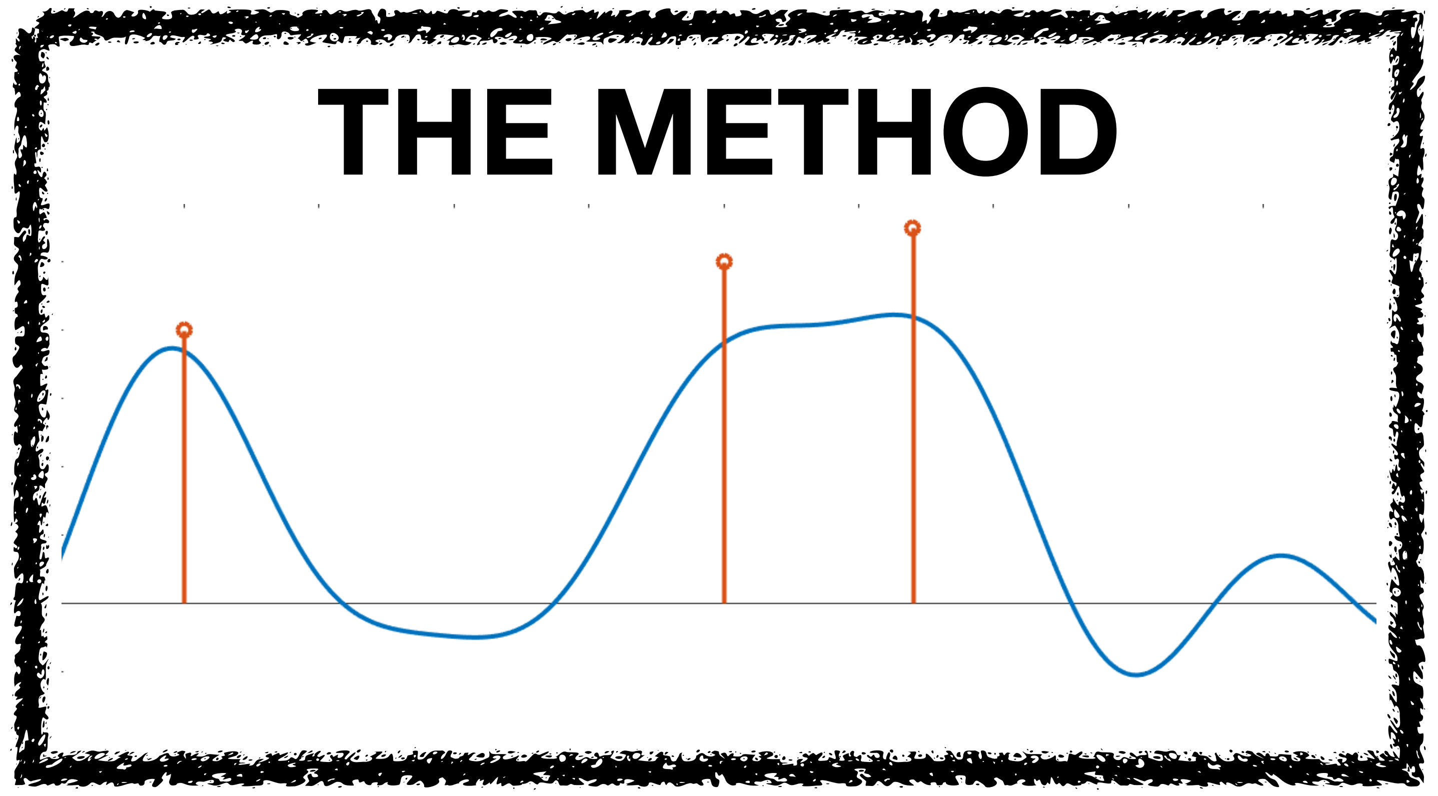
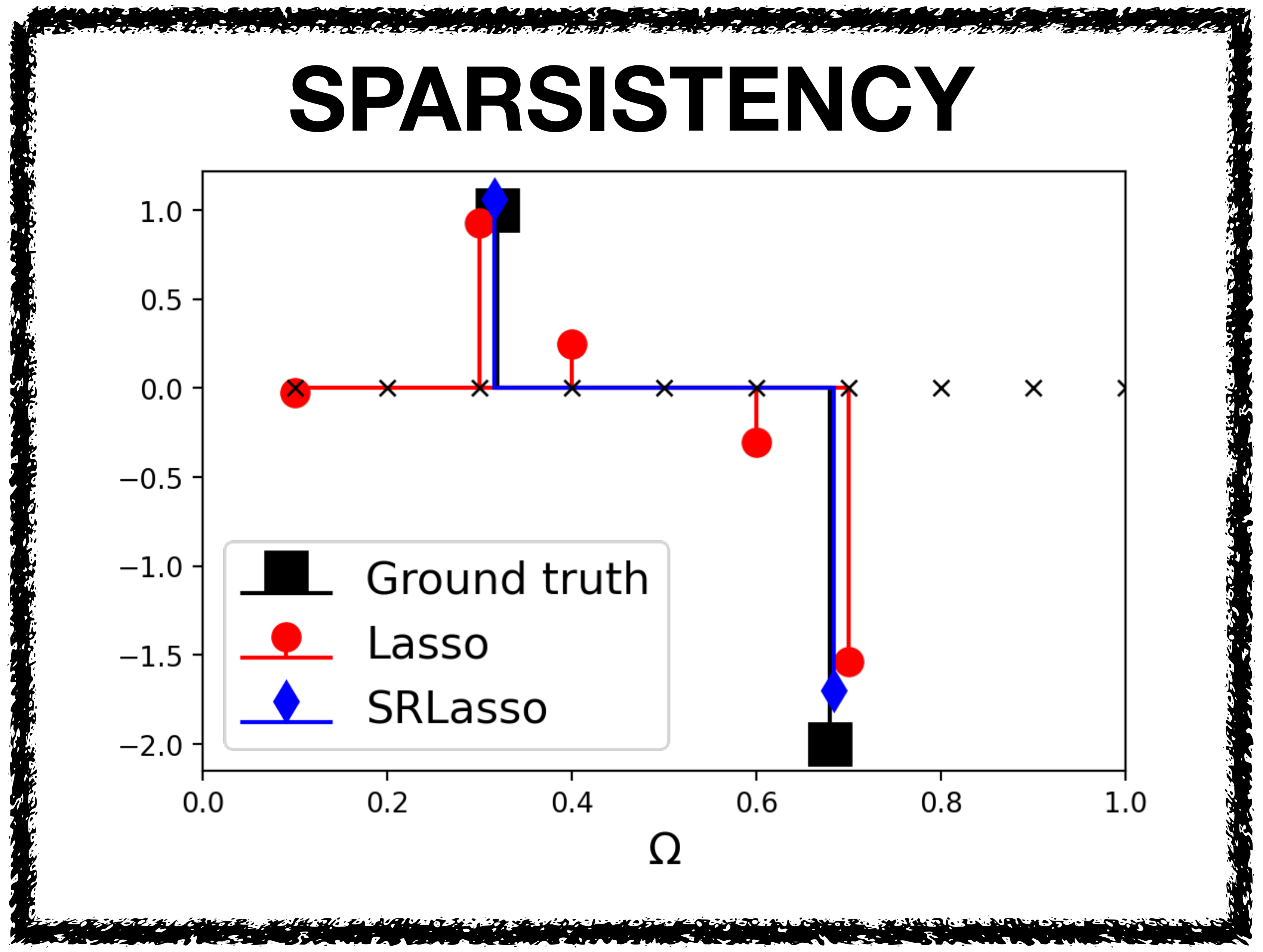
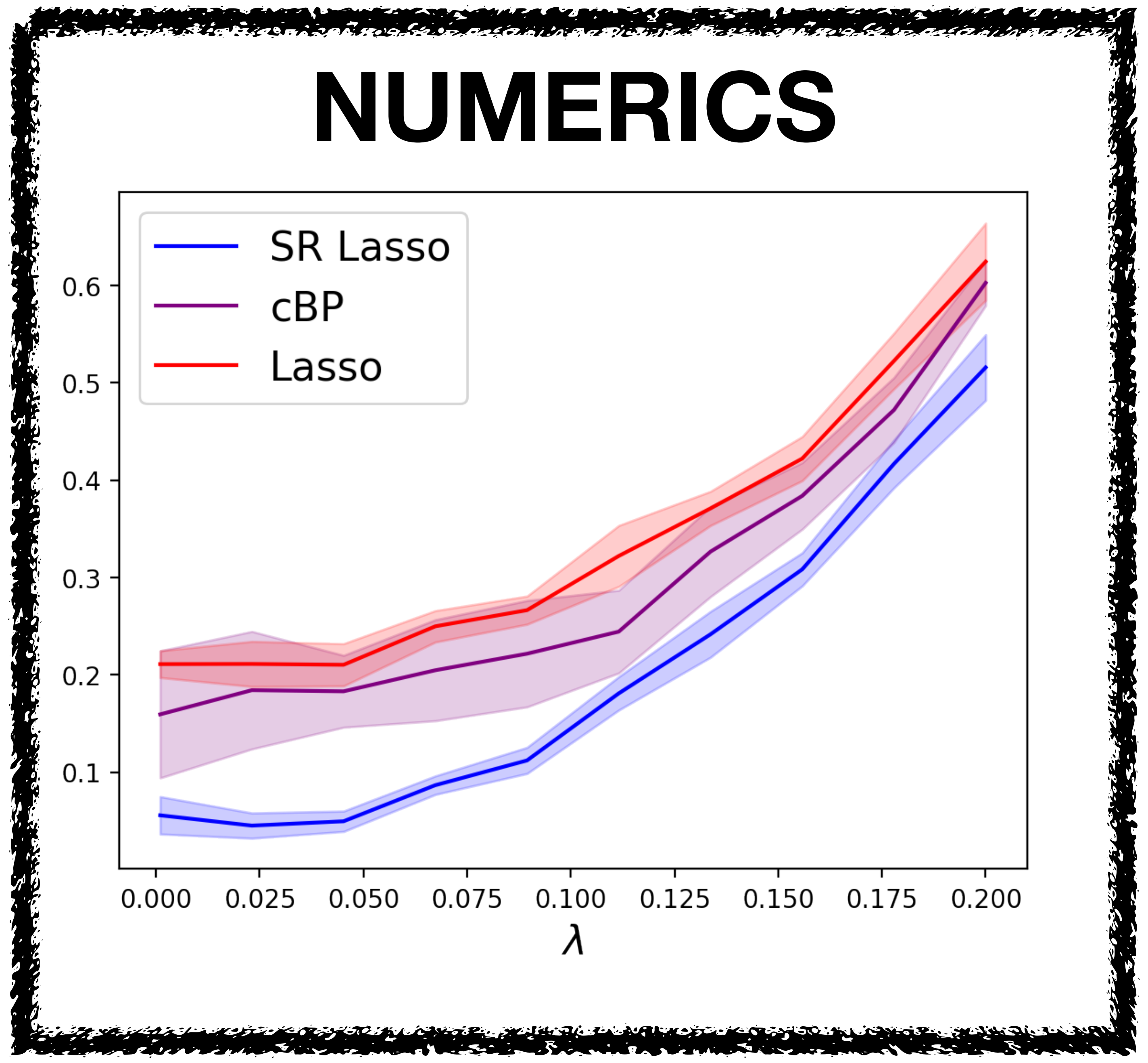
Super-resolved Lasso

Clarice Poon (U. Warwick) and Gabriel Peyré (ENS Paris)

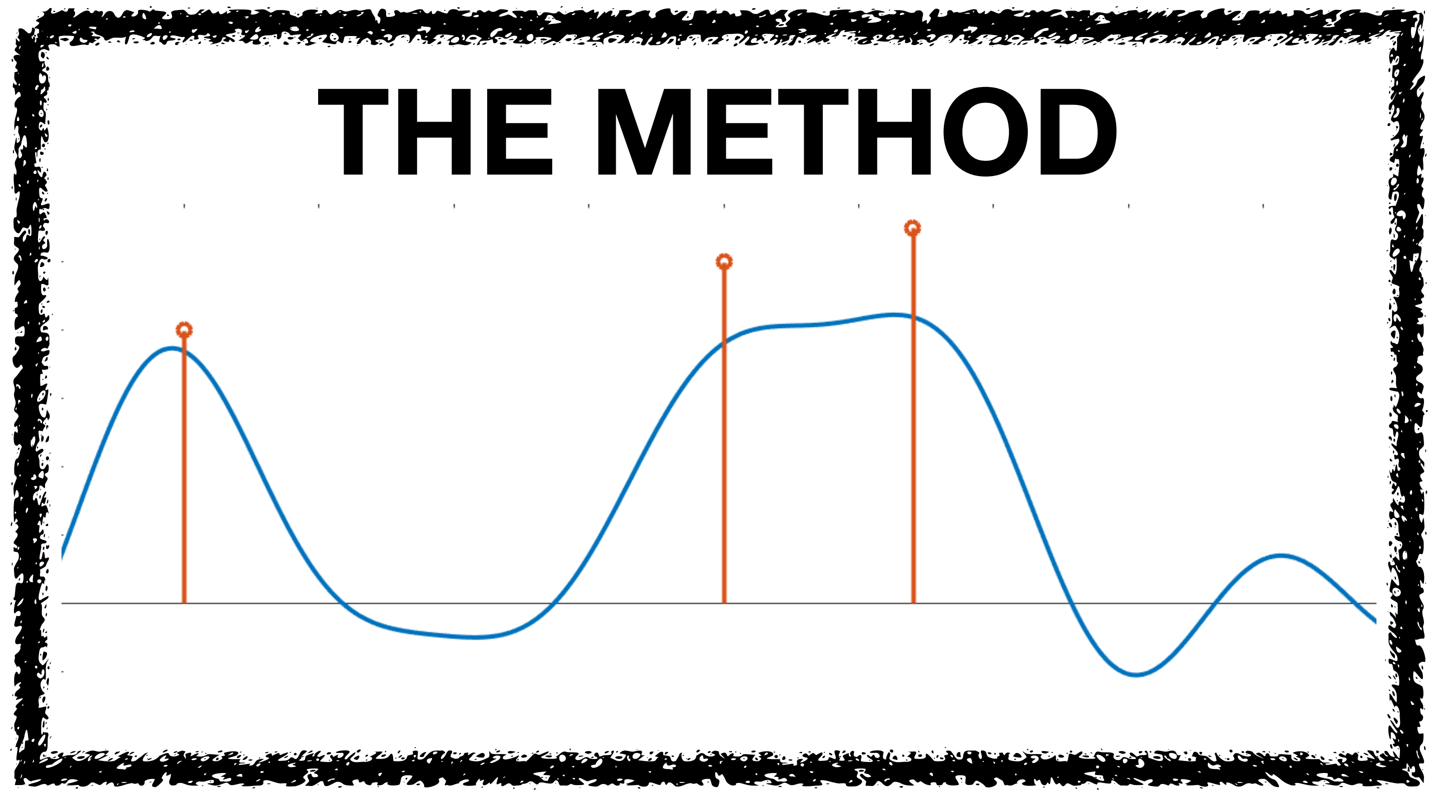
Outline


THE METHOD


SPARSISTENCY
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
Ω
Ground truth
Lasso
SRLasso


NUMERICS
SR Lasso
cBP
Lasso
0.6
0.5
0.4
0.3
0.2
0.1
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
λ

Outline

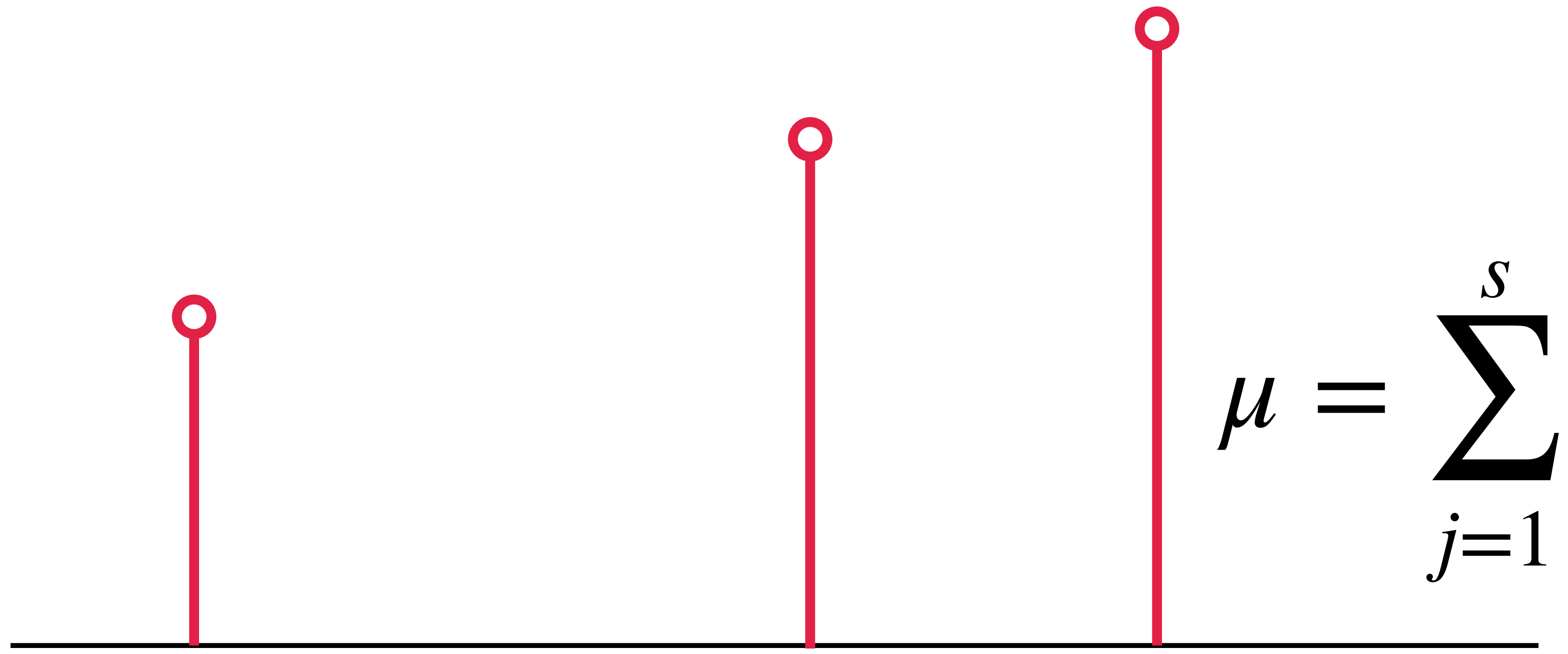

THE METHOD

SPARSISTENCY
Ground truth
Lasso
SRLasso
Ω

NUMERICS
SR Lasso
cBP
Lasso
λ

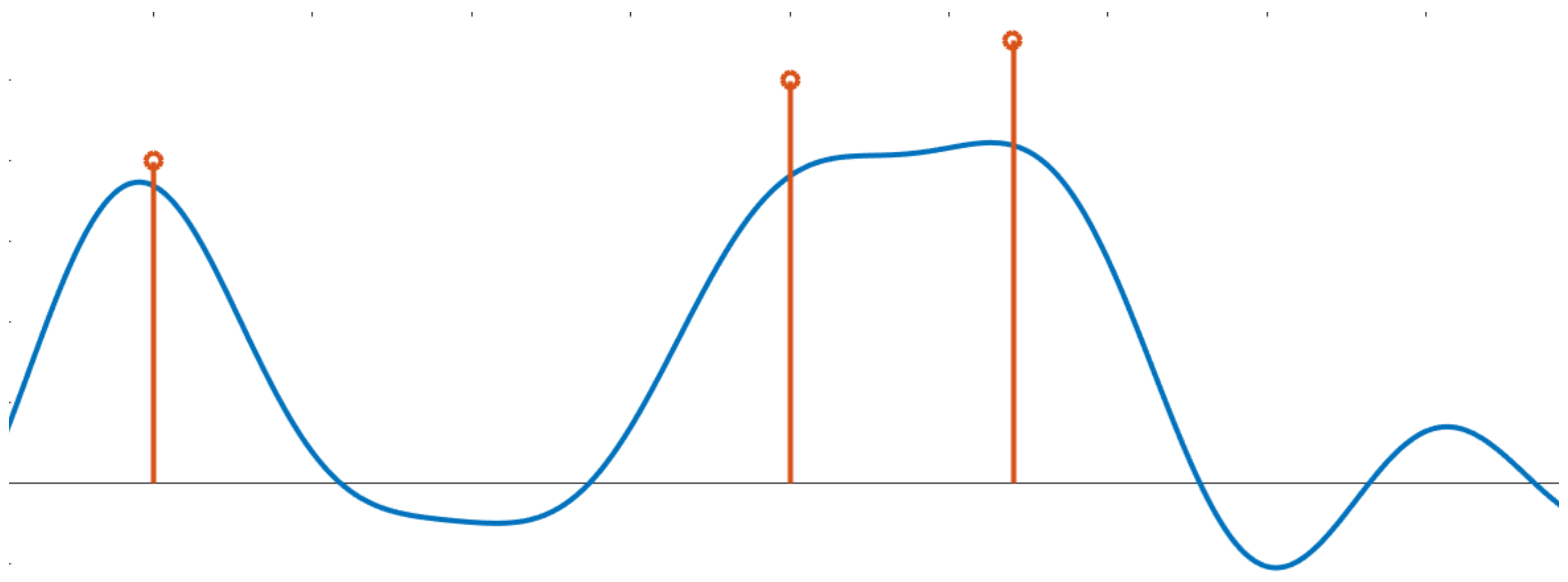
Super-resolution of point sources

Goal: Recover a sum of Diracs/spikes

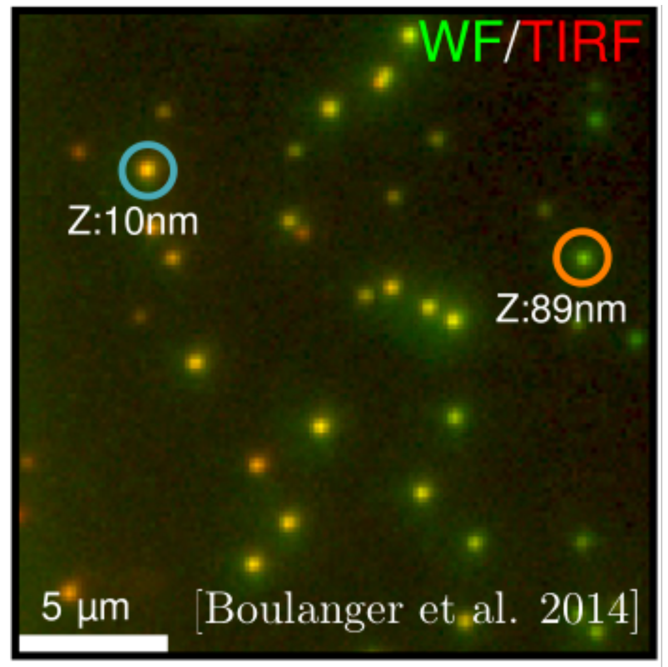


$$\mu = \sum_{j=1}^{s} a_j \delta_{x_j}$$

Observe: $y = \int \phi(x) d\mu(x) + \text{noise}$



Examples:



[Boulanger et. al. 2014]

$$y(x) = \sum_{j=1}^{s} a_j \exp(-\|x - x_j\|^2/\sigma) + \text{noise}$$

$$y_k = \sum_{j=1}^{s} a_j \exp(i2\pi k x_j) + \text{noise},$$

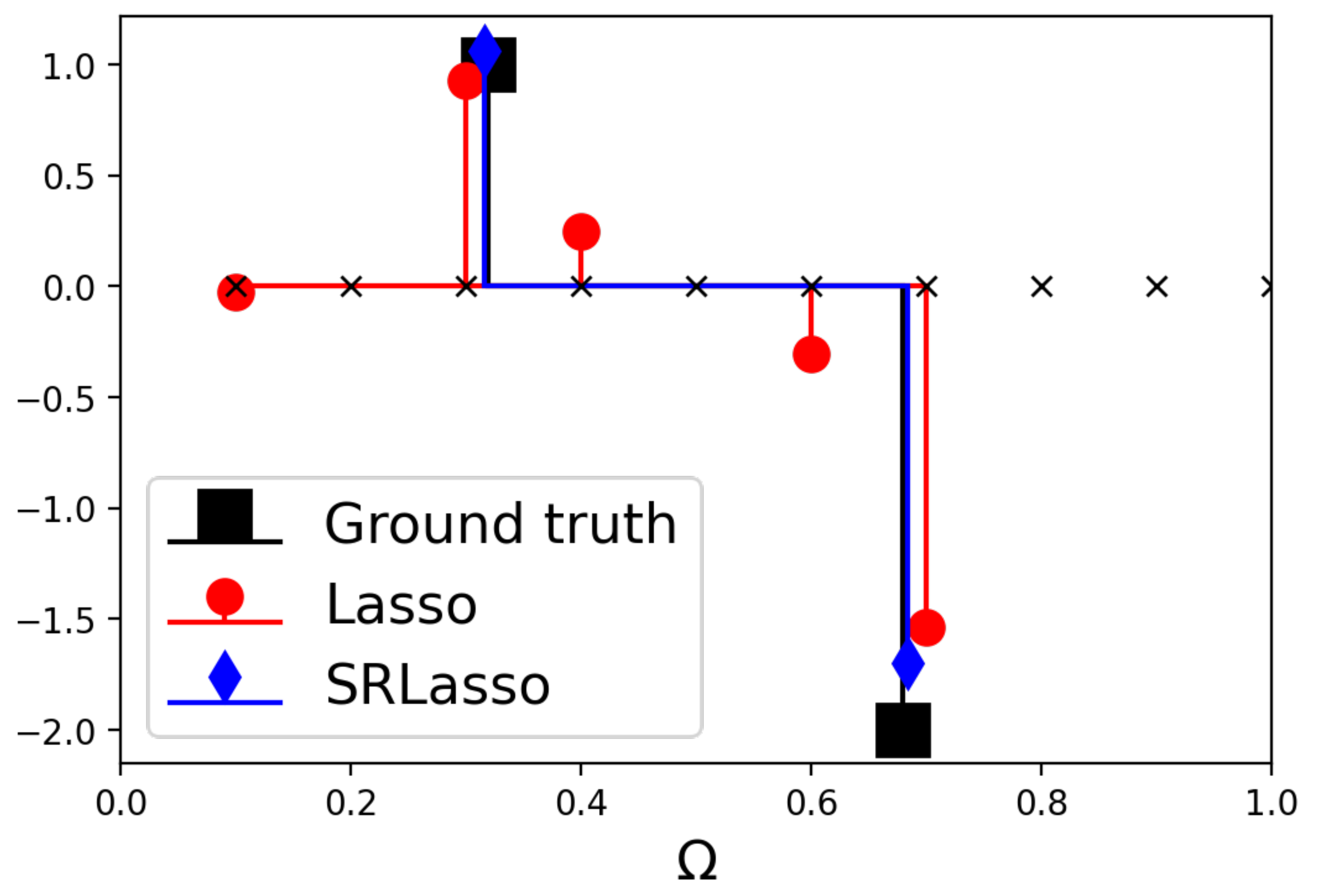
for $|k| \le f_c$

Lasso

Discretise on grid $\{x_j : \; j = 1,\ldots,N\}$:

$$\Phi\mu = \int \phi(x)\mu(dx) \approx \sum_{j=1}^{N} \phi(x_j)\beta_j =: X\beta$$

Sparse regularisation: $\min_{\beta \in \mathbb{R}^N} \|\beta\|_1 + \frac{1}{2\lambda}\|X\beta - y\|_2^2$

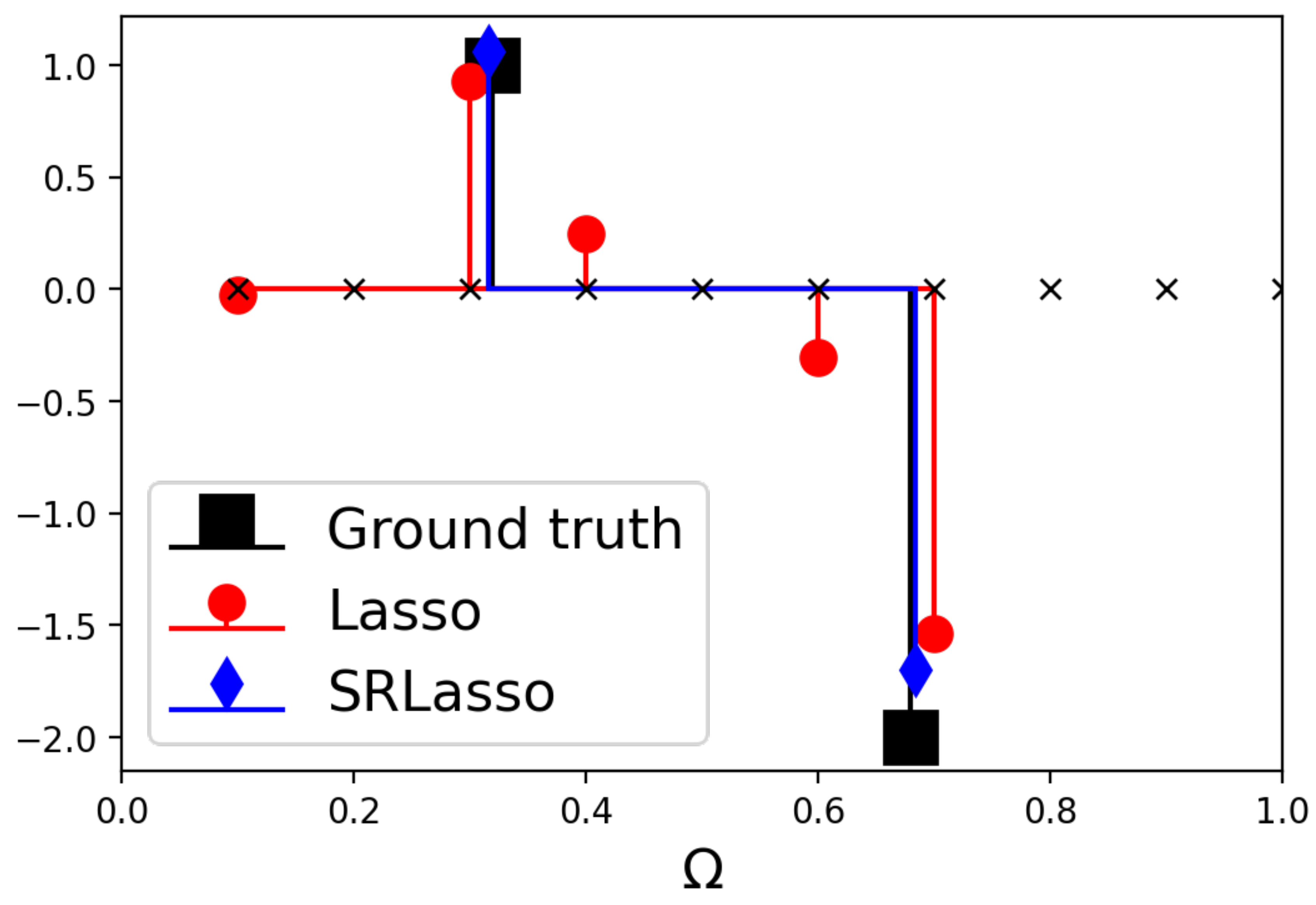


Lasso

Discretise on grid $\{x_j : \ j = 1,\ldots,N\}$:

$$\Phi\mu = \int \phi(x)\mu(dx) \approx \sum_{j=1}^{N} \phi(x_j)\beta_j =: X\beta$$

Sparse regularisation: $\min_{\beta\in\mathbb{R}^N} \|\beta\|_1 + \frac{1}{2\lambda}\|X\beta - y\|_2^2$



✓ This is relatively simple to solve with wide choice of algorithms.

✗
- Algorithms become slow when grid is too fine (high coherence in columns of X)
- Quantisation effects [Duval & Peyre '17]

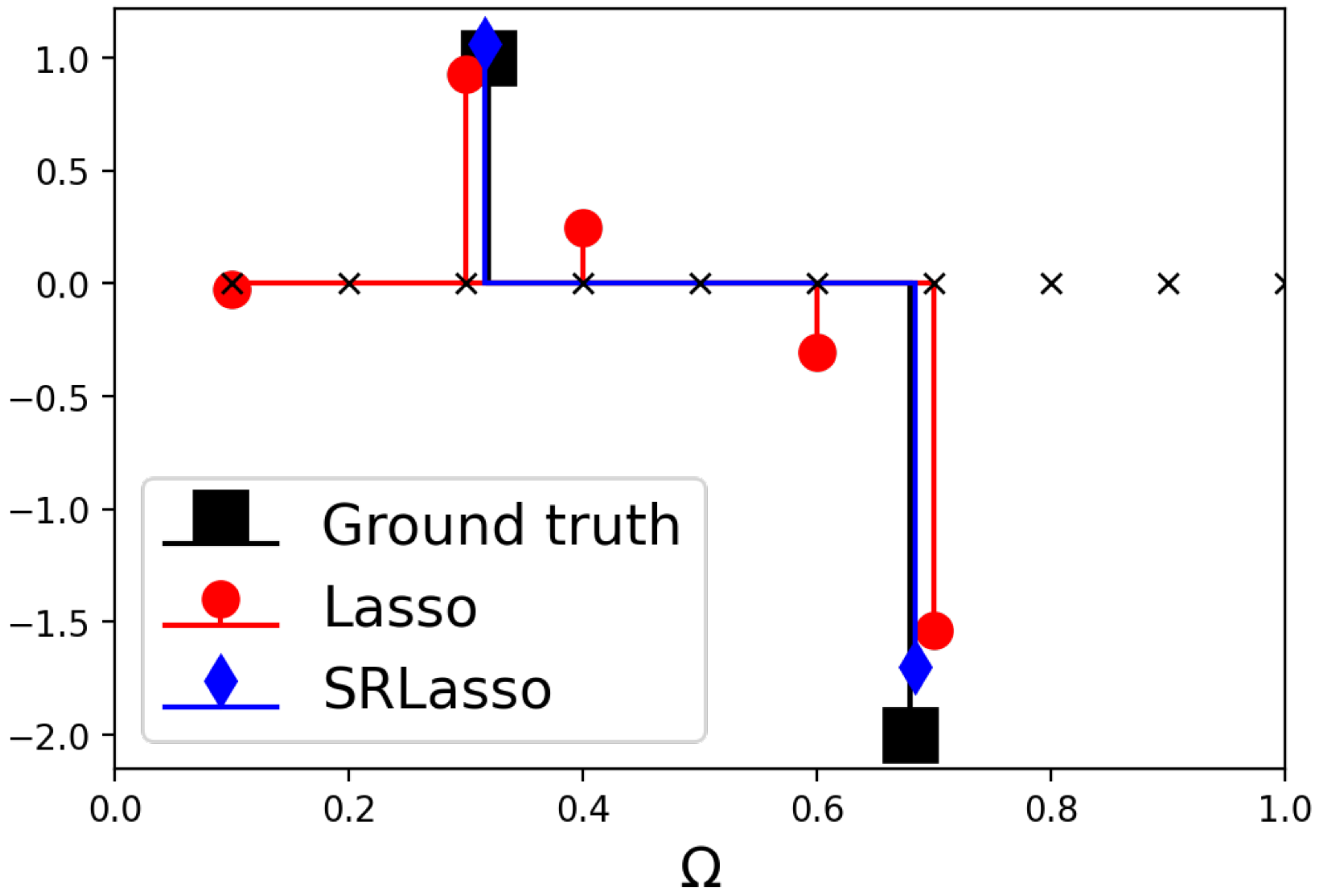
Ref: Duval & Peyré. "Sparse spikes super-resolution on thin grids I" *Inverse Problems* (2017)

Lasso

Discretise on grid $\{x_j : \; j = 1,\ldots,N\}$:

$$\Phi\mu = \int \phi(x)\mu(dx) \approx \sum_{j=1}^{N} \phi(x_j)\beta_j =: X\beta$$

Sparse regularisation: $\min_{\beta\in\mathbb{R}^N} \|\beta\|_1 + \frac{1}{2\lambda}\|X\beta - y\|_2^2$



✓ This is relatively simple to solve with wide choice of algorithms.

✗
- Algorithms become slow when grid is too fine (high coherence in columns of X)
- Quantisation effects [Duval & Peyre ’17]

Off-the-grid approaches such as Prony methods and Beurling Lasso (direct formulation in the space of measures) resolve the issue of quantisation effects, but Lasso is still widely used due to its simplicity.

Ref: Duval & Peyré. "Sparse spikes super-resolution on thin grids I" *Inverse Problems* (2017)

Continuous Basis Pursuit [Ekanadham et al ’11]

Ground truth is **off-the-grid**: $\mu = \sum_{j=1}^{s} a_j \delta_{x_j + t_j}$ where $|t_j| \leq h/2$

Taylor expand: $y = \sum_j a_j \phi(x_j + t_j) \approx \sum_j a_j \phi(x_j) + a_j t_j \phi'(x_j) + \mathcal{O}(h^2)$

Ref: Ekanadham, Tranchina & Simoncelli. Recovery of sparse translation-invariant signals with continuous basis pursuit. *IEEE transactions on signal processing (2011)*

Continuous Basis Pursuit [Ekanadham et al ’11]

Ground truth is **off-the-grid**: $\mu = \sum_{j=1}^{s} a_j \delta_{x_j + t_j}$ where $|t_j| \leq h/2$

Taylor expand: $y = \sum_j a_j \phi(x_j + t_j) \approx \sum_j a_j \phi(x_j) + a_j t_j \phi'(x_j) + \mathcal{O}(h^2)$

$$\min_{a \in \mathbb{R}_+^N, b \in \mathbb{R}^N} \frac{1}{2} \|y - \Phi_X a - \Phi_X' b\|^2 + \lambda \|a\|_1 \quad \text{s.t.} \quad |b_j| \leq \frac{h}{2} a_j$$

$\Phi_X = [\phi(x_j)]_{j=1}^N$ and $\Phi_X' = [\phi'(x_j)]_{j=1}^N$

Ref: Ekanadham, Tranchina & Simoncelli. Recovery of sparse translation-invariant signals with continuous basis pursuit. *IEEE transactions on signal processing (2011)*

Continuous Basis Pursuit [Ekanadham et al '11]

Ground truth is **off-the-grid**: $\mu = \sum_{j=1}^{s} a_j \delta_{x_j + t_j}$ where $|t_j| \leq h/2$

Taylor expand: $y = \sum_j a_j \phi(x_j + t_j) \approx \sum_j a_j \phi(x_j) + a_j t_j \phi'(x_j) + \mathcal{O}(h^2)$

$$\min_{a \in \mathbb{R}^N_+, b \in \mathbb{R}^N} \frac{1}{2} \|y - \Phi_X a - \Phi'_X b\|^2 + \lambda \|a\|_1 \quad \text{s.t.} \quad |b_j| \leq \frac{h}{2} a_j$$

$\Phi_X = [\phi(x_j)]_{j=1}^N$ and $\Phi'_X = [\phi'(x_j)]_{j=1}^N$

✓ Convex formulation

$$\min_{r, l \in \mathbb{R}^N_+} \lambda \|r\|_1 + \lambda \|l\|_1 + \frac{1}{2} \left\| y - \begin{pmatrix} \Phi_X + \frac{h}{2}\Phi'_X & \Phi_X - \frac{h}{2}\Phi'_X \end{pmatrix} \begin{pmatrix} r \\ l \end{pmatrix} \right\|^2$$

✗ Restrictions

- Works only for **non-negative** signals $a \geq 0$
- Unstable when the grid is too fine [Duval & Peyre '17]

Ref: Ekanadham, Tranchina & Simoncelli. Recovery of sparse translation-invariant signals with continuous basis pursuit. *IEEE transactions on signal processing (2011)*
Ref: Duval & Peyré. "Sparse spikes super-resolution on thin grids II" *Inverse Problems (2017)*

Super-resolved Lasso

Unconstrained optimisation problem

$$\min_{a,b\in\mathbb{R}^N} \frac{1}{2}\|y - \Phi_X a - \tau\Phi'_X b\|^2 + \lambda \sum_{j=1}^{N} \sqrt{a_j^2 + b_j^2}$$

- Performance depends on appropriately weighting Φ'_X.
- Define normalised derivative $\psi(x) := \phi'(x)/\|\phi'(x)\|$ and let $\Psi_X = [\psi(x_j)]_{j=1}^N$

Super-resolved Lasso

Unconstrained optimisation problem

$$\min_{a,b\in\mathbb{R}^N} \frac{1}{2}\|y - \Phi_X a - \tau\Phi'_X b\|^2 + \lambda\sum_{j=1}^{N}\sqrt{a_j^2 + b_j^2}$$

- Performance depends on appropriately weighting Φ'_X.
- Define normalised derivative $\psi(x) := \phi'(x)/\|\phi'(x)\|$ and let $\Psi_X = [\psi(x_j)]_{j=1}^N$

$$\min_{a,b\in\mathbb{R}^N} \frac{1}{2}\|y - \Phi_X a - \tau\Psi_X b\|^2 + \lambda\sum_{j=1}^{N}\sqrt{a_j^2 + b_j^2}$$

Solution interpretation

- Parameter $\tau \in [0,1]$ controls how far we move inside the grid.
- Solution $\mu = \sum_{j=1}^{N} a_j\delta_{x_j+t_j}$ where $t_j = \dfrac{\tau b_j}{a_j\|\phi'(x_j)\|}$

Super-resolved Lasso

Unconstrained optimisation problem

$$\min_{a,b\in\mathbb{R}^N} \frac{1}{2}\|y - \Phi_X a - \tau\Phi_X' b\|^2 + \lambda \sum_{j=1}^{N} \sqrt{a_j^2 + b_j^2}$$

- Performance depends on appropriately weighting Φ_X'.
- Define normalised derivative $\psi(x) := \phi'(x)/\|\phi'(x)\|$ and let $\Psi_X = [\psi(x_j)]_{j=1}^N$

$$\min_{a,b\in\mathbb{R}^N} \frac{1}{2}\|y - \Phi_X a - \tau\Psi_X b\|^2 + \lambda \sum_{j=1}^{N} \sqrt{a_j^2 + b_j^2}$$

Solution interpretation

- Parameter $\tau \in [0,1]$ controls how far we move inside the grid.
- Solution $\mu = \sum_{j=1}^{N} a_j \delta_{x_j + t_j}$ where $t_j = \dfrac{\tau b_j}{a_j \|\phi'(x_j)\|}$

Can handle arbitrarily signed signals (including complex signs)!

In practice, we choose τ to be 1 or very close to 1.

Group-Lasso

Let $\Gamma = \begin{bmatrix}\Phi_X & \tau\Psi_X\end{bmatrix}$ and write $z = \begin{pmatrix} a \\ b \end{pmatrix}$. Then,

Group Lasso: $$\min_z \frac{1}{2}\|\Gamma z - y\|^2 + \lambda\|z\|_{1,2}$$

Group-Lasso

Let $\Gamma = \begin{bmatrix} \Phi_X & \tau\Psi_X \end{bmatrix}$ and write $z = \begin{pmatrix} a \\ b \end{pmatrix}$. Then,

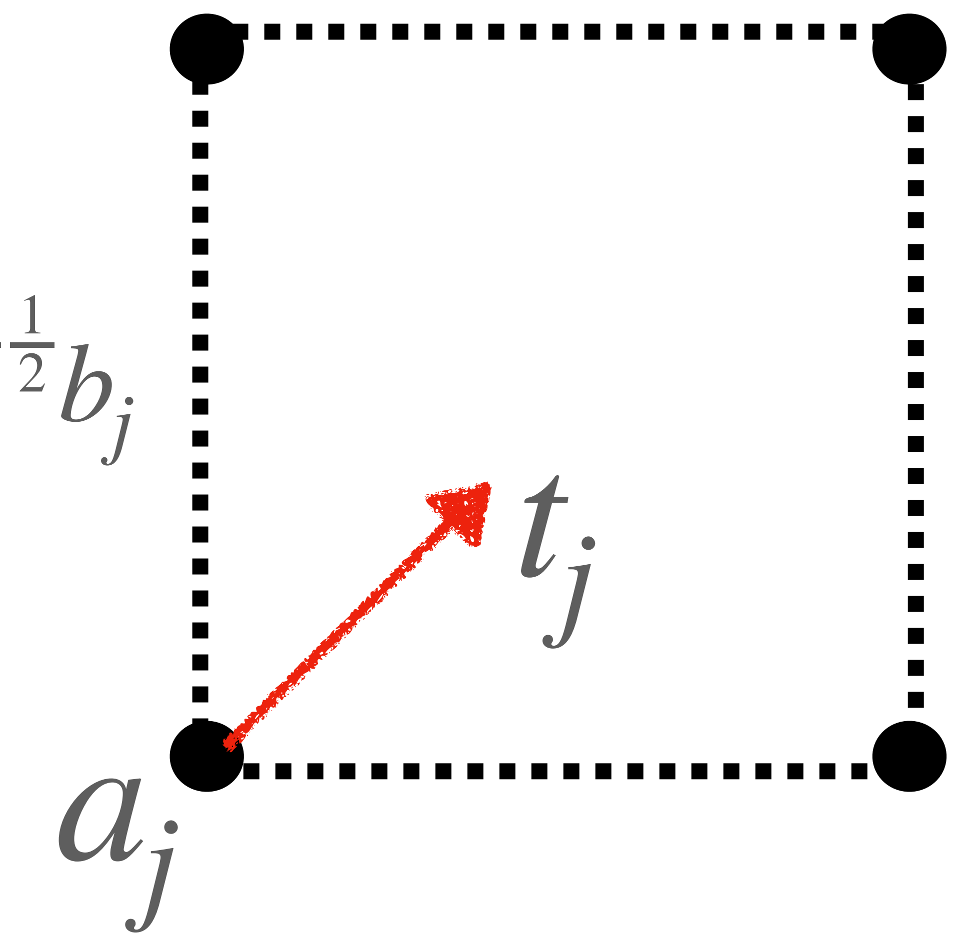
Group Lasso: $$\min_{a\in\mathbb{R}^N, b\in\mathbb{R}^{dN}} \frac{1}{2}\|\Gamma\begin{pmatrix} a \\ b \end{pmatrix} - y\|^2 + \lambda \sum_i \sqrt{a_i^2 + \|b_i\|^2}$$

Multivariate setting:

- $\Psi_X b = \sum_{i=1}^N b_i^\top (G_{x_i}^{-1/2} \nabla\phi(x_i))$, with $b_i \in \mathbb{R}^d$ and $G_x = \nabla\phi(x)\nabla\phi(x)^\top$

  - Normalisation to ensure that $\Gamma^\top\Gamma$ is block identity I_{d+1}.
  - For translation invariant kernels, $\langle\phi(x), \phi(z)\rangle = \kappa(x-y)$
    $G_x = -\nabla^2\kappa(0)$ is constant.

Shift $t_j = \frac{\tau}{a_j} G_{x_j}^{-\frac{1}{2}} b_j$



Group-Lasso

Let $\Gamma = \begin{bmatrix} \Phi_X & \tau \Psi_X \end{bmatrix}$ and write $z = \begin{pmatrix} a \\ b \end{pmatrix}$. Then,

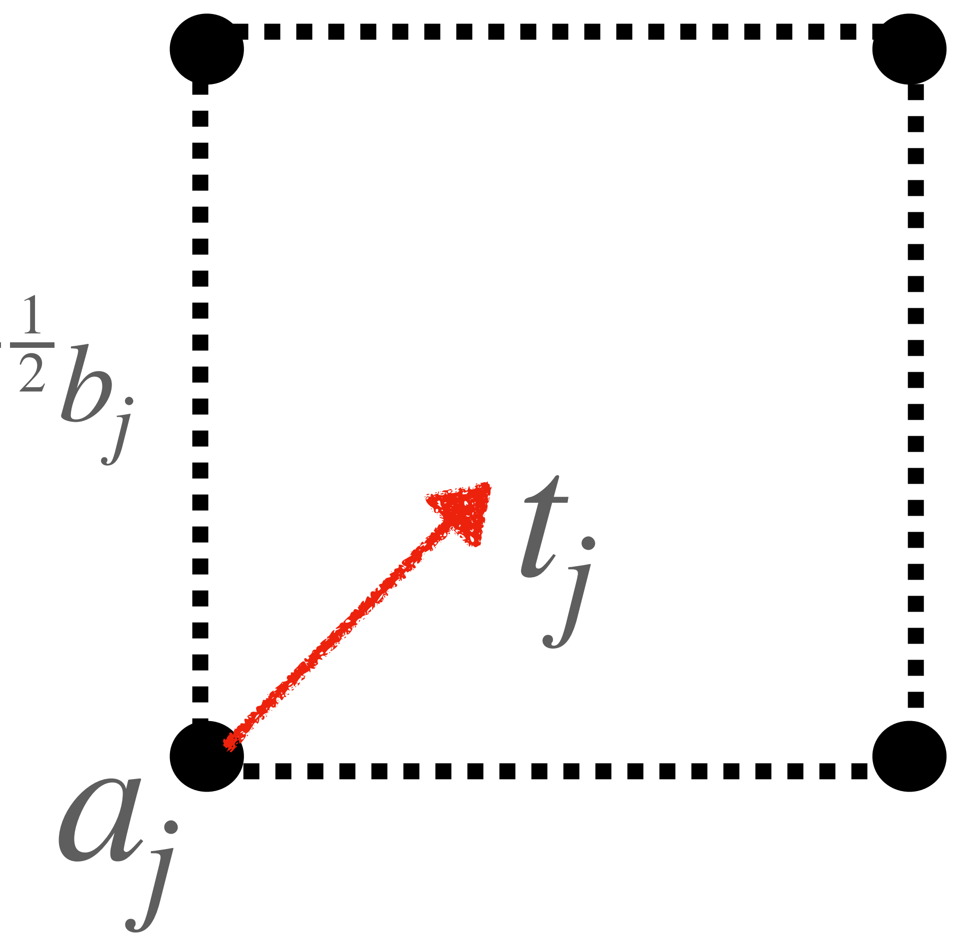
Group Lasso: $$\min_{a \in \mathbb{R}^N, b \in \mathbb{R}^{dN}} \frac{1}{2} \|\Gamma \begin{pmatrix} a \\ b \end{pmatrix} - y\|^2 + \lambda \sum_i \sqrt{a_i^2 + \|b_i\|^2}$$

Multivariate setting:

- $\Psi_X b = \sum_{i=1}^{N} b_i^\top (G_{x_i}^{-1/2} \nabla \phi(x_i))$, with $b_i \in \mathbb{R}^d$ and $G_x = \nabla \phi(x) \nabla \phi(x)^\top$

  - Normalisation to ensure that $\Gamma^\top \Gamma$ is block identity I_{d+1}.
  - For translation invariant kernels, $\langle \phi(x), \phi(z) \rangle = \kappa(x - y)$
    $G_x = - \nabla^2 \kappa(0)$ is constant.

Shift $t_j = \frac{\tau}{a_j} G_{x_j}^{-\frac{1}{2}} b_j$

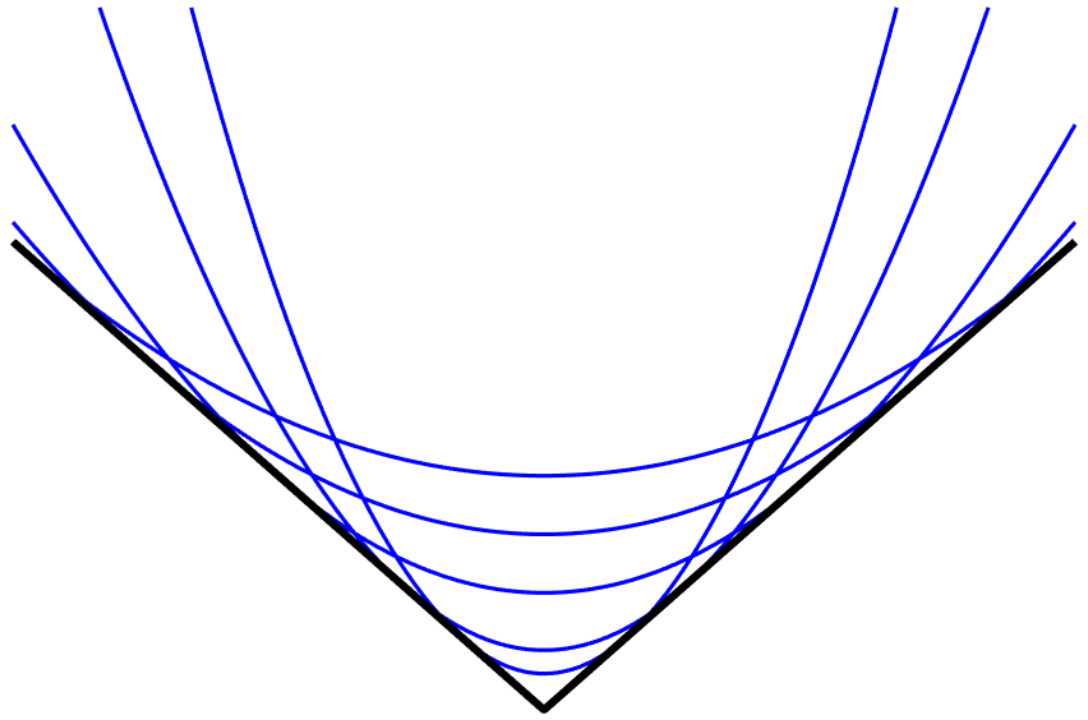


There are a wide range of optimization methods for handling this group Lasso problem.

L-BFGS solver for SR-Lasso (VarPro)

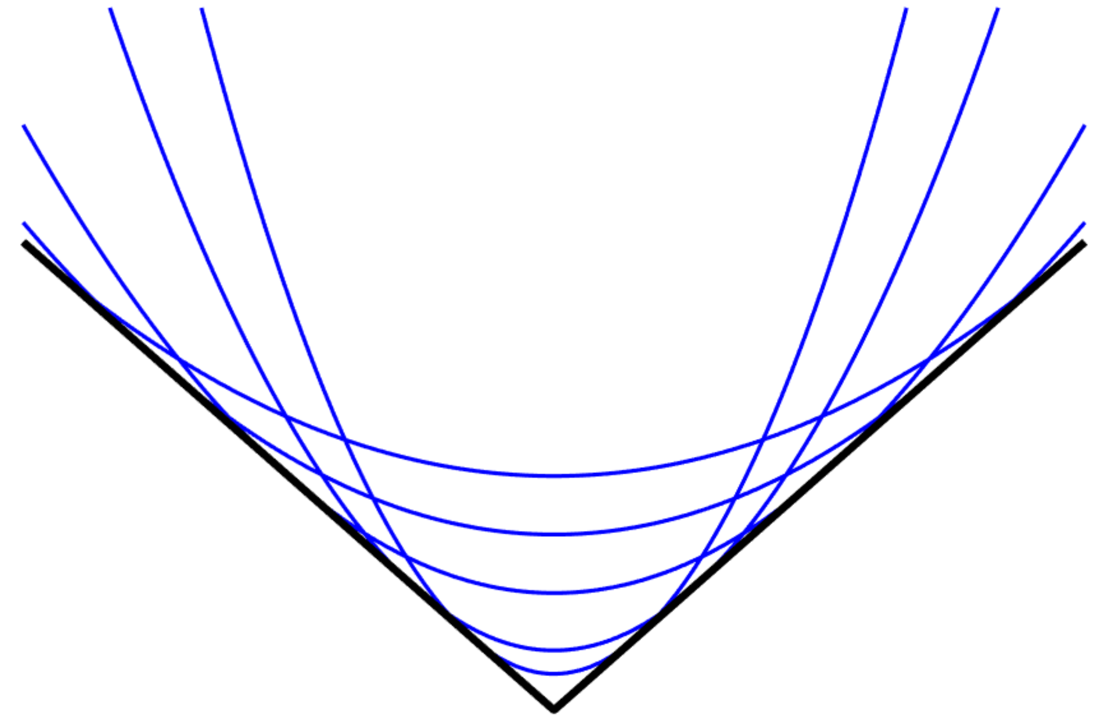
Group Lasso: $$\min_z \frac{1}{2}\|\Gamma z - y\|^2 + \lambda \|z\|_{1,2}$$

$$\sum_{j=1}^{n} \|z_{g_i}\| = \min_{z_{g_i}=v_i u_{g_i}} \frac{1}{2} v_i^2 + \frac{1}{2}\|u_{g_i}\|^2$$



Smooth formulation: $\min_{u,v} G(u,v)$ where $G(u,v) = \frac{1}{2}\|\Gamma(v_i u_{g_i})_i - y\|^2 + \frac{\lambda}{2}\|u\|^2 + \frac{\lambda}{2}\|v\|^2$

L-BFGS solver for SR-Lasso (VarPro)



$$\sum_{j=1}^{n} \|z_{g_i}\| = \min_{z_{g_i}=v_i u_{g_i}} \frac{1}{2} v_i^2 + \frac{1}{2}\|u_{g_i}\|^2$$

Group Lasso: $$\min_z \frac{1}{2}\|\Gamma z - y\|^2 + \lambda \|z\|_{1,2}$$

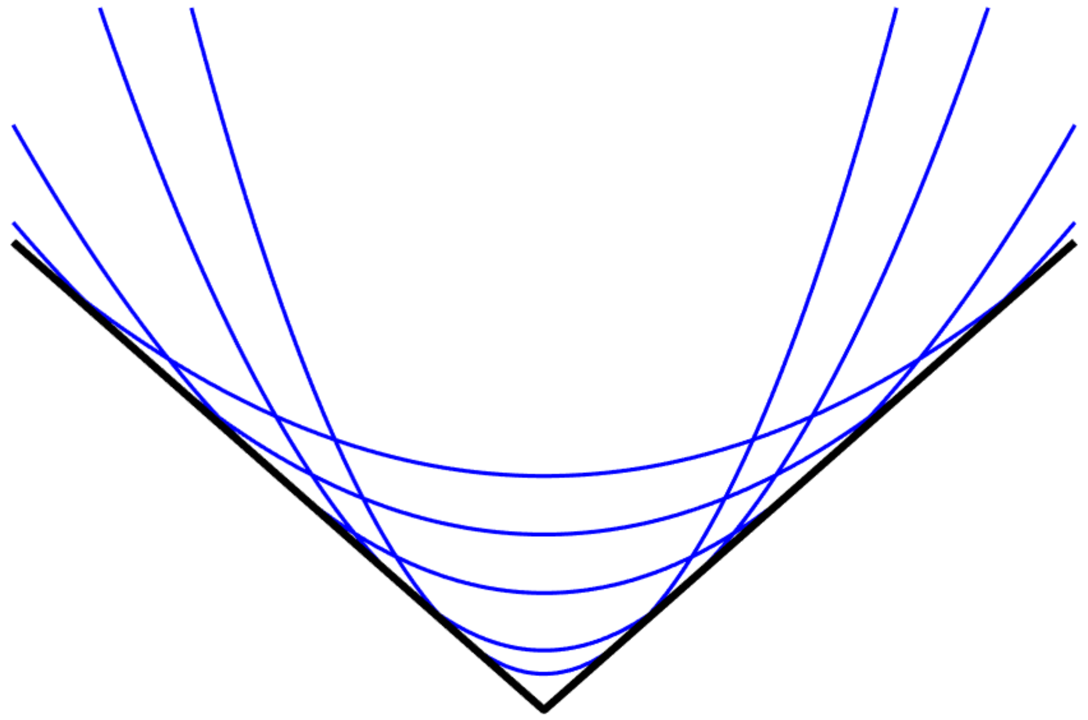
Smooth formulation: $$\min_{u,v} G(u, v) \quad \text{where} \quad G(u, v) = \frac{1}{2}\|\Gamma(v_i u_{g_i})_i - y\|^2 + \frac{\lambda}{2}\|u\|^2 + \frac{\lambda}{2}\|v\|^2$$

"VarPro": $$\min_v f(v) \quad \text{where} \quad f(v) = \min_u \frac{1}{2}\|\Gamma(v_i u_{g_i})_i - y\|^2 + \frac{\lambda}{2}\|u\|^2 + \frac{\lambda}{2}\|v\|^2$$

Ref. Golub and Pereyra. "Differentiation of pseudoinverses, separable nonlinear least square problems and other tales." Academic Press. (1976)

Poon, Clarice, and Gabriel Peyré. "Smooth bilevel programming for sparse regularization." *Neurips* (2021)

L-BFGS solver for SR-Lasso (VarPro)



$$\sum_{j=1}^{n} \|z_{g_i}\| = \min_{z_{g_i}=v_i u_{g_i}} \frac{1}{2} v_i^2 + \frac{1}{2}\|u_{g_i}\|^2$$

Group Lasso: $\min_{z} \frac{1}{2}\|\Gamma z - y\|^2 + \lambda \|z\|_{1,2}$

Smooth formulation: $\min_{u,v} G(u,v)$ where $G(u,v) = \frac{1}{2}\|\Gamma(v_i u_{g_i})_i - y\|^2 + \frac{\lambda}{2}\|u\|^2 + \frac{\lambda}{2}\|v\|^2$

"VarPro": $\min_{v} f(v)$ where $f(v) = \min_{u} \frac{1}{2}\|\Gamma(v_i u_{g_i})_i - y\|^2 + \frac{\lambda}{2}\|u\|^2 + \frac{\lambda}{2}\|v\|^2$

One can prove that all saddle points are strict
(gradient descent always converge to global min)

Apply L-BFGS to this smooth function!

Ref. Golub and Pereyra. "Differentiation of pseudoinverses, separable nonlinear least square problems and other tales." Academic Press. (1976)

Poon, Clarice, and Gabriel Peyré. "Smooth bilevel programming for sparse regularization." *Neurips* (2021)

Some properties of the Hadamard Parametrization

All stationary points of f are either global minima or **strict saddles** ($\nabla^2 f$ has at least one negative eigenvalue).

Lee et al (2017): Gradient descent almost always avoid strict saddles.

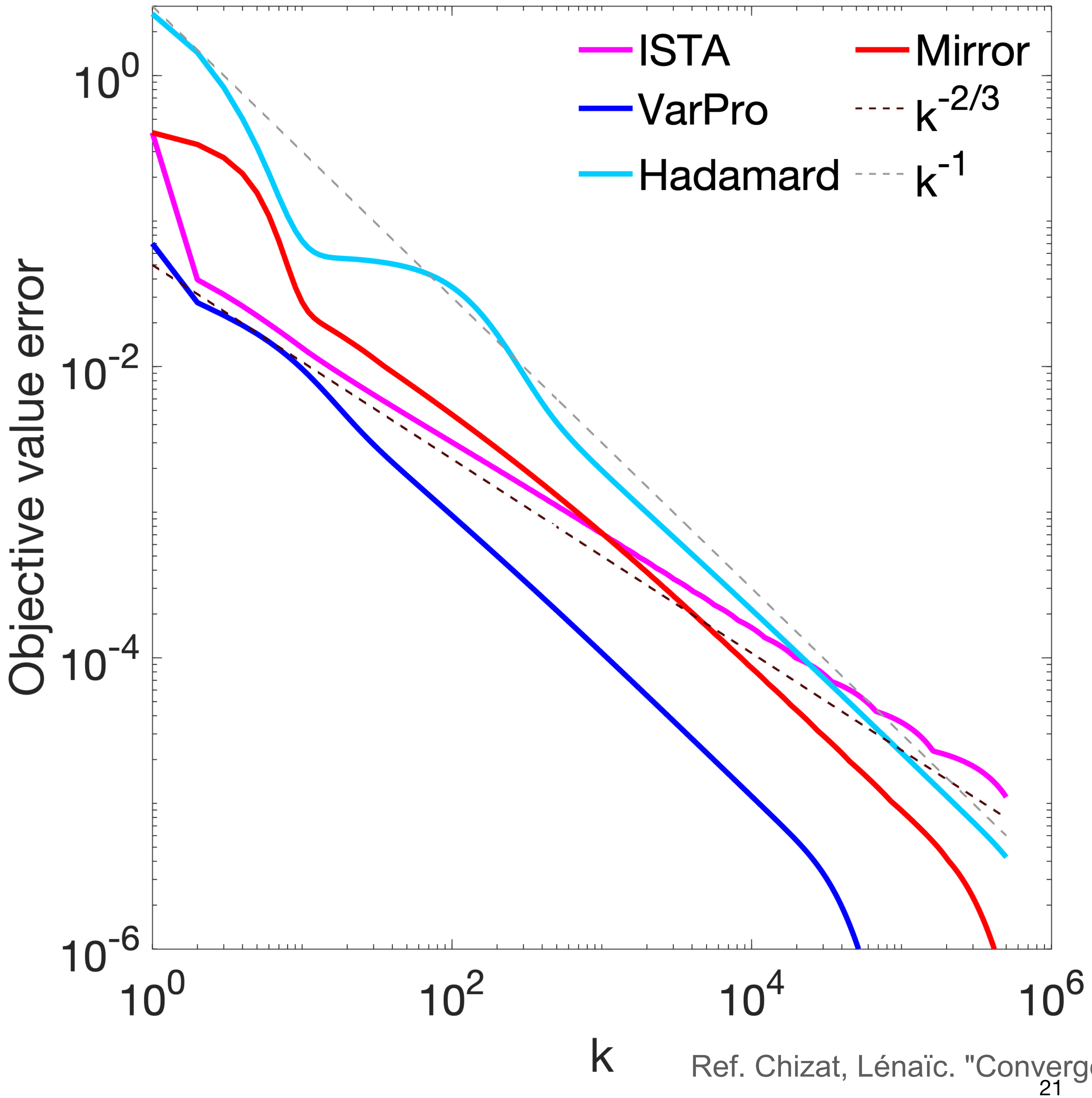
Well known: for $f(v) = \min_u G(u, v)$, $\nabla^2 f$(v) is the Schur complement of $\nabla^2 G(u, v)$ and it is always no worse conditioned.

In the case of the Lasso: $\dfrac{\mathrm{Cond}(\nabla^2 f)}{\mathrm{Cond}(\nabla^2 G)} = \mathcal{O}(\lambda)$

Lipschitz constant of ∇f is independent of discretisation of Γ.

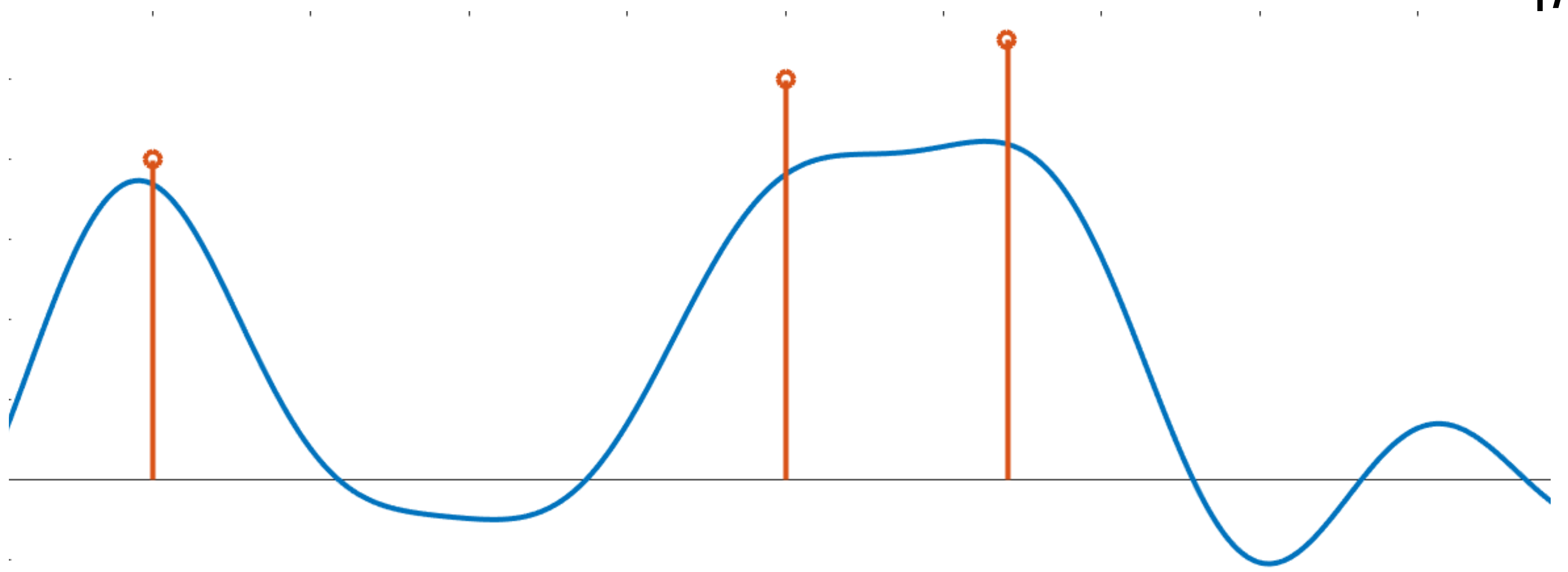
Ref. Poon & Peyré Smooth over-parameterized solvers for non-smooth structured optimization. *Math. Prog. (2023)*

Observation



Solving the Lasso with a discretised Fourier operator ($n = 500$):

$$\text{Column i}: \quad \Gamma_i = \left(\exp(2\pi\sqrt{-1}ik/n)\right)_{|k|\leq m}$$



Observations:

- ISTA converges at $\mathcal{O}(k^{-2/3})$ while proximal mirror descent converges at $\mathcal{O}(k^{-1})$ as shown by Chizat 2021.
- The Hadamard parameterisations also converge at $\mathcal{O}(k^{-1})$

Ref. Chizat, Lénaïc. "Convergence rates of gradient methods for convex optimization in the space of measures."

Mirror flow interpretation

Hadamard parametrised gradient flow

$$\min_z \lambda \|z\|_1 + F(z) = \min_{u,v} \frac{\lambda}{2}(\|u\|^2 + \|v\|^2) + F(u \odot v)$$

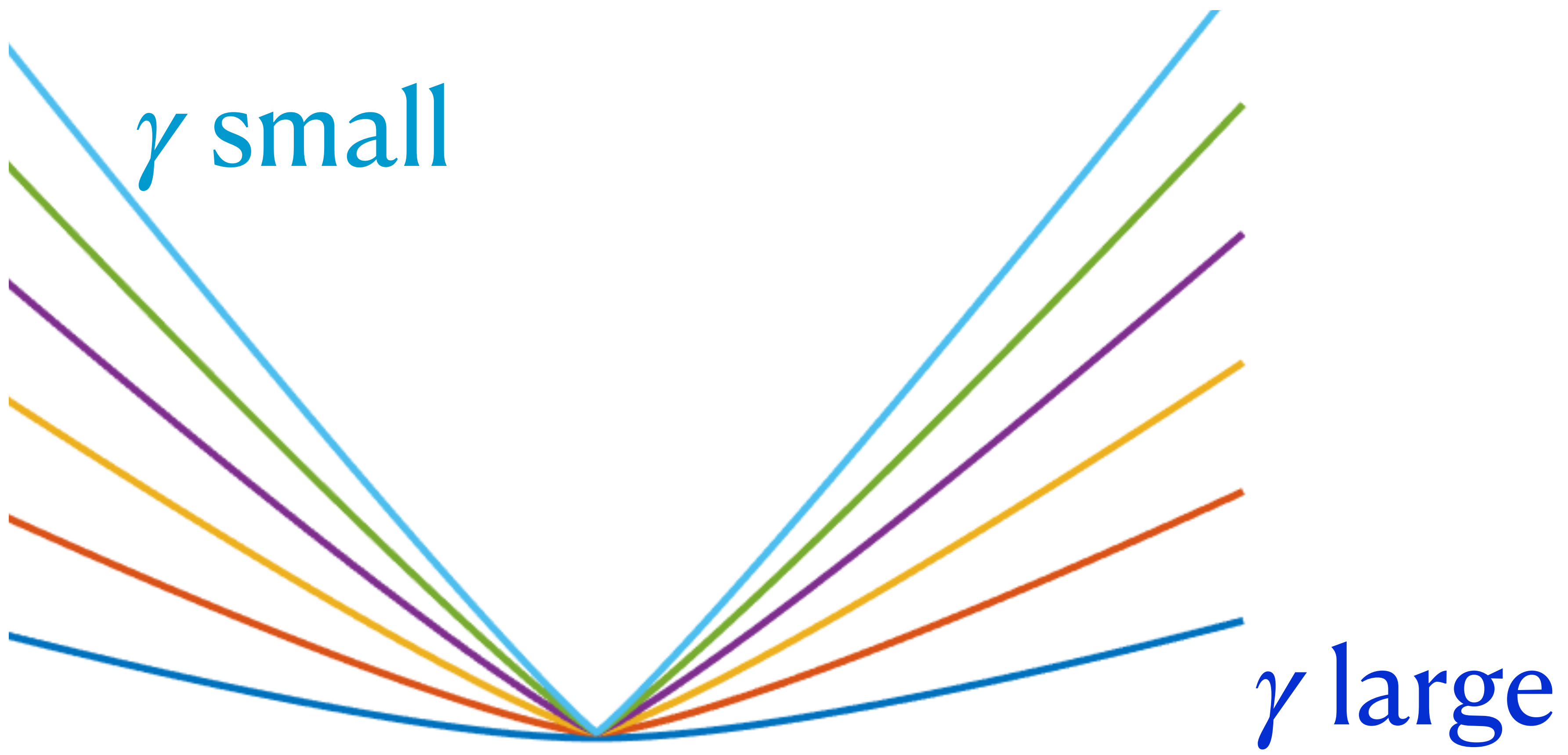
$$\begin{cases} \dot{u}_t = -\tau(\lambda u_t + v_t \nabla F(u_t \odot v_t)) \\ \dot{v}_t = -\tau(\lambda v_t + u_t \nabla F(u_t \odot v_t)) \end{cases}$$

Let $z(t) := u(t) \odot v(t)$, then:

$$\frac{d}{dt} \nabla \eta_{\gamma(t)}(z(t)) = -2 \nabla F(z(t))$$

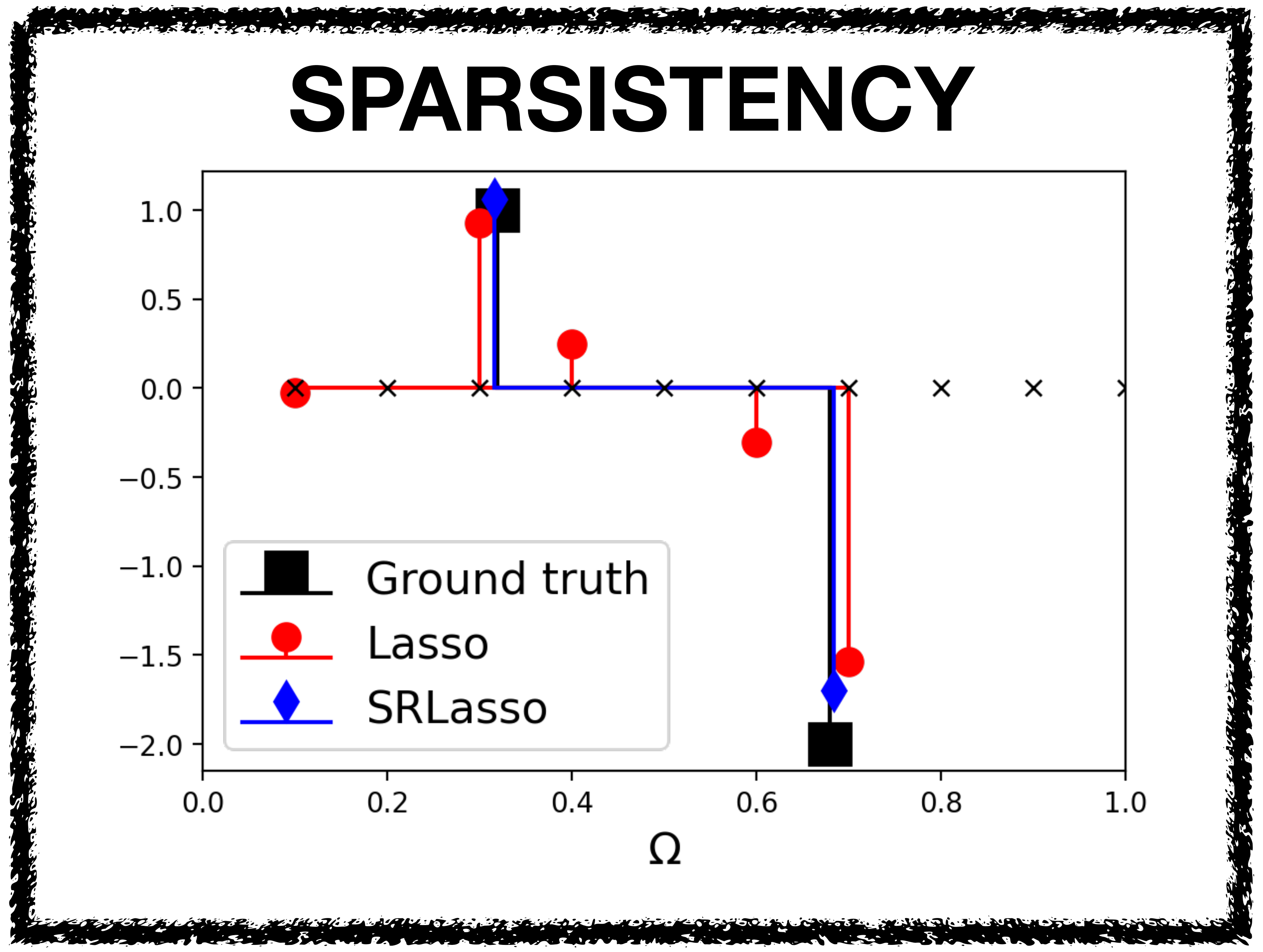
$$\gamma(t) = \frac{1}{2} e^{-2\lambda t} | u(0) - v(0) |$$

$$\eta_\gamma(z) = \gamma \operatorname{arsinh}(z/\gamma) - \sqrt{z^2 + \gamma^2} + \gamma$$



Outline

THE METHOD


SPARSISTENCY
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0
Ω
Ground truth
Lasso
SRLasso

NUMERICS
SR Lasso
cBP
Lasso
0.6
0.5
0.4
0.3
0.2
0.1
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
λ

Sparsistency:

Do we recover the correct number of Diracs?

$$\mu = \sum_{j \in I} a_j \delta_{x_j + t_j}$$

<u>Lasso [Wainwright '08]:</u>

$$\min_z \lambda \|z\|_1 + \frac{1}{2}\|\Gamma z - y\|^2$$

If Diracs are on the grid, need

- $\mathrm{Ker}(\Gamma_I) = \{0\}$.
- define $p_L := \Gamma_I^{*,\dagger} \mathrm{Sign}(a_I)$.

$$\|\Gamma_{I^c}^* p\|_\infty = \max_{j \notin I} |\phi(x_j)^\top p_L| < 1$$

<u>C-BP [Duval & Peyre '17]:</u>

- Let $\Gamma_I := [(\Phi_X)_I, (\Psi_X)_I]$ be injective
- define $p_V = \Gamma_I^{*,\dagger} \begin{pmatrix} 1_N \\ 0_N \end{pmatrix}$ and $\eta(x) = \phi(x)^\top p_V$.

$$\max_{j \notin I} \eta(x_j) \pm \frac{h}{2} \eta'(x_j) < 1$$

This does not hold in general, in particular, fails for translation invariant operators such as Gaussian when grid is too fine.

- Cannot handle $t_j \neq 0$.
- Even if $t_j = 0$, does not hold when grid is too fine [Duval & Peyre '17]

Refined condition for the group-Lasso

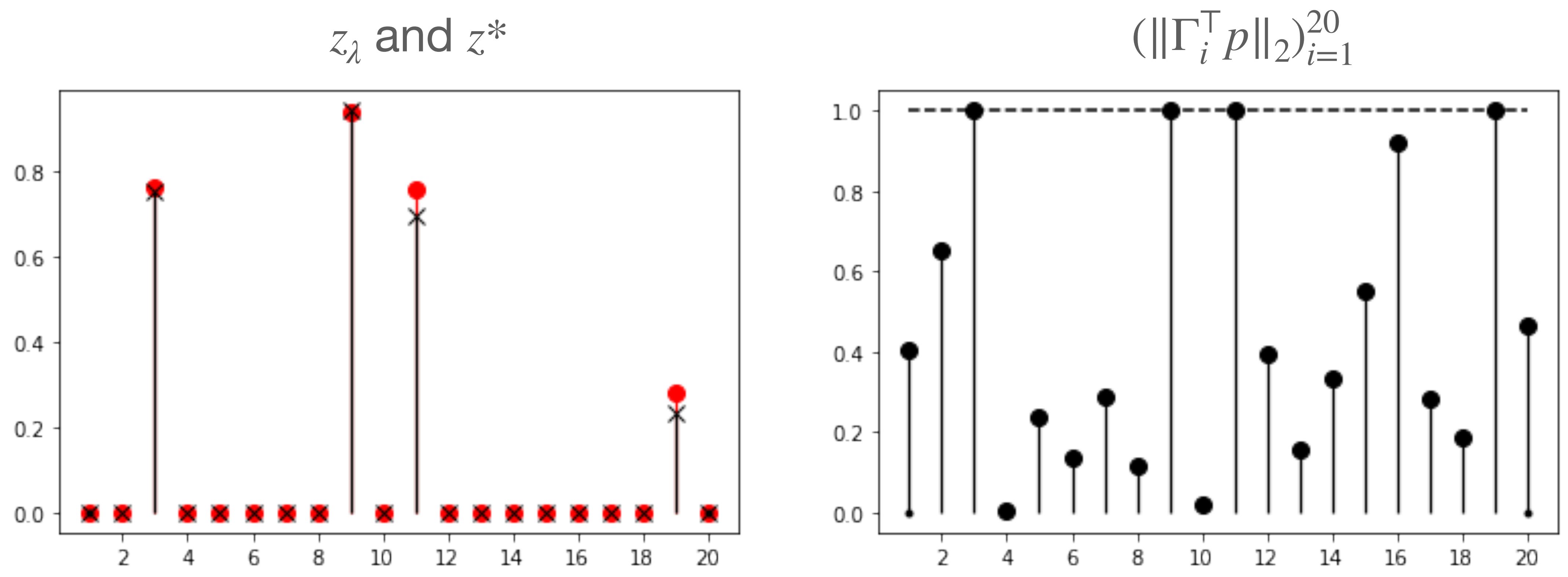
$$\min_z \lambda \sum_i \|z_i\|_2 + \frac{1}{2}\| \sum_i \Gamma_i z_i - y\|^2$$

(N) $\mathrm{Ker}(\Gamma_I) \cap \mathrm{Ker}(Q_{z^*}) = \{0\}$ where

$$Q_z = \mathrm{diag}\left(I - \frac{1}{\|z\|^2} z z^\top\right)$$

(IC) Define $p = \Gamma_I^{*,\dagger}\mathrm{Sign}(z^*)$ where $\mathrm{Sign}(z)_i = \frac{z_i}{\|z_i\|}$.

$$\|\Gamma_{I^c}^* p\|_{\infty,2} = \max_{j \notin I} \|\Gamma_i^\top p\|_2 < 1$$



- $\Gamma \in \mathbb{R}^{4\times 40}$. Each group is size 2.
- True signal is supported on 4 groups, and has sparsity $4 \times 2 = 8$.
- No injectivity restricted to the support, but the signal can be stably recovered via group-Lasso!

*NB: (N) is **necessary** for uniqueness of group Lasso. Ref. Fadili, Nghia, and Tran. "Sharp, strong and unique minimizers for low complexity robust recovery." Information and Inference (2023)*

Refined condition for the group-Lasso

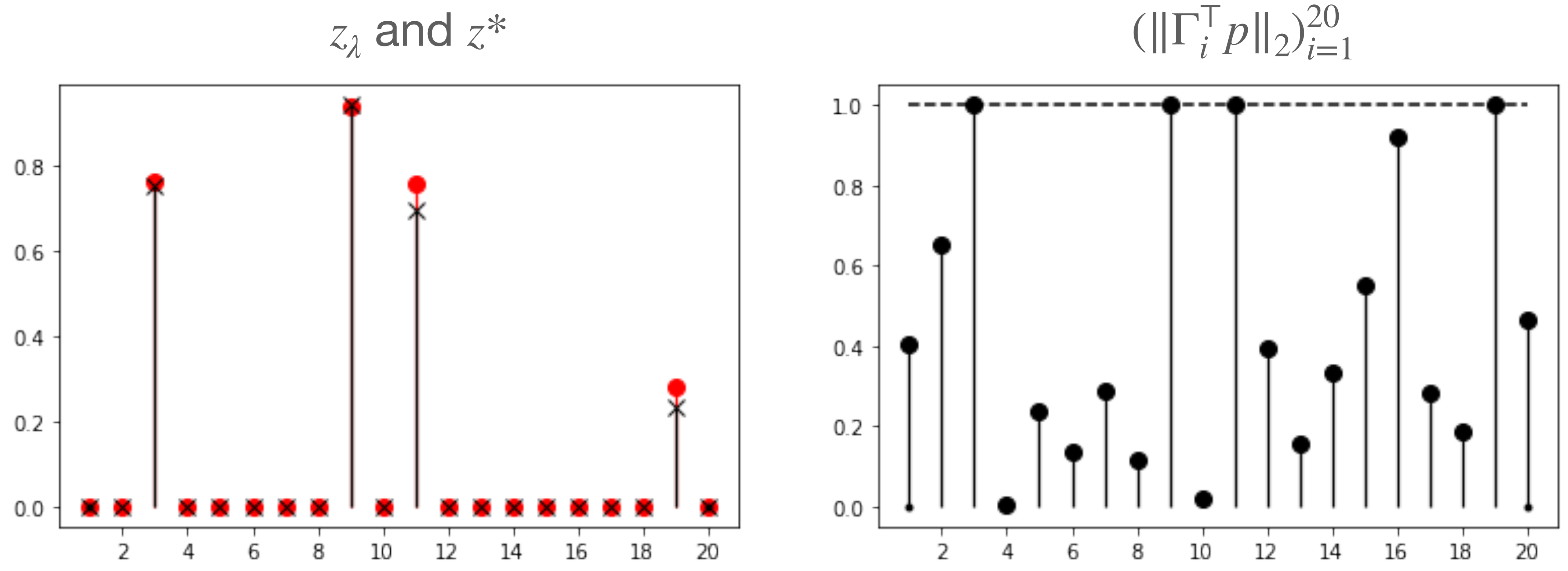
$$\min_z \lambda \sum_i \|z_i\|_2 + \frac{1}{2}\|\sum_i \Gamma_i z_i - y\|^2$$

(N) $\mathrm{Ker}(\Gamma_I) \cap \mathrm{Ker}(Q_{z^*}) = \{0\}$ where

$$Q_z = \mathrm{diag}\left(I - \frac{1}{\|z\|^2} zz^\top\right)$$

(IC) Define $p = \Gamma_I^{*,\dagger}\mathrm{Sign}(z^*)$ where $\mathrm{Sign}(z)_i = \dfrac{z_i}{\|z_i\|}$.

$$\|\Gamma_{I^c}^* p\|_{\infty,2} = \max_{j \notin I} \|\Gamma_i^\top p\|_2 < 1$$



- $\Gamma \in \mathbb{R}^{4\times 40}$. Each group is size 2.
- True signal is supported on 4 groups, and has sparsity $4 \times 2 = 8$.
- No injectivity restricted to the support, but the signal can be stably recovered via group-Lasso!

Theorem [Poon & Peyre '23]:

Under assumptions (N) and (IC), if $y = \Gamma z^* + w$, for all $\|w\|/\lambda$ and λ sufficiently small, there is a unique solution $z_{\lambda,w}$ with support I and $\|z_{\lambda,w} - z^*\| = \mathcal{O}(\lambda)$

*NB: (N) is **necessary** for uniqueness of group Lasso. Ref. Fadili, Nghia, and Tran. "Sharp, strong and unique minimizers for low complexity robust recovery." Information and Inference (2023)*

Sparsistency for SR-Lasso

Ground truth: $\mu = \sum_{j=1}^{s} a_j \delta_{u_j}$ where $u_j = x_{k_j} + t_j$ with $I := \{k_j\}$

$$\Phi\mu \approx \Gamma_I \begin{pmatrix} a \\ b \end{pmatrix} + \mathcal{O}(\|t\|_\infty^2) \text{ where } b_j = a_j \tau^{-1} \|\phi'(x_j)\|$$

Certificate

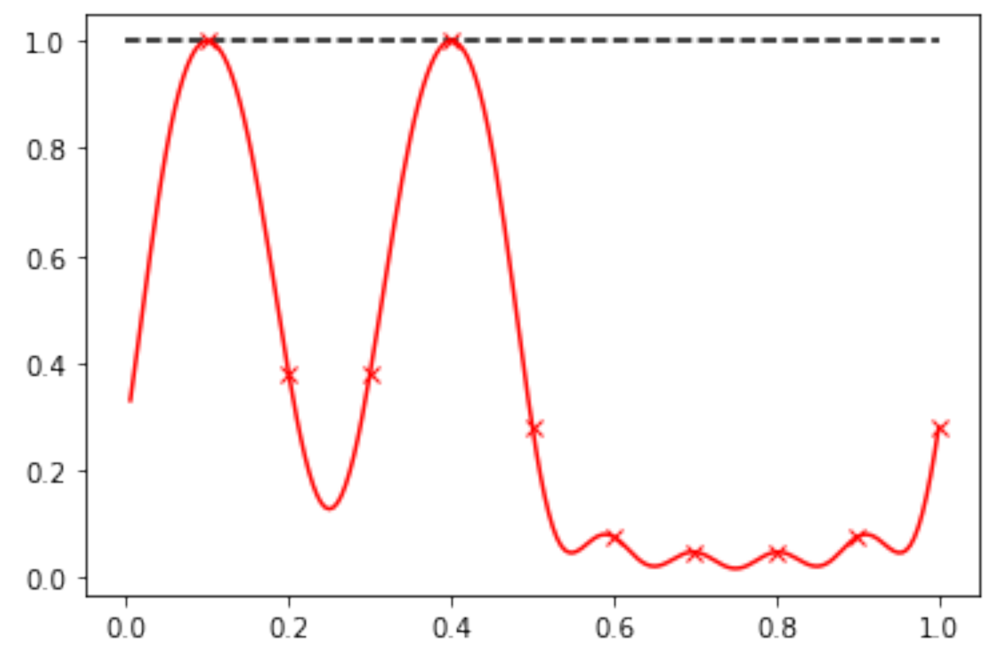
$f(x) = |\phi(x)^\top p|^2 + \tau^2 |\psi(x)^\top p|^2$ where

$p = \Gamma_I^{*,\dagger} \mathrm{Sign}((a,b))$

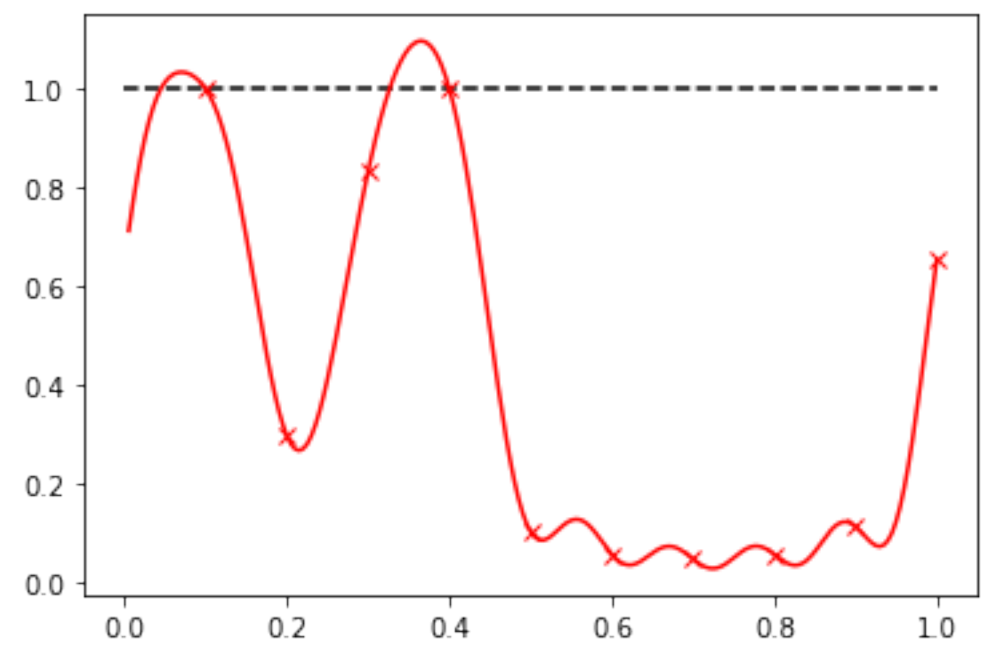
Condition for sparsistency

- for all $j \in I,\ f(x_j) = 1$
- for all $k \notin I,\ f(x_k) < 1$

If b = 0 (i.e. on the grid), then the condition holds under sufficient separation of $\{u_j\}$.



In general, $f(x) > 1$ near x_i for $i \in I$. Choose grid sufficiently coarse (depends on the shift) to have a non-degenerate certificate.



Translation invariant case

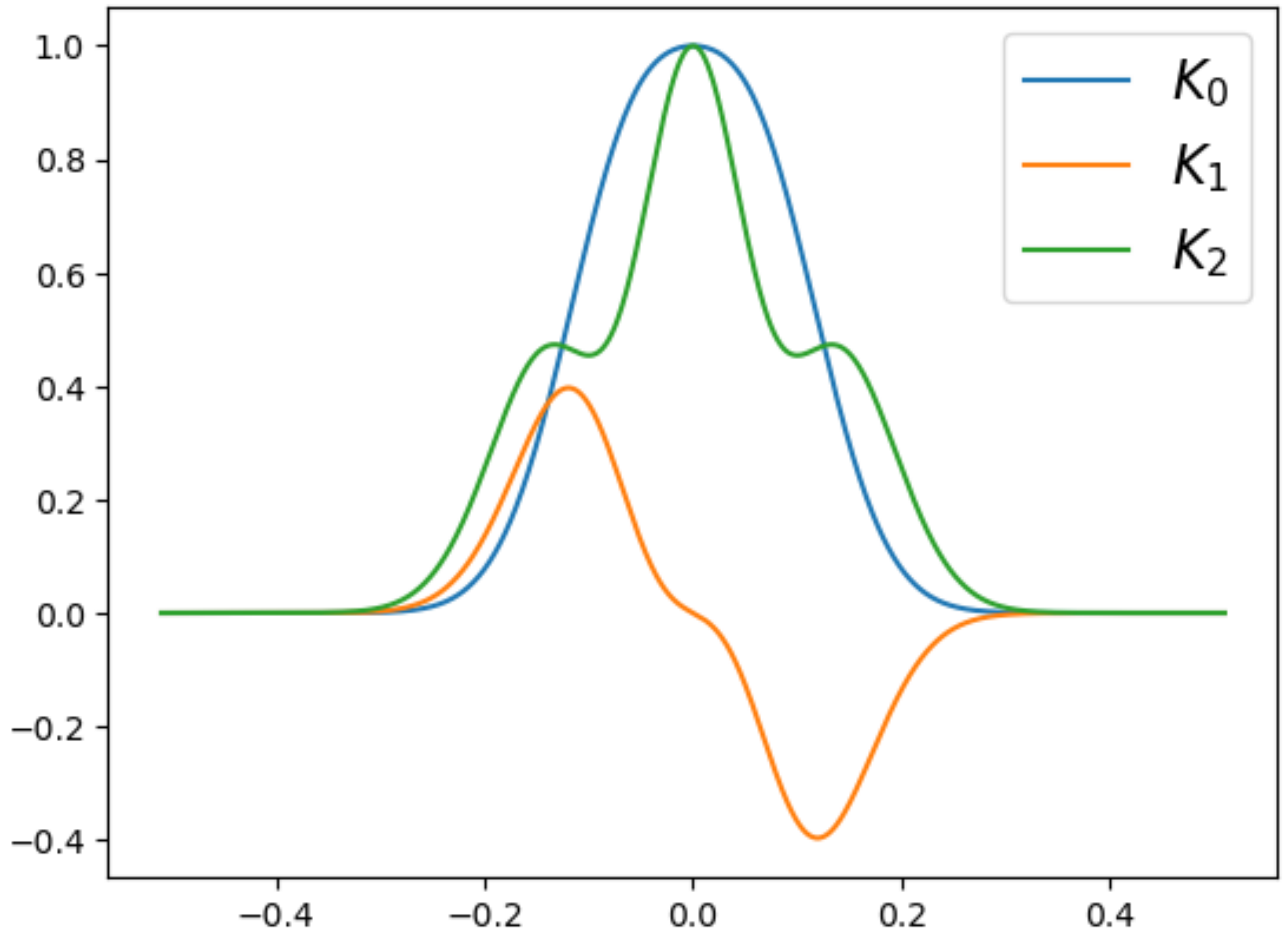
Minimum separation: $\Delta_{\min} := \min_{i\neq j} |u_j - u_i|$ $\quad 1-\tau^2 \gtrsim \sup\{ \sum_{i=1}^{s} |\tilde\kappa_\ell(z_0 - z_i)|;\ \min_{i\neq j}|z_i - z_j| \geq \frac{1}{2}\Delta_{\min}\}$

Kernel functions :

$$K_0(x) = \kappa(x)^2 + \tau^2 \tilde\kappa_1(x)^2$$

$$K_1(x) = \tilde\kappa_1(x)(\kappa(x) + \tau^2\tilde\kappa_2(x))$$

$$K_2(x) = \tilde\kappa_2(x) + \tau^{-2}\tilde\kappa_1(x)^2$$



$\forall |x| \leq r$

$K_0''(x) - \gamma K_1''(x) < K_0''(0)$

and $\quad K_2''(x) < K_2''(0)$

$\forall |x| \geq r$

$|K_0(x) - \gamma K_1(x)| < 1$

and $\quad |K_2(x)| < 1$

Theorem [Poon & Peyré ’23]:

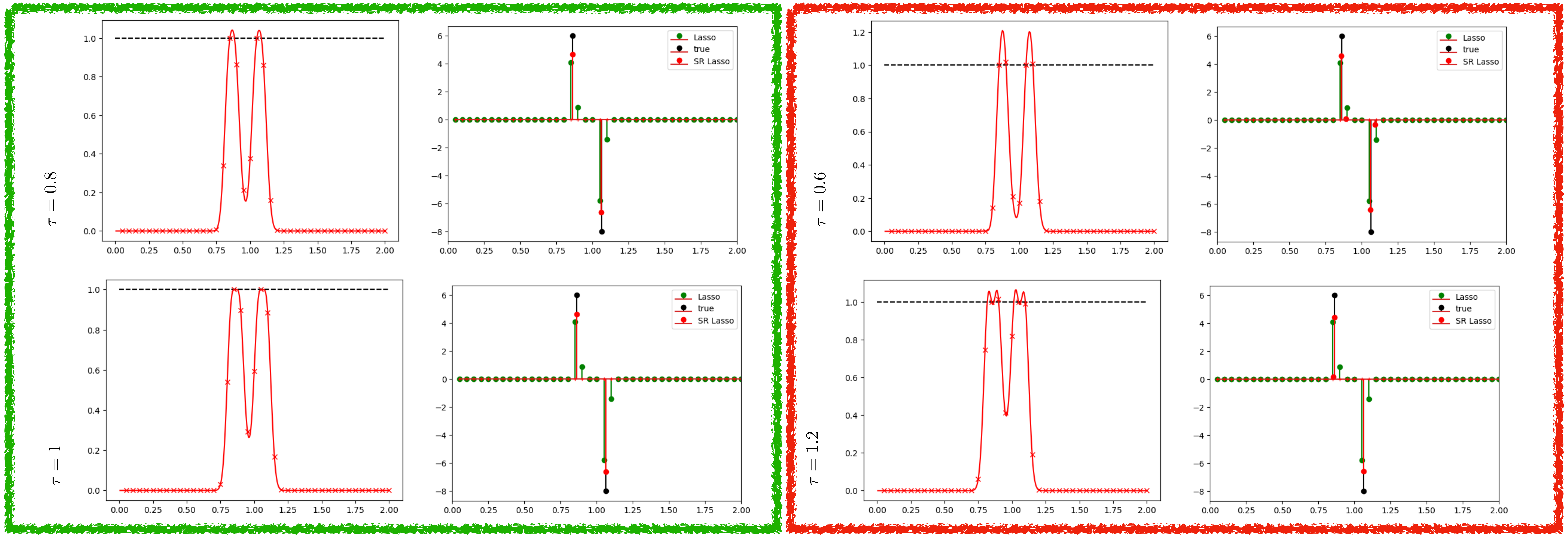
Stable support recovery provided that $t \lesssim h(1-\tau^2)$ and $\lambda \sim (1-\tau^2)h\sqrt{\kappa''(0)}$

Gaussian case $\phi(x) = \exp(-x^2/\sigma^2)$

Theorem [Poon & Peyre ’23]

Let $\tau \in (0.8,1)$. Constant is $C \sim 1 - \tau^2$. Then, SR-Lasso recovers s Diracs stably if

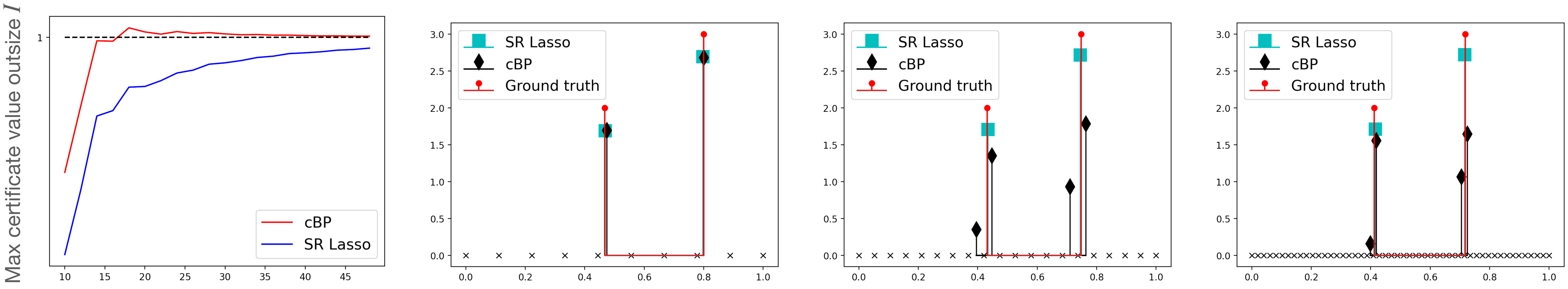
$$\min_{i \neq j} |u_j - u_i| \gtrsim \sigma\sqrt{|\log(C)|}, \quad |t_j| \leq C\min(h, \sigma), \;\; \lambda \sim Ch/\sigma$$



Distance between spikes is 4σ, $h = \sigma$ and spikes are $0.25h$ inside the grid.

Comparison against certificate for c-BP

Recovering 2 spikes from samples of Gaussian convolution $\phi(x) = (\exp(-(x - x_j)^2/\sigma)_{j \in [m]}$



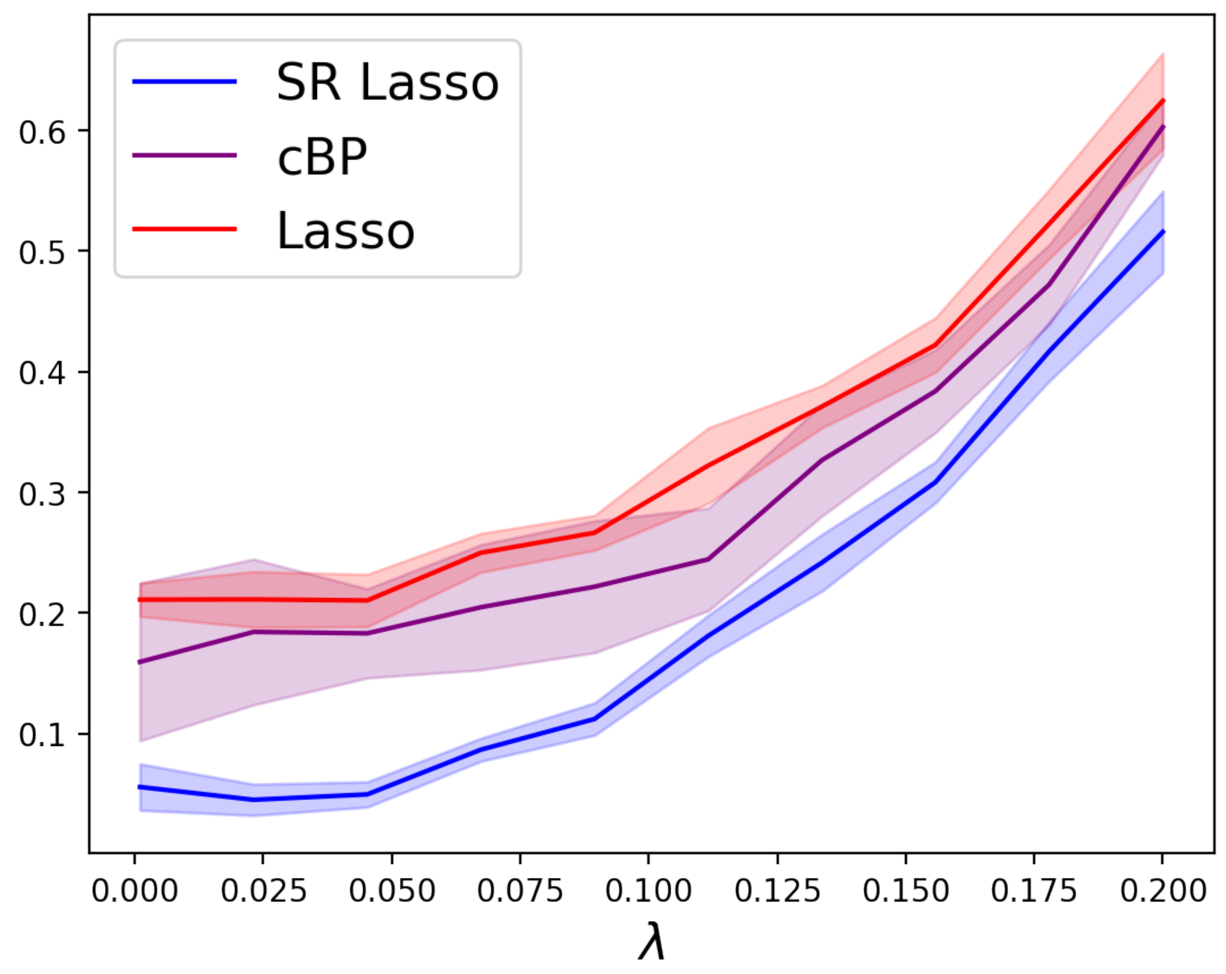
N $\quad N = 10 \quad N = 20 \quad N = 40$

Outline

THE METHOD

SPARSISTENCY

NUMERICS



A few numerical results

<u>Evaluation metric MMD.</u> Given a measure μ and a function $k(x, y)$,

$$\mathrm{MMD}_k^2(\mu) := \int k(x, y) d\mu(x) d\mu(y)$$

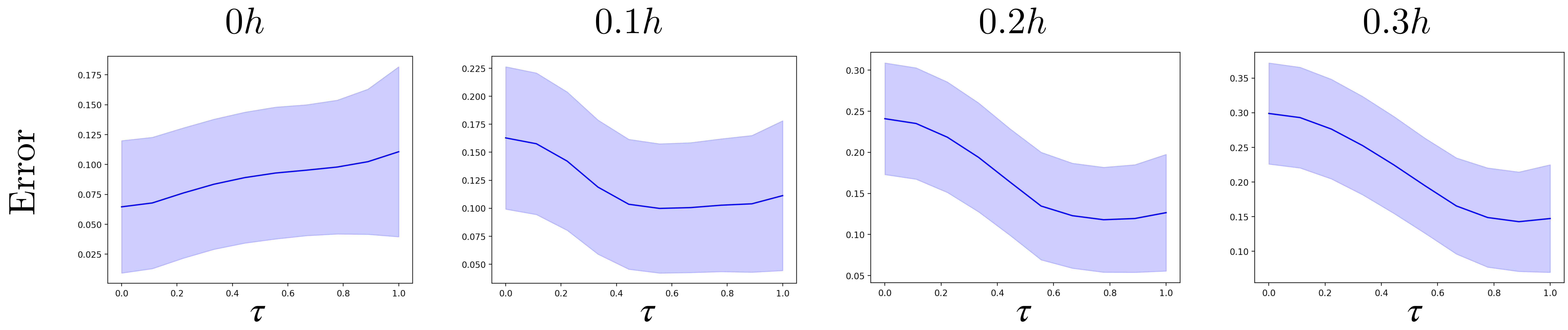
We use the *Laplace kernel* $k(x, y) = \exp(-\|x - y\|)$ and consider

$$\boxed{D(\mu, \nu) = \|\mu - \nu\|_k^2}$$

NB: Our loss term can be seen as minimising MMD with kernel $k(x, y) := \phi(x)^\top \phi(y)$.

Laplace kernel: slow Fourier decay—> measures the match on 'higher order moments'.

Impact of τ



$\phi(x) = (\exp(2\pi i k x))_{|k| \leq f_c}$

Result averaged over 10 random noisy instantiations $y = \Phi\mu_0 + \text{noise}$.
Images on 2nd row is showing one noisy instantiation for visualisation of the reconstructions.

1D Gaussian sampling (positive signs)

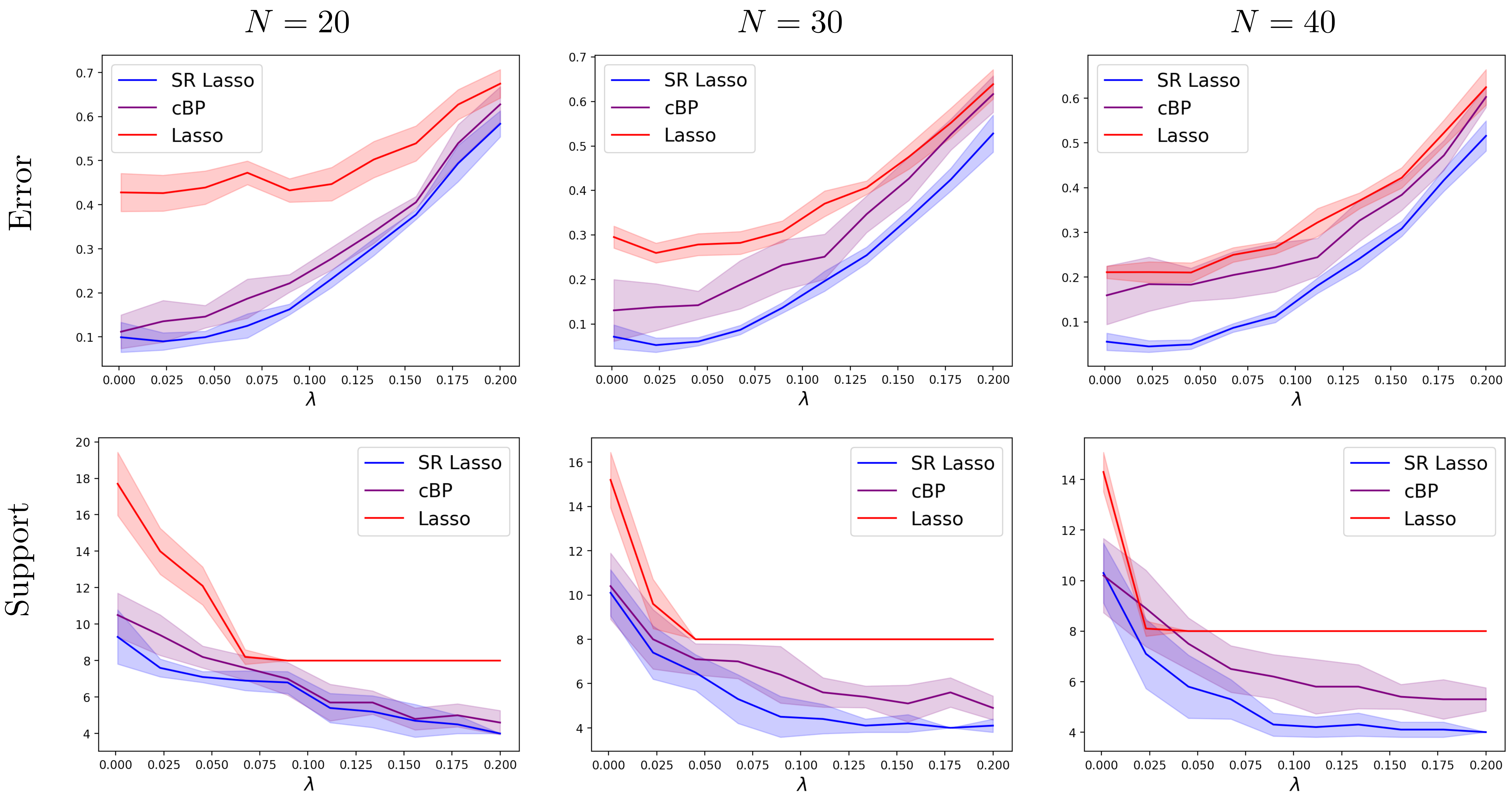


Figure 9: 1D comparison for the recovery of 4 positive spikes in the case of Gaussian sampling.

2D Fourier sampling (complex signs)

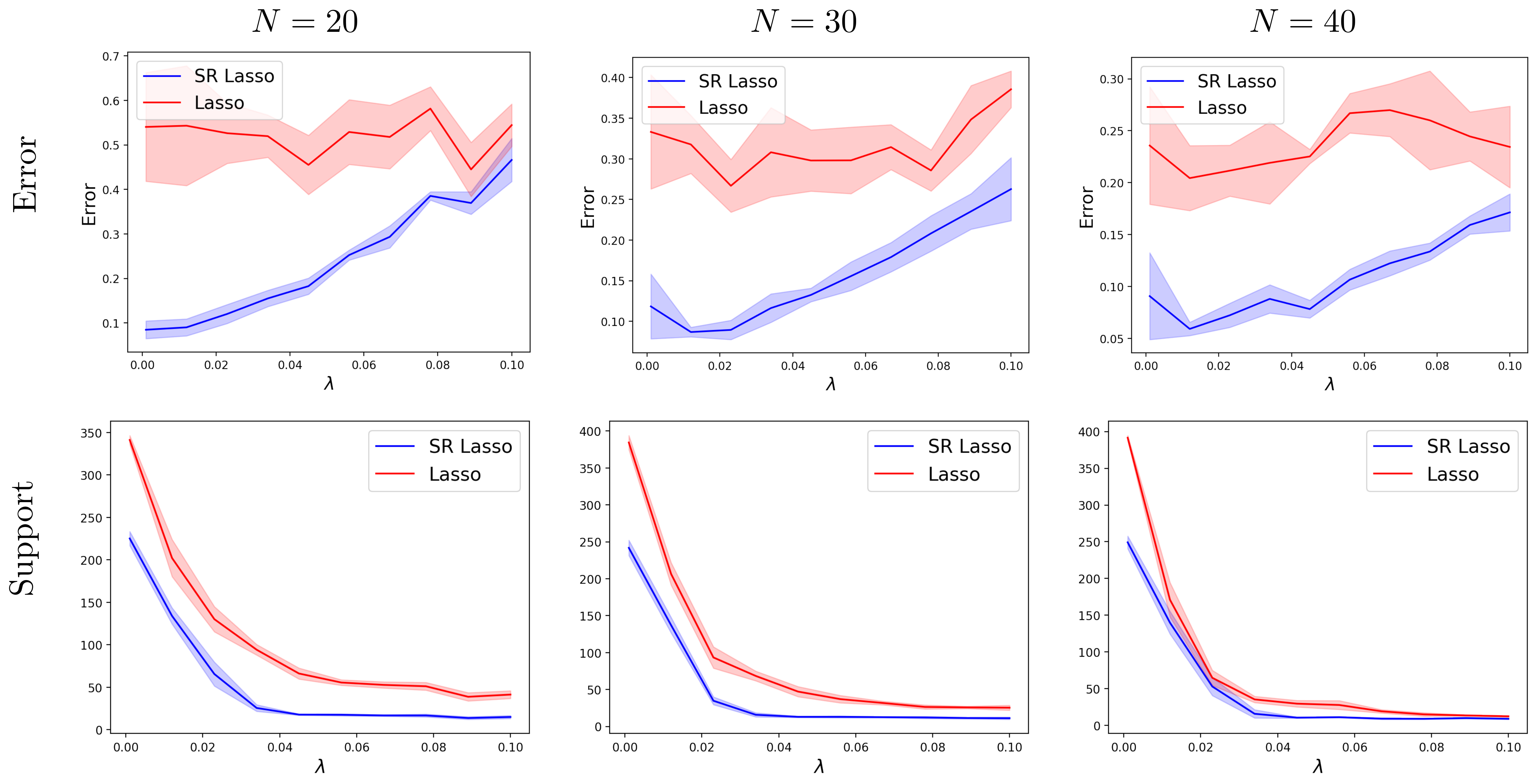


Figure 10: 2D comparison for the recovery of 3 signed spikes in the case of Fourier sampling.

3D “Microscopy example”

$$\phi(\Omega, R) = \left(\exp\left(- \frac{\|\Omega_i - \Omega\|^2}{(2\sigma^2)} \right) \exp(-R_j R) \right)_{(i,j)\in[M]\times[T]}$$

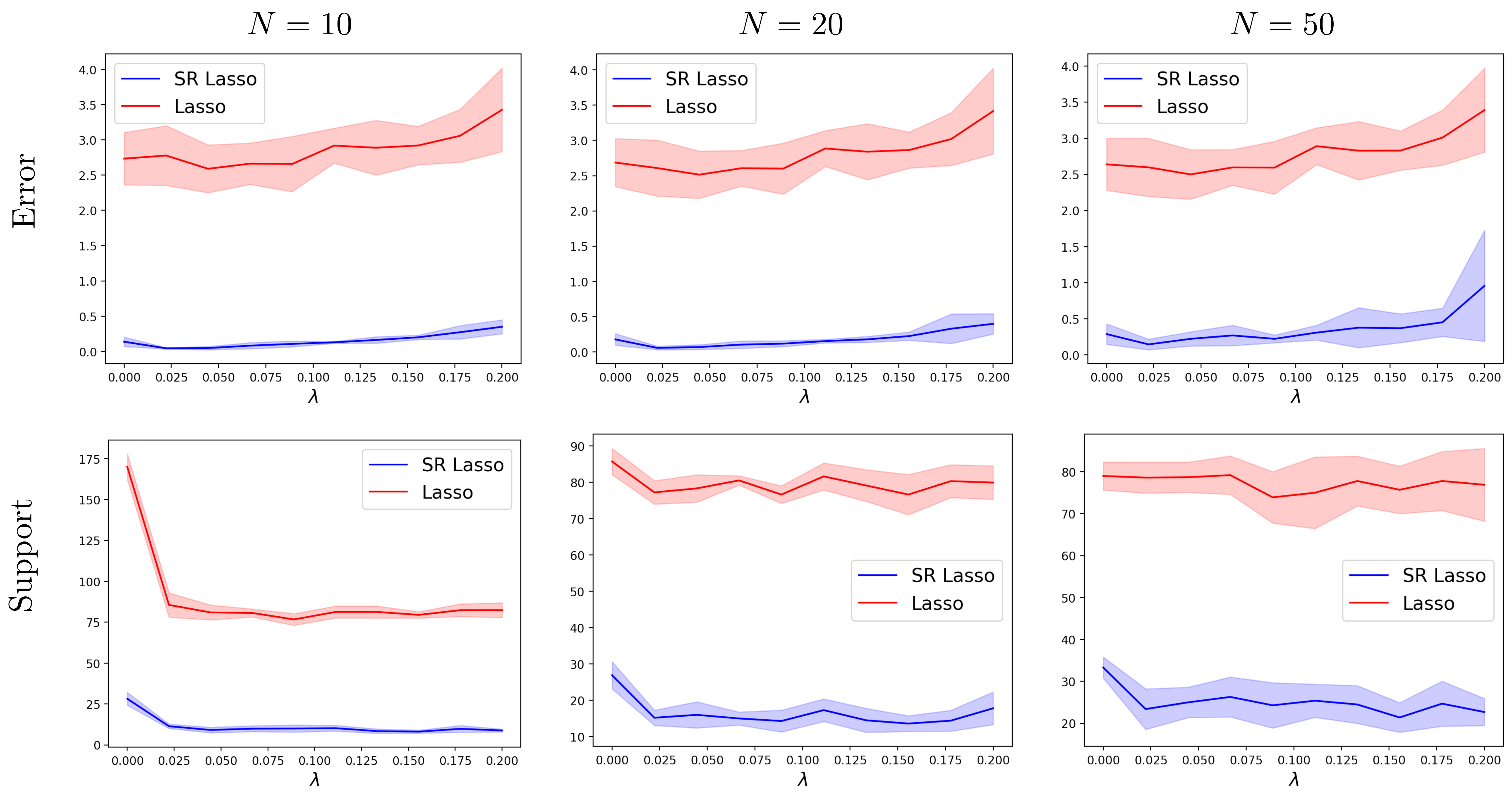


Figure 11: 3D comparison for the recovery of 3 signed spikes in the case of Gaussian-Laplace sampling.

Summary

- SR-Lasso is a 'half-way' option for resolving sparse sum of Diracs. Only require evaluation of ϕ and ϕ' on a fixed grid.
- Inspired by C-BP but can handle arbitrary signs and dimensions.
- Able to recover shifts that are $\mathcal{O}(h)$ while retaining sparsistency.
- Since this is a standard group-Lasso problem, we can exploit existing solvers. We advocate the use of VarPro with L-BFGS.