,I}B_{\lambda_2^0,I})^{-1}B^T_{\lambda_2^0,I}y_{\lambda_2^0}\right)_j}{\left((B^T_{\lambda_2^0,I}B_{\lambda_2^0,I})^{-1}\sigma\right)_j} & \text{if } j \in I. \end{cases}$$

Extension of the Lasso path

Multi-penalty framework

- The nodes are calculated by inductive verification of the KKT conditions:
  - Find $\lambda_1^{i+1}(\lambda_2^0)$ from $\lambda_1^i(\lambda_2^0)$, I, σ via iterative verification of KKT conditions, here I **is a support** and σ **is a sign pattern.**
  - Entries are in the support as long as

$$|B^T_{\lambda_2^0,j} \cdot \mathrm{Residual}(B_{\lambda_2^0}, y_{\lambda_2^0}, \lambda_1)| = \lambda_1.$$

- Candidates for new nodes:

$$\tilde{\lambda}_1^j(I,\sigma,\lambda_2^0) := \begin{cases} \dfrac{B^T_{\lambda_2^0,j}(\mathbb{I}-B_{\lambda_2^0,I}(B^T_{\lambda_2^0,I}B_{\lambda_2^0,I})^{-1}B^T_{\lambda_2^0,I})y_{\lambda_2^0}}{\pm 1-B^T_{\lambda_2^0,j}B_{\lambda_2^0,I}(B^T_{\lambda_2^0,I}B_{\lambda_2^0,I})^{-1}\sigma} & \text{if } j \notin I \\ \dfrac{\left((B^T_{\lambda_2^0,I}B_{\lambda_2^0,I})^{-1}B^T_{\lambda_2^0,I}y_{\lambda_2^0}\right)_j}{\left((B^T_{\lambda_2^0,I}B_{\lambda_2^0,I})^{-1}\sigma\right)_j} & \text{if } j \in I. \end{cases}$$

- Choose

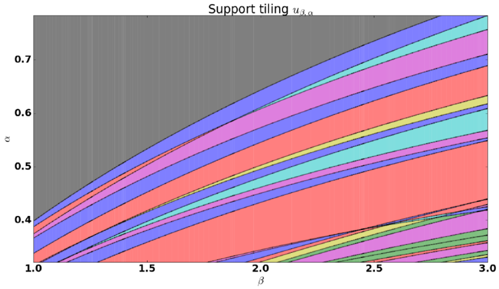
$$\lambda_1^{i+1} = \max_j\{\tilde{\lambda}_1^j(I,\sigma,\lambda_2^0) < \lambda_i(\lambda_2^0)\}$$

and recompute solution (I,τ).

Extension of the Lasso path

Multi-penalty framework

- Building upon the Lasso path, we can create tiles, which contain parameter regions leading to the same support and sign pattern;
- Results on the tiles structure using a directed multi-graph framework;
- An algorithm for efficient computation of the tiles over the whole range of the parameters.



Numerical experiments

The model problem

$$y = T(u^\dagger + v^\dagger),$$

where $T \in \mathbb{R}^{m\times N}$ is a measurement operator, $u^\dagger$ is a sparse vector and $v^\dagger$ is a noise vector.

Numerical experiments

The model problem

$$y = T(u^\dagger + v^\dagger),$$

where $T \in \mathbb{R}^{m\times N}$ is a measurement operator, $u^\dagger$ is a sparse vector and $v^\dagger$ is a noise vector.

Data:

- 100 problems for each configuration;
- three different measurement operators: **Gaussian, random circulant matrices,** and **Gamma/Gaussian matrices** of different sizes;
- $u^\dagger$ are randomly generated with entries uniformly sampled from $(1.5, 5)$;
- $v^\dagger$ are randomly generated with entries uniformly sampled from $(-0.2, 0.2)$;
- β range is $(10^{-6}, 100)$;
- support size of $u^\dagger$ is known;
- compared to ℓ_1-regularization, IHT with warm start, OMP, preconditioned Lasso;

Numerical experiments

Multi-penalty algorithm with adaptive parameter choice

- Construct the graph/tiles attainable for any (λ_1, λ_2) up to the given support size.
- Support selection via criterion:
  - For each tiling $\tau(I, \sigma)$ calculate

$$SNR(\tau) = \frac{\min_{j \in I} |[u_I]_j|}{\|v_I\|_\infty}, \text{ where}$$
$$u_I = \operatorname*{argmin}_{u:\mathrm{supp}(u)=I} \|Tu - y\|^2$$
$$v_I = T^\dagger(y - Tu_I).$$

Numerical experiments

Multi-penalty algorithm with adaptive parameter choice

- Construct the graph/tiles attainable for any (λ_1, λ_2) up to the given support size.
- Support selection via criterion:
  - For each tiling $\tau(I, \sigma)$ calculate

$$SNR(\tau) = \frac{\min_{j \in I} |[u_I]_j|}{\|v_I\|_\infty}, \text{ where}$$

$$u_I = \underset{u:\mathrm{supp}(u)=I}{\operatorname{argmin}} \|Tu - y\|^2$$

$$v_I = T^\dagger(y - Tu_I).$$

  - Select

$$I^* = \arg \max_{\tau, |I|=s} SNR(\tau).$$

Comparison with the single-parameter counterparts

Accuracy wrt to varying support size

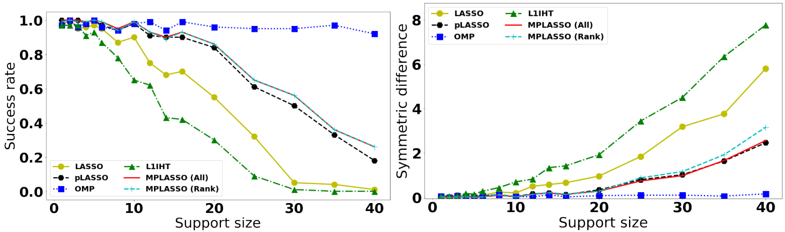


Figure: Accuracy of the support recovery for Gaussian random matrices $A \in \mathbb{R}^{600\times2500}$ and varying support sizes: (a) success rate (b) symmetric difference.

Comparison with the single-parameter counterparts

Accuracy wrt to varying support size

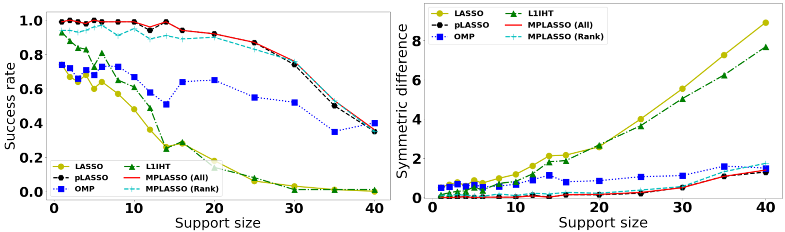


Figure: Accuracy of the support recovery for random circulant matrices $A \in \mathbb{R}^{900\times2500}$ and varying support sizes s: (a) success rate (b) symmetric difference.

Comparison with the single-parameter counterparts

Accuracy wrt to varying support size

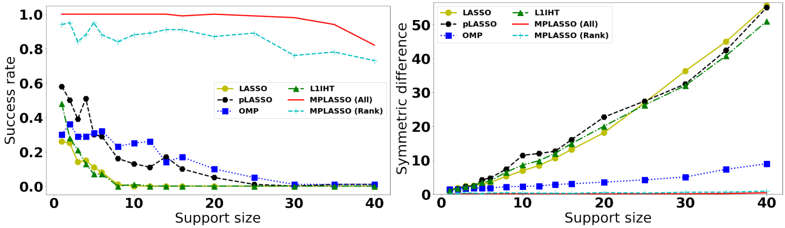


Figure: Accuracy of the support recovery for Gamma/Gaussian matrices $A \in \mathbb{R}^{900 \times 2500}$ and varying support sizes s: (a) success rate (b) symmetric difference.

Comparison with the single-parameter counterparts

Accuracy wrt to varying noise

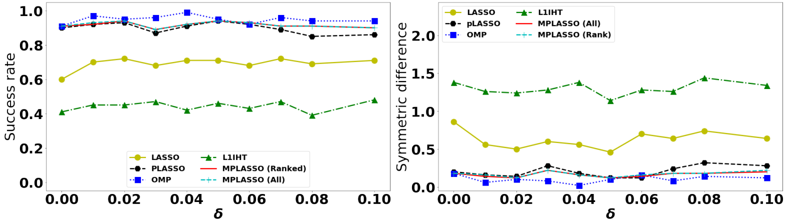


Figure: Accuracy of the support recovery for Gaussian random matrices $A \in \mathbb{R}^{600\times2500}$ and varying measurement noise: (a) success rate (b) symmetric difference.

Comparison with the single-parameter counterparts

Accuracy wrt to varying noise

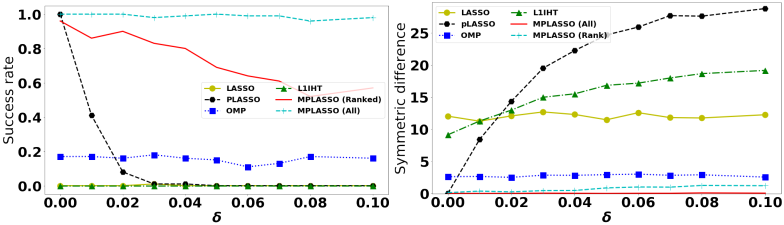


Figure: Accuracy of the support recovery for Gamma/Gaussian random matrices $A \in \mathbb{R}^{900\times 2500}$ and varying measurement noise: (a) success rate (b) symmetric difference.

Robust recovery of low-rank matrices

Joint work with Johannes Maly and Massimo Fornasier, TU Munich

Fornasier, Maly, VN, submitted, '18.

Robust recovery of low-rank matrices

- We are interested in designing an algorithm for recovery low-rank matrices from linear noisy measurements

$$y = \mathcal{A}(X) + \eta = \frac{1}{\sqrt{m}} \begin{pmatrix} \langle A_1, X \rangle_F \\ \vdots \\ \langle A_m, X \rangle_F \end{pmatrix} + \eta,$$

  - $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ is a linear measurement operator.
  - $X \in \mathbb{R}^{n_1 \times n_2}$ is an unknown low-rank matrix with approximately sparse singular vectors such that

$$X = \sum_{r=1}^{R} u^r (v^r)^T \text{ and } \|v^r\|_1 \leqslant \sqrt{s} \|v^r\|_2 \text{ for } r = 1, \dots R.$$

Robust recovery of low-rank matrices

- We are interested in designing an algorithm for recovery low-rank matrices from linear noisy measurements

$$y = \mathcal{A}(X) + \eta = \frac{1}{\sqrt{m}} \begin{pmatrix} \langle A_1, X \rangle_F \\ \vdots \\ \langle A_m, X \rangle_F \end{pmatrix} + \eta,$$

  - $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ is a linear measurement operator.
  - $X \in \mathbb{R}^{n_1 \times n_2}$ is an unknown low-rank matrix with approximately sparse singular vectors such that

$$X = \sum_{r=1}^{R} u^r (v^r)^T \text{ and } \|v^r\|_1 \leqslant \sqrt{s}\|v^r\|_2 \text{ for } r = 1, \ldots R.$$

- Low-rank matrix recovery / blind deconvolution:
  - Oymak et. al., ’15; Lee et al., ’13 and ’17; Ahmed, Recht, and Romberg, ’14, ...

Robust recovery of low-rank matrices

- We are interested in designing an algorithm for recovery low-rank matrices from linear noisy measurements

$$y = \mathcal{A}(X) + \eta = \frac{1}{\sqrt{m}} \begin{pmatrix} \langle A_1, X \rangle_F \\ \vdots \\ \langle A_m, X \rangle_F \end{pmatrix} + \eta,$$

  - $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ is a linear measurement operator.
  - $X \in \mathbb{R}^{n_1 \times n_2}$ is an unknown low-rank matrix with approximately sparse singular vectors such that

$$X = \sum_{r=1}^{R} u^r (v^r)^T \text{ and } \|v^r\|_1 \leqslant \sqrt{s}\|v^r\|_2 \text{ for } r = 1, \ldots R.$$

- Low-rank matrix recovery / blind deconvolution:
  - Oymak et. al., '15; Lee et al., '13 and '17; Ahmed, Recht, and Romberg, '14, ...
- We consider the multi-penalty functional

$$J^R_{\alpha,\beta}(u^1, \ldots, u^R, v^1, \ldots, v^R) := \left\| y - \mathcal{A}\left(\sum_{r=1}^{R} u^r (v^r)^T \right) \right\|_2^2 + \alpha \sum_{r=1}^{R} \|u^r\|_2^2 + \beta \sum_{r=1}^{R} \|v^r\|_1,$$

Theoretical results

$$\begin{cases} u_{k+1}^{l} &= \operatorname{argmin}_u \left\| (y - \mathcal{A}\,[\sum_{r=2}^{R} u_k^r {v_k^r}^T\,]\,) - \mathcal{A}\,(u {v_k^l}^T\,) \right\|_2^2 + \alpha\,\|u\|_2^2 + \frac{1}{2\lambda_k^l}\,\|u - u_k^l\|_2^2, \\ v_{k+1}^{l} &= \operatorname{argmin}_v \left\| (y - \mathcal{A}\,[\sum_{r=2}^{R} u_k^r {v_k^r}^T\,]\,) - \mathcal{A}\,(u_{k+1}^l v^T\,) \right\|^2 + \beta\,\|v\|_1 + \frac{1}{2\mu_k^l}\,\|v - v_k^l\|_2^2, \end{cases}$$

Theoretical results

$$\begin{cases} u_{k+1}^{l} & = \operatorname{argmin}_u \left\| (y - \mathcal{A}\,[\sum_{r=2}^{R} u_k^r {v_k^r}^T\,]\,) - \mathcal{A}\,(u {v_k^l}^T\,) \right\|_2^2 + \alpha \|u\|_2^2 + \frac{1}{2\lambda_k^l} \|u - u_k^l\|_2^2, \\ v_{k+1}^{l} & = \operatorname{argmin}_v \left\| (y - \mathcal{A}\,[\sum_{r=2}^{R} u_k^r {v_k^r}^T\,]\,) - \mathcal{A}\,(u_{k+1}^l v^T\,) \right\|^2 + \beta \|v\|_1 + \frac{1}{2\mu_k^l} \|v - v_k^l\|_2^2, \end{cases}$$

Theorem (Strong Convergence)

The algorithm produces $(u_k^{(1)} \ldots v_k^{(R)})$ *which converges to a global minimizer of* $J_{\alpha,\beta}^R$.

Theoretical results

$$\begin{cases} u^l_{k+1} &= \mathrm{argmin}_u \left\| (y - \mathcal{A}\,[\sum_{r=2}^{R} u^r_k {v^r_k}^T]) - \mathcal{A}\,(u {v^l_k}^T) \right\|_2^2 + \alpha \|u\|_2^2 + \frac{1}{2\lambda^l_k} \|u - u^l_k\|_2^2, \\ v^l_{k+1} &= \mathrm{argmin}_v \left\| (y - \mathcal{A}\,[\sum_{r=2}^{R} u^r_k {v^r_k}^T]) - \mathcal{A}\,(u^l_{k+1} v^T) \right\|^2 + \beta \|v\|_1 + \frac{1}{2\mu^l_k} \|v - v^l_k\|_2^2, \end{cases}$$

Theorem (Strong Convergence)

The algorithm produces $(u_k^{(1)} \ldots v_k^{(R)})$ *which converges to a global minimizer of* $J^R_{\alpha,\beta}$.

Lemma (Number of measurements)

An operator $\mathcal{A}$ *fulfils the rank-R approximately* (s_1, s_2)*-sparse RIP w.h.p. with* γ *constant for*

$$m \gtrsim \gamma^{-4} R^3 (s_1 + s_2) \log^3 (\max\{n_1, n_2\})$$

if all A_i *for* $1 \leqslant i \leqslant m$ *have iid subgaussian entries and*

$$\|u^r\|_1 / \|u^r\|_2 \leqslant \sqrt{s_1}, \|v^r\|_1 / \|v^r\|_2 \leqslant \sqrt{s_2}.$$

Theoretical results

$$\begin{cases} u^l_{k+1} &= \operatorname{argmin}_u \left\| (y - \mathcal{A}[\sum_{r=2}^R u^r_k v^{r\,T}_k]) - \mathcal{A}(u v^{l\,T}_k) \right\|_2^2 + \alpha \|u\|_2^2 + \frac{1}{2\lambda^l_k} \|u - u^l_k\|_2^2, \\ v^l_{k+1} &= \operatorname{argmin}_v \left\| (y - \mathcal{A}[\sum_{r=2}^R u^r_k v^{r\,T}_k]) - \mathcal{A}(u^l_{k+1} v^T) \right\|^2 + \beta \|v\|_1 + \frac{1}{2\mu^l_k} \|v - v^l_k\|_2^2, \end{cases}$$

Theorem (Strong Convergence)

The algorithm produces $(u_k^{(1)} \ldots v_k^{(R)})$ *which converges to a global minimizer of* $J^R_{\alpha,\beta}$.

Lemma (Number of measurements)

An operator $\mathcal{A}$ *fulfils the rank-R approximately* (s_1, s_2)*-sparse RIP w.h.p. with* γ *constant for*

$$m \gtrsim \gamma^{-4} R^3 (s_1 + s_2) \log^3 (\max\{n_1, n_2\})$$

if all A_i *for* $1 \leqslant i \leqslant m$ *have iid subgaussian entries and*

$$\|u^r\|_1 / \|u^r\|_2 \leqslant \sqrt{s_1}, \|v^r\|_1 / \|v^r\|_2 \leqslant \sqrt{s_2}.$$

Theorem (Error estimations for approximately sparse matrices)

Assume $\alpha = \beta = \|\eta\|_2^2 / \|\hat{X}\|_{\frac{2}{3}}^{\frac{2}{3}} < 1$ *and that* $\mathcal{A}$ *satisfies RIP with some constant* γ. *Then*

$$\|\hat{X} - X_{\alpha,\beta}\|_F \leqslant (2\sqrt{CR^{2/3}s^{1/3}} + 2)\|\eta\|_2 + \sqrt{\gamma}$$

Theoretical results

$$\begin{cases} u^l_{k+1} & = \operatorname{argmin}_u \left\| (y - \mathcal{A}\,[\sum_{r=2}^R u^r_k {v^r_k}^T\,]\,) - \mathcal{A}\,(u{v^l_k}^T\,) \right\|_2^2 + \alpha\,\|u\|_2^2 + \frac{1}{2\lambda^l_k}\,\|u - u^l_k\|_2^2, \\ v^l_{k+1} & = \operatorname{argmin}_v \left\| (y - \mathcal{A}\,[\sum_{r=2}^R u^r_k {v^r_k}^T\,]\,) - \mathcal{A}\,(u^l_{k+1}v^T\,) \right\|^2 + \beta\,\|v\|_1 + \frac{1}{2\mu^l_k}\,\|v - v^l_k\|_2^2, \end{cases}$$

Theorem (Strong Convergence)

The algorithm produces $(u_k^{(1)} \dots v_k^{(R)})$ *which converges to a global minimizer of* $J^R_{\alpha,\beta}$.

Lemma (Number of measurements)

An operator $\mathcal{A}$ *fulfils the rank-R approximately* (s_1, s_2)*-sparse RIP w.h.p. with* γ *constant for*

$$m \gtrsim \gamma^{-4} R^3 (s_1 + s_2) \log^3 (\max\{n_1, n_2\})$$

if all A_i *for* $1 \leqslant i \leqslant m$ *have iid subgaussian entries and*

$$\|u^r\|_1 / \|u^r\|_2 \leqslant \sqrt{s_1}, \|v^r\|_1 / \|v^r\|_2 \leqslant \sqrt{s_2}.$$

Theorem (Error estimations for approximately sparse matrices)

Assume $\alpha = \beta = \|\eta\|_2^2 / \|\hat{X}\|_{\frac{2}{3}}^{\frac{2}{3}} < 1$ *and that* $\mathcal{A}$ *satisfies RIP with some constant* γ. *Then*

$$\|\hat{X} - X_{\alpha,\beta}\|_F \leqslant (2\sqrt{CR^{2/3}s^{1/3}} + 2)\|\eta\|_2 + \sqrt{\gamma} \leqslant (2\sqrt{s^{1/3}} + 2)\|\eta\|_2 + \sqrt{\gamma}.$$

Numerical experiments

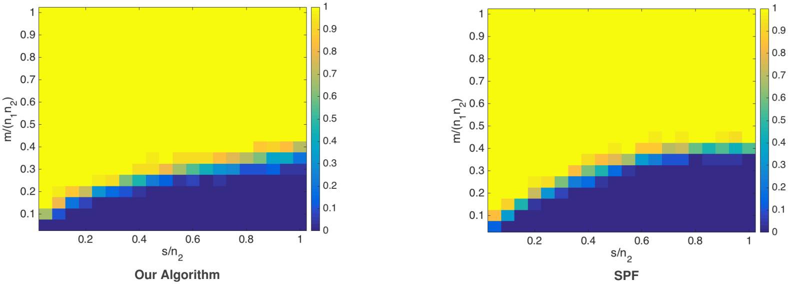
We compare our algorithm to Sparse Power Factorization (SPF) [Lee et al. ’17], the so far stand-alone algorithm for low-rank recovery under additional sparsity constraints.

Numerical experiments

We compare our algorithm to Sparse Power Factorization (SPF) [Lee et al. '17], the so far stand-alone algorithm for low-rank recovery under additional sparsity constraints.

Data:

- 30 iterations, $X \in \mathbb{R}^{16\times 100}$ with $\|X\|_F = 10$;
- $R = 1$, $\|\eta\| = 0.3\|X\|_F$;
- success if $\|X - X_{\text{approx}}\|_F / \|X\|_F \leqslant 0.4$.



Conclusions and Outlook

- We have presented a unified framework for solution of the inverse problems of unmixing type by means of multi-penalty regularization;
- Both theoretical and numerical results show superiority of the multi-penalty regularization compared to its single-penalty counterparts;
- Adaptive parameter choice rule allows for efficient and accurate selection of the optimal parameters.

Conclusions and Outlook

- We have presented a unified framework for solution of the inverse problems of unmixing type by means of multi-penalty regularization;
- Both theoretical and numerical results show superiority of the multi-penalty regularization compared to its single-penalty counterparts;
- Adaptive parameter choice rule allows for efficient and accurate selection of the optimal parameters.

Open questions:

- Theoretical results for

$$J(u,v) = \frac{1}{2}\|T(u+v) - y^\delta\|_Y^2 + \lambda_1\|u\|_{\ell_1} + \lambda_2\|v\|_{\ell_\infty}.$$

Conclusions and Outlook

- We have presented a unified framework for solution of the inverse problems of unmixing type by means of multi-penalty regularization;
- Both theoretical and numerical results show superiority of the multi-penalty regularization compared to its single-penalty counterparts;
- Adaptive parameter choice rule allows for efficient and accurate selection of the optimal parameters.

Open questions:

- Theoretical results for

$$J(u,v) = \frac{1}{2}\|T(u+v) - y^\delta\|_Y^2 + \lambda_1\|u\|_{\ell_1} + \lambda_2\|v\|_{\ell_\infty}.$$

- Numerical and theoretical framework for the matrix recovery from incomplete data y by the multi-penalty regularization of the type

$$J(u,v) = \|T(\sum_{r=1}^{R} u^r \cdot (v^r)^T) - y\|_F^2 + \lambda_1 \sum_{r=1}^{R} \|u^r\|_{\ell_q}^q + \lambda_2 \sum_{r=1}^{R} \|v^r\|_{\ell_p}^p,$$

Acknowledgements

- Timo Klock, Simula Research Lab
- Markus Grasmair, NTNU
- Massimo Fornasier, TU Munich
- Steffen Peter, TU Munich
- Johannes Maly, TU Munich
- Sergei Pereverzyev, RICAM

Code + More results:

`github.com/soply/mpgraph`

`github.com/soply/sparse_encoder_testsuite`

Sponsored by **Norwegian Research Council, Deutsche Forschungsgemeinschaft** (DFG) and **Fonds zur Förderung der wissenschaftlichen Forschung** (FWF).

Acknowledgements

- Timo Klock, Simula Research Lab
- Markus Grasmair, NTNU
- Massimo Fornasier, TU Munich
- Steffen Peter, TU Munich
- Johannes Maly, TU Munich
- Sergei Pereverzyev, RICAM

Code + More results:

`github.com/soply/mpgraph`

`github.com/soply/sparse_encoder_testsuite`

Sponsored by **Norwegian Research Council, Deutsche Forschungsgemeinschaft** (DFG) and **Fonds zur Förderung der wissenschaftlichen Forschung** (FWF).

Thank you very much for your attention!

Literature

The results presented are sampled from recent papers:

1. *Multi-penalty regularization with a component-wise penalization* (with S. Pereverzyev), Inverse Problems (2013).
2. *Parameter choice strategies for multi-penalty regularization* (with M. Fornasier, S. Pereverzyev), SIAM Journal on Numerical Analysis (2014).
3. *Minimization of multi-penalty functionals by alternating iterative thresholding and optimal parameter choices* (with S. Peter), Inverse Problems (2014).
4. *Conditions on optimal support recovery in unmixing problems by means of multi-penalty regularization* (with M. Grasmair), Inverse Problems (2016).
5. *Extension of the Lasso path algorithm to multi-parameter regularization* (with M. Grasmair and T. Klock), submitted (2017).
6. *A-T-LAS$_{2;1}$: A Multi-Penalty Approach to Compressed Sensing of Low-Rank Matrices with Sparse Singular Vectors* (with M. Fornasier and J. Maly), submitted (2018).