Geometry & Symmetry in Short-and-Sparse Deconvolution

Han-Wen (Henry) Kuo

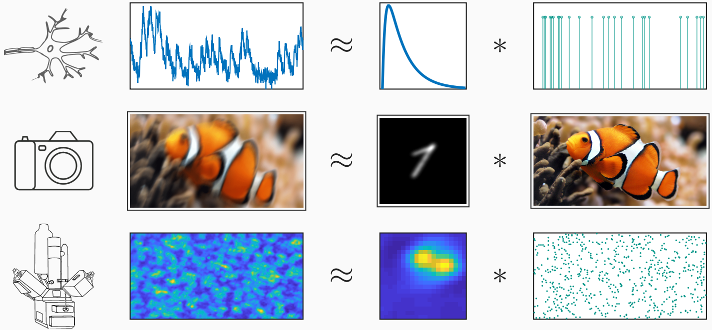
Aug, 07, 2019

Signal with Repeating Short Pattern

SIGNALS CONTAINING **SHORT** **REPEATED** PATTERN:

Signal with Repeating Short Pattern

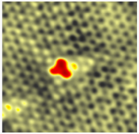
SIGNALS CONTAINING **SHORT** **REPEATED** PATTERN:



Short-and-Sparse Signals

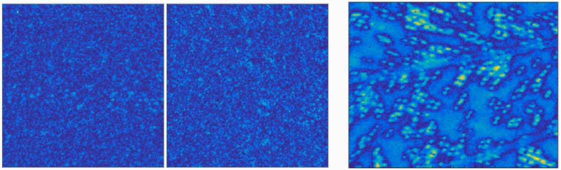
DEFECTS IN CRYSTAL LATTICE FROM STM SIGNAL

Defect signature effects material properties (superconductivity, semiconductivity, etc..)



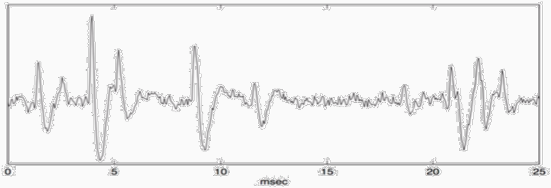
Doped Graphine

REPEATING DEFECTS



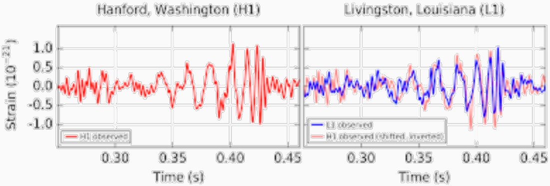
Short-and-Sparse Signals

TEMPORAL PATTERN IN SPIKE SORTING & CALCIUM IMAGING



Neurons transmit information via firing pattern

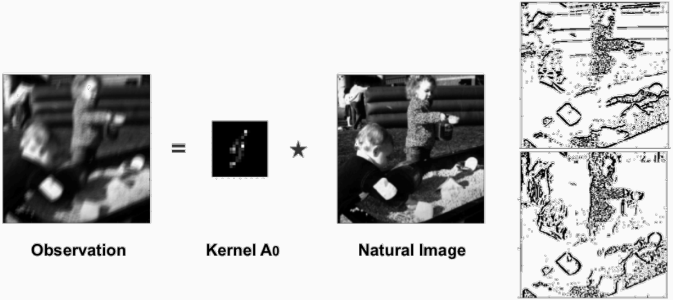
EVENT PATTERN IN LIGO



Black hole merger has characteristic gravitational wave

Short-and-Sparse Signals

Image Deblurring



- Small blurring kernel
- Sparse image gradient

Short-and-Sparse Deconvolution (SaSD) Model

ANALYSIS SETTING:

GIVEN **OBSERVATION** $\boldsymbol{y} = \boldsymbol{a}_0 * \boldsymbol{x}_0 \in \mathbb{R}^n$, $p \ll n$

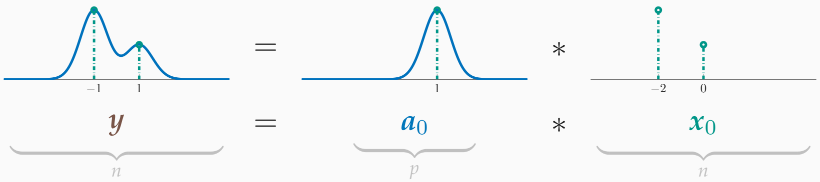
DECONVOLVE **SHORT** $\boldsymbol{a}_0 \in \mathbb{R}^p$ AND **SPARSE** $\boldsymbol{x}_0 \in \mathbb{R}^n$ SIGNALS

Short-and-Sparse Deconvolution (SaSD) Model

ANALYSIS SETTING:

GIVEN **OBSERVATION** $y = a_0 * x_0 \in \mathbb{R}^n$, $p \ll n$

DECONVOLVE **SHORT** $a_0 \in \mathbb{R}^p$ AND **SPARSE** $x_0 \in \mathbb{R}^n$ SIGNALS



In analysis the convolution $*$ is circular†

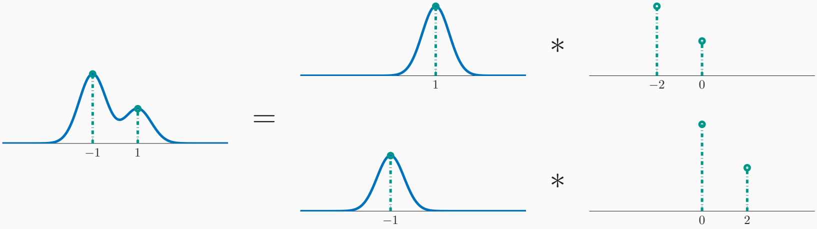
† In practice it can be either circular or direct

Symmetric Solutions in SaSD

ALL **SHIFTED** & **SCALED** (a_0, x_0) ARE SOLUTIONS

Symmetric Solutions in SaSD

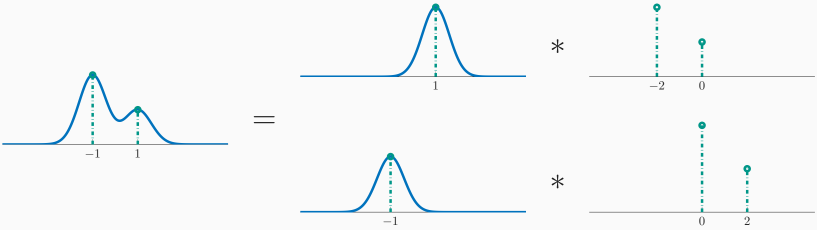
All shifted & scaled $(\boldsymbol{a}_0, \boldsymbol{x}_0)$ are solutions



$$\boldsymbol{y} = \alpha\, s_{\ell}[\boldsymbol{a}_0] * (1/\alpha)\, s_{-\ell}[\boldsymbol{x}_0]$$

Symmetric Solutions in SaSD

All shifted & scaled (a_0, x_0) are solutions



$$y \quad = \quad \alpha\, s_\ell[a_0] \quad * \quad (1/\alpha)\, s_{-\ell}[x_0]$$

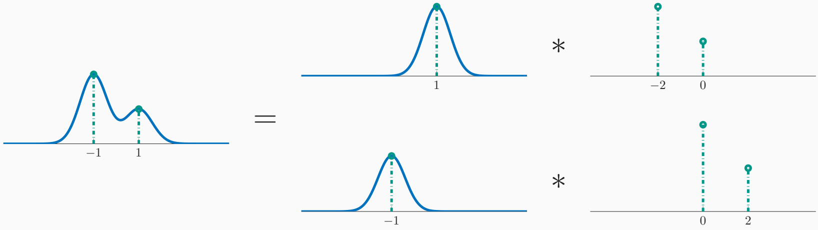
$s_i[a_0] \in \mathbb{R}^{3p}$ is shift of a_0 by ℓ indices†:

$$s_\ell[a_0] \;=\; [\,\underbrace{0,\ldots,0}_{p+\ell},\; a_0\,,\; \underbrace{0,\ldots,0}_{p-\ell}\,]$$

† In analysis $s_\ell[x_0]$ is circular shift of x_0 by ℓ

Symmetric Solutions in SaSD

All shifted & scaled $(\boldsymbol{a}_0, \boldsymbol{x}_0)$ are solutions



$$\boldsymbol{y} = \alpha\, s_\ell[\boldsymbol{a}_0] * (1/\alpha)\, s_{-\ell}[\boldsymbol{x}_0]$$

We have many possible solutions ... but it is ok!

Symmetric Solutions in SaSD

All **shifted** & **scaled** $(\boldsymbol{a}_0, \boldsymbol{x}_0)$ are solutions

$$\boldsymbol{y} = \alpha\, s_\ell[\boldsymbol{a}_0] * (1/\alpha)\, s_{-\ell}[\boldsymbol{x}_0]$$

We have many possible solutions ... but it is ok!

Find $(\widehat{\boldsymbol{a}}, \widehat{\boldsymbol{x}})$ as SaSD solution where:

- Fix scale $\|\widehat{\boldsymbol{a}}\|_2 = 1$
- Accept every signed shift $\widehat{\boldsymbol{a}} = \pm s_\ell[\boldsymbol{a}_0]$ as solution

Algorithm: Bilinear Lasso

NATURAL, EFFECTIVE ALGORITHM—**BILINEAR LASSO**

$$\min_{a \in \mathbb{S}^{3p-1},\, x \in \mathbb{R}^n} \underbrace{\lambda \|x\|_1}_{\text{sparsity surrogate}} + \underbrace{\frac{1}{2}\|a * x - y\|_F^2}_{\text{data fidelity}}$$

Algorithm: Bilinear Lasso

NATURAL, EFFECTIVE ALGORITHM—**BILINEAR LASSO**

$$\min_{\boldsymbol{a}\in\mathbb{S}^{3p-1},\ \boldsymbol{x}\in\mathbb{R}^n} \underbrace{\lambda \|\boldsymbol{x}\|_1}_{\text{sparsity surrogate}} + \underbrace{\frac{1}{2}\|\boldsymbol{a} * \boldsymbol{x} - \boldsymbol{y}\|_F^2}_{\text{data fidelity}}$$

FIND **ONE OF THE MINIMIZERS** $(\widehat{\boldsymbol{a}}, \widehat{\boldsymbol{x}})$ SOLVES SASD

Algorithm: Bilinear Lasso

NATURAL, EFFECTIVE ALGORITHM—**BILINEAR LASSO**

$$\min_{a \in \mathbb{S}^{3p-1},\, x \in \mathbb{R}^n} \underbrace{\lambda \|x\|_1}_{\text{sparsity surrogate}} + \underbrace{\frac{1}{2} \|a * x - y\|_F^2}_{\text{data fidelity}}$$

FIND **ONE OF THE MINIMIZERS** $(\widehat{a}, \widehat{x})$ SOLVES SASD

Caveats:

1. Fix scale $\Longrightarrow$ optimize a over sphere where $\|a\|_2 = 1$
2. Accept shifts $\Longrightarrow$ optimize a at higher dimension space $\mathbb{R}^{3p}$ †

† This space contains all shifts: $\{s_{-p}[a_0], \ldots, s_p[a_0]\}$

Analysis of Algorithm: Approximate Bilinear Lasso

APPROXIMATION...

$$\min_{\boldsymbol{a}\in\mathbb{S}^{3p-1},\, \boldsymbol{x}\in\mathbb{R}^n} \lambda\rho(\boldsymbol{x}) + \frac{1}{2}\left\|\boldsymbol{a} * \boldsymbol{x} - \boldsymbol{y}\right\|_F^2$$

Analysis of Algorithm: Approximate Bilinear Lasso

APPROXIMATION...

$$\min_{\boldsymbol{a}\in\mathbb{S}^{3p-1},\, \boldsymbol{x}\in\mathbb{R}^n} \lambda\rho(\boldsymbol{x}) + \frac{1}{2}\left\|\boldsymbol{a} * \boldsymbol{x} - \boldsymbol{y}\right\|_F^2$$

$$= \min_{\boldsymbol{a}\in\mathbb{S}^{3p-1}} \left(\min_{\boldsymbol{x}\in\mathbb{R}^n} \lambda\rho(\boldsymbol{x}) + \frac{1}{2}\left\|\boldsymbol{a} * \boldsymbol{x} - \boldsymbol{y}\right\|_F^2\right)$$

Analysis of Algorithm: Approximate Bilinear Lasso

Approximation...

$$\min_{\boldsymbol{a}\in\mathbb{S}^{3p-1},\, \boldsymbol{x}\in\mathbb{R}^n} \lambda\rho(\boldsymbol{x}) + \frac{1}{2}\left\|\boldsymbol{a} * \boldsymbol{x} - \boldsymbol{y}\right\|_F^2$$

$$= \min_{\boldsymbol{a}\in\mathbb{S}^{3p-1}} \left(\min_{\boldsymbol{x}\in\mathbb{R}^n} \lambda\rho(\boldsymbol{x}) + \frac{1}{2}\left\|\boldsymbol{a} * \boldsymbol{x} - \boldsymbol{y}\right\|_F^2 \right)$$

$$= \min_{\boldsymbol{a}\in\mathbb{S}^{3p-1}} \left(\min_{\boldsymbol{x}\in\mathbb{R}^n} \lambda\rho(\boldsymbol{x}) + \frac{1}{2}\left\|\boldsymbol{a} * \boldsymbol{x}\right\|_F^2 + \frac{1}{2}\left\|\boldsymbol{y}\right\|_F^2 - \langle \boldsymbol{a} * \boldsymbol{x}, \boldsymbol{y} \rangle \right)$$

$$\approx \min_{\boldsymbol{a}\in\mathbb{S}^{3p-1}} \left(\min_{\boldsymbol{x}\in\mathbb{R}^n} \lambda\rho(\boldsymbol{x}) + \underbrace{\frac{1}{2}\left\|\boldsymbol{a}\right\|_F^2 + \frac{1}{2}\left\|\boldsymbol{x}\right\|_F^2}_{\text{Accurate if } \boldsymbol{a}\approx\boldsymbol{\delta} \text{ or } \boldsymbol{x} \text{ highly sparse}} + \frac{1}{2}\left\|\boldsymbol{y}\right\|_F^2 - \langle \boldsymbol{a} * \boldsymbol{x}, \boldsymbol{y} \rangle \right)$$

Analysis of Algorithm: Approximate Bilinear Lasso

Approximation...

$$\min_{\boldsymbol{a}\in\mathbb{S}^{3p-1},\,\boldsymbol{x}\in\mathbb{R}^n} \lambda\rho(\boldsymbol{x}) + \frac{1}{2}\|\boldsymbol{a}*\boldsymbol{x}-\boldsymbol{y}\|_F^2$$

$$= \min_{\boldsymbol{a}\in\mathbb{S}^{3p-1}} \left(\min_{\boldsymbol{x}\in\mathbb{R}^n} \lambda\rho(\boldsymbol{x}) + \frac{1}{2}\|\boldsymbol{a}*\boldsymbol{x}-\boldsymbol{y}\|_F^2 \right)$$

$$= \min_{\boldsymbol{a}\in\mathbb{S}^{3p-1}} \left(\min_{\boldsymbol{x}\in\mathbb{R}^n} \lambda\rho(\boldsymbol{x}) + \frac{1}{2}\|\boldsymbol{a}*\boldsymbol{x}\|_F^2 + \frac{1}{2}\|\boldsymbol{y}\|_F^2 - \langle \boldsymbol{a}*\boldsymbol{x}, \boldsymbol{y}\rangle \right)$$

$$\approx \min_{\boldsymbol{a}\in\mathbb{S}^{3p-1}} \left(\min_{\boldsymbol{x}\in\mathbb{R}^n} \lambda\rho(\boldsymbol{x}) + \underbrace{\frac{1}{2}\|\boldsymbol{a}\|_F^2 + \frac{1}{2}\|\boldsymbol{x}\|_F^2}_{\text{Accurate if } \boldsymbol{a}\approx\boldsymbol{\delta} \text{ or } \boldsymbol{x} \text{ highly sparse}} + \frac{1}{2}\|\boldsymbol{y}\|_F^2 - \langle \boldsymbol{a}*\boldsymbol{x}, \boldsymbol{y}\rangle \right)$$

$$= \min_{\boldsymbol{a}\in\mathbb{S}^{3p-1}} \underbrace{\left(\min_{\boldsymbol{x}\in\mathbb{R}^n} \lambda\rho(\boldsymbol{x}) - \langle \boldsymbol{a}*\boldsymbol{x}, \boldsymbol{y}\rangle + \frac{1}{2}\|\boldsymbol{x}\|_F^2 \right)}_{\varphi_{ABL}(\boldsymbol{a})} + \text{contant}$$

φ_{ABL} : Approximate Bilinear Lasso objective

ρ : Smooth sparsity surrogate

Analysis of Algorithm: Approximate Bilinear Lasso

THEORY: STUDY APPROXIMATE BILINEAR LASSO

$$\min_{\boldsymbol{a} \in \mathbb{S}^{3p-1}} \left(\min_{\boldsymbol{x} \in \mathbb{R}^n} \lambda \boldsymbol{\rho}(\boldsymbol{x}) + \frac{1}{2} \|\boldsymbol{x}\|_2^2 + \langle \boldsymbol{a} * \boldsymbol{x}, \boldsymbol{y} \rangle \right)$$

$$=: \boxed{\min_{\boldsymbol{a}} \quad \varphi_{\text{ABL}}(\boldsymbol{a}) \quad s.t. \quad \boldsymbol{a} \in \mathbb{S}^{3p-1}}$$

Analysis of Algorithm: Approximate Bilinear Lasso

THEORY: STUDY APPROXIMATE BILINEAR LASSO

$$\min_{\boldsymbol{a}\in\mathbb{S}^{3p-1}} \left(\min_{\boldsymbol{x}\in\mathbb{R}^n} \lambda\boldsymbol{\rho}(\boldsymbol{x}) + \frac{1}{2}\|\boldsymbol{x}\|_2^2 + \langle \boldsymbol{a} * \boldsymbol{x}, \boldsymbol{y}\rangle \right)$$

$$=: \boxed{\min_{\boldsymbol{a}} \quad \varphi_{\mathrm{ABL}}(\boldsymbol{a}) \quad s.t. \quad \boldsymbol{a} \in \mathbb{S}^{3p-1}}$$

Caveats:

- Performance is worse then Bilinear Lasso.....
- $\varphi_{ABL}(\boldsymbol{a})$ is min. of convex function of $\boldsymbol{x}$ that is easier to study

Analysis of Algorithm: Approximate Bilinear Lasso

THEORY: STUDY APPROXIMATE BILINEAR LASSO

$$\min_{\boldsymbol{a}\in\mathbb{S}^{3p-1}} \left(\min_{\boldsymbol{x}\in\mathbb{R}^n} \lambda\rho(\boldsymbol{x}) + \frac{1}{2}\|\boldsymbol{x}\|_2^2 + \langle \boldsymbol{a} * \boldsymbol{x}, \boldsymbol{y}\rangle \right)$$

$$=: \boxed{\min_{\boldsymbol{a}} \quad \varphi_{\mathrm{ABL}}(\boldsymbol{a}) \quad s.t. \quad \boldsymbol{a} \in \mathbb{S}^{3p-1}}$$

Caveats:

- Performance is worse then Bilinear Lasso.....
- $\varphi_{ABL}(\boldsymbol{a})$ is min. of convex function of $\boldsymbol{x}$ that is easier to study

Toward analysis:

- Study the geometry landscape of φ_{ABL} over sphere

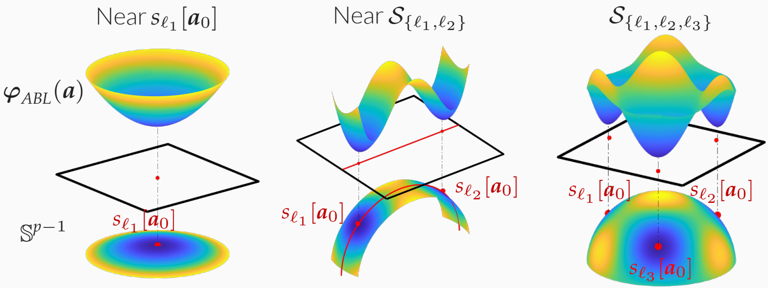
Geometry of Approximate Bilinear Lasso-1

LANDSCAPE OF φ_{ABL} NEAR SHIFTS SUBSPACE OVER SPHERE

Geometry of Approximate Bilinear Lasso-1

LANDSCAPE OF φ_{ABL} NEAR SHIFTS SUBSPACE OVER SPHERE

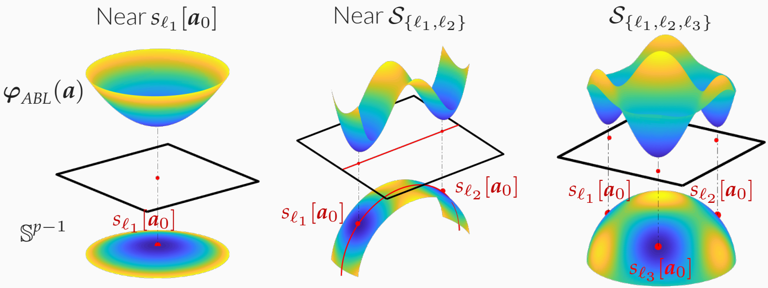
Shifts subspace: $\mathcal{S}_{\{\ell_1,\dots,\ell_\tau\}} = \mathrm{span}\left\{s_{\ell_1}[\boldsymbol{a}_0], \cdots, s_{\ell_\tau}[\boldsymbol{a}_0]\right\}$



Geometry of Approximate Bilinear Lasso-1

LANDSCAPE OF φ_{ABL} NEAR SHIFTS SUBSPACE OVER SPHERE

Shifts subspace: $\mathcal{S}_{\{\ell_1,\ldots,\ell_\tau\}} = \operatorname{span}\{s_{\ell_1}[\boldsymbol{a}_0], \cdots, s_{\ell_\tau}[\boldsymbol{a}_0]\}$



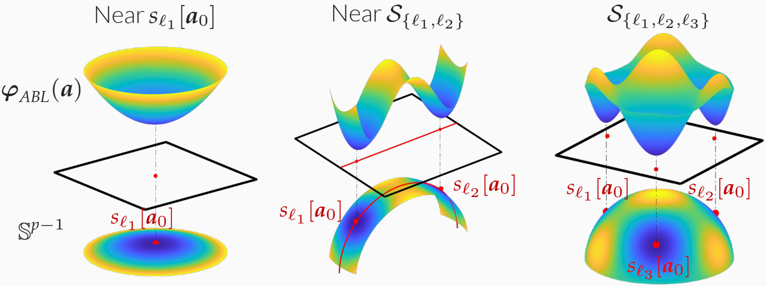
Left: $\varphi_{ABL}(\boldsymbol{a})$ near one shift over sphere

- Strongly convex
- Local minimizer is near $s_i[\boldsymbol{a}_0]$ (a good solution!)

Geometry of Approximate Bilinear Lasso-2

LANDSCAPE OF φ_{ABL} NEAR SHIFTS SUBSPACE OVER SPHERE

Shifts subspace: $\mathcal{S}_{\{\ell_1,\dots,\ell_\tau\}} = \mathrm{span}\left\{s_{\ell_1}[\boldsymbol{a}_0], \cdots, s_{\ell_\tau}[\boldsymbol{a}_0]\right\}$



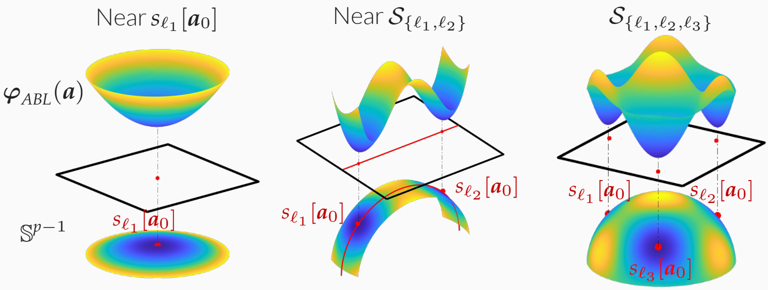
Mid: $\varphi_{ABL}(\boldsymbol{a})$ near two shifts over sphere

- Negative curvature in between shifts breaks the symmetry
- Positive curvature away from shifts subspace

Geometry of Approximate Bilinear Lasso-3

LANDSCAPE OF φ_{ABL} NEAR SHIFTS SUBSPACE OVER SPHERE

Shifts subspace: $\mathcal{S}_{\{\ell_1,\dots,\ell_\tau\}} = \operatorname{span}\{s_{\ell_1}[\boldsymbol{a}_0], \cdots, s_{\ell_\tau}[\boldsymbol{a}_0]\}$



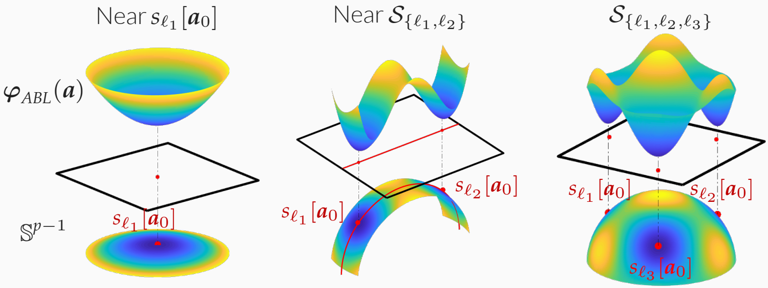
Right: $\varphi_{ABL}(\boldsymbol{a})$ over three shifts and sphere

- Convex-concave-convex geometry in higher dimension
- Every pair of shifts has similar geometry as (Mid)

Geometry of Approximate Bilinear Lasso-4

LANDSCAPE OF φ_{ABL} NEAR SHIFTS SUBSPACE OVER SPHERE

Shifts subspace: $\mathcal{S}_{\{\ell_1,\dots,\ell_\tau\}} = \operatorname{span}\{s_{\ell_1}[\boldsymbol{a}_0], \cdots, s_{\ell_\tau}[\boldsymbol{a}_0]\}$



CONCLUDE:

- LOCAL MINIMIZERS ARE NEAR SHIFTS
- NEGATIVE CURVATURE BREAKS SYMMETRY BTWN SHIFTS

Geometry of Approximate Bilinear Lasso-5

GEOMETRY OF φ_{ABL} IS IDEAL FOR OPTIMIZATION

IN **UNION OF SUBSPACES** OF **HIGH DIMENSION**

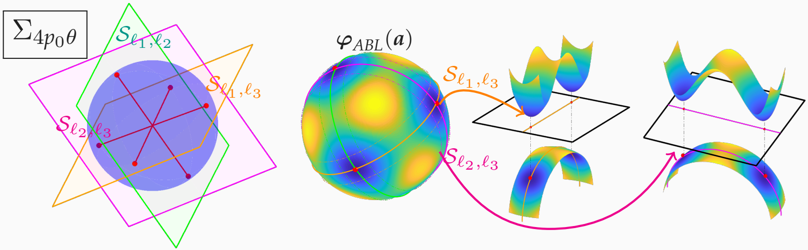
...but not global

Geometry of Approximate Bilinear Lasso-5

GEOMETRY OF φ_{ABL} IS IDEAL FOR OPTIMIZATION

IN **UNION OF SUBSPACES** OF **HIGH DIMENSION**

...but not global

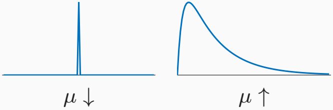


$\boxed{\Sigma_{4p_0\theta}}$: UoS spanned by $4p_0\theta$ shifts of all combination

When does φ_{ABL} has good geometry?-1

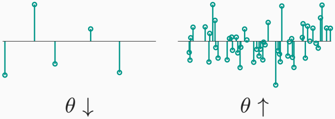
Shift–coherence μ of $\boldsymbol{a}_0$:

$$\mu \quad = \quad \max_{i \neq j} \left| \left\langle s_i[\boldsymbol{a}_0]\,,\, s_j[\boldsymbol{a}_0] \right\rangle \right|$$



Sparsity θ of $\boldsymbol{x}_0$:

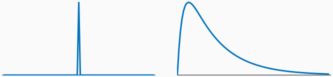
$$\boldsymbol{x}_0 \quad \sim \quad \text{Bernoulli–Gaussian}(\theta)$$



When does φ_{ABL} has good geometry?-1

SHIFT–COHERENCE μ OF $\boldsymbol{a}_0$:

$$\mu \quad = \quad \max_{i \neq j} \left| \left\langle s_i[\boldsymbol{a}_0]\,,\, s_j[\boldsymbol{a}_0] \right\rangle \right|$$



$\mu\downarrow$ $\mu\uparrow$

SPARSITY θ OF $\boldsymbol{x}_0$:

$$\boldsymbol{x}_0 \quad \sim \quad \text{Bernoulli–Gaussian}(\theta)$$

$\theta\downarrow$ $\theta\uparrow$

SASD IS **HARDER** IF...

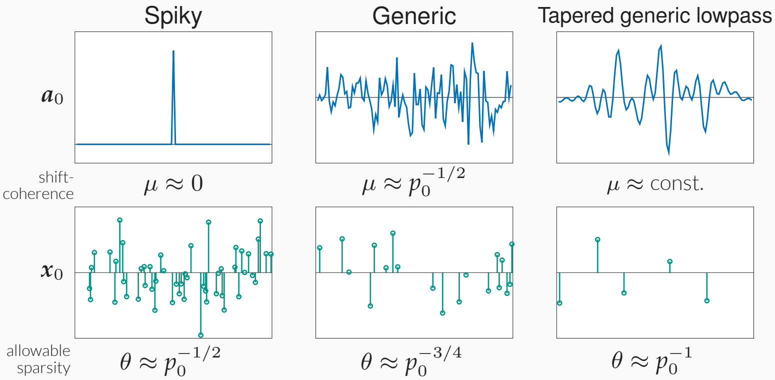
- COHERENCE $\mu\uparrow$ --- Solutions closer on sphere
- SPARSITY $\theta\uparrow$ ------- More unknowns
- $\boldsymbol{a}_0$ LENGTH $p\uparrow$ ----- More unknowns
- $\boldsymbol{y}$ LENGTH $n\downarrow$ ------- Fewer observations

When does φ_{ABL} has good geometry?-2

SPARSITY–COHERENCE TRADEOFF:

When does φ_{ABL} has good geometry?-2

SPARSITY–COHERENCE TRADEOFF:



If μ of $\boldsymbol{a}_0$ increases from $0 \nearrow 1$, than θ of $\boldsymbol{x}_0$ decreases from $\frac{1}{\sqrt{p_0}} \searrow \frac{1}{p_0}$

Algorithm---Initialization

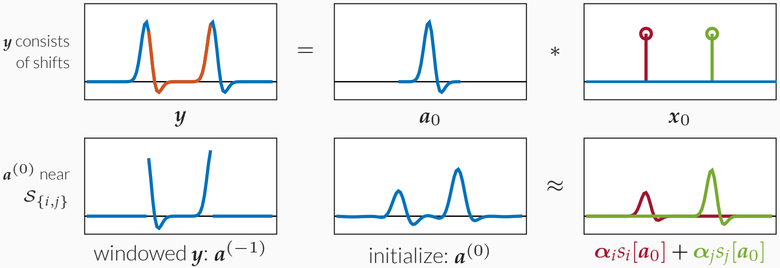
START $a^{(0)}$ NEAR SHIFTS SUBSPACE WITH **CHUNK OF SIGNAL y**

...signal $\boldsymbol{y}$ chunk is sum of few (truncated) shifts

Algorithm---Initialization

START $a^{(0)}$ NEAR SHIFTS SUBSPACE WITH CHUNK OF SIGNAL y

...signal $\boldsymbol{y}$ chunk is sum of few (truncated) shifts

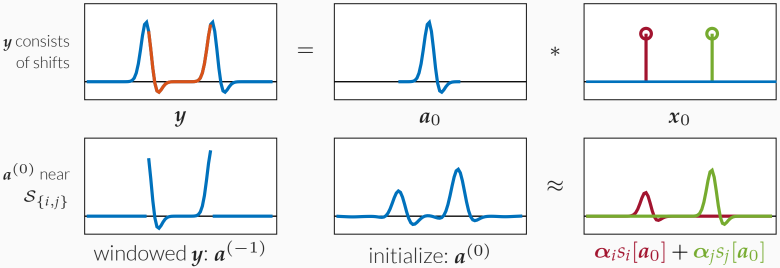


- In analysis: $\boldsymbol{a}^{(0)} = -\mathcal{P}_{\mathbb{S}^{3p-1}} \nabla \varphi_{ABL} \mathcal{P}_{\mathbb{S}^{3p-1}} \left([\mathbf{0}^p, \boldsymbol{y}_1, \cdots \boldsymbol{y}_p, \mathbf{0}^p] \right)$

Algorithm---Initialization

START $a^{(0)}$ NEAR SHIFTS SUBSPACE WITH CHUNK OF SIGNAL y

...signal $\boldsymbol{y}$ chunk is sum of few (truncated) shifts



- In analysis: $\boldsymbol{a}^{(0)} = -\mathcal{P}_{\mathbb{S}^{3p-1}} \nabla \varphi_{ABL} \mathcal{P}_{\mathbb{S}^{3p-1}} \left([\mathbf{0}^p, \boldsymbol{y}_1, \cdots \boldsymbol{y}_p, \mathbf{0}^p] \right)$
- In practice: $\boldsymbol{a}^{(0)}$ is normalized $[\mathbf{0}^p, \boldsymbol{y}_1, \cdots \boldsymbol{y}_p, \mathbf{0}^p]$

Algorithm---Retractive Minimization

SMALL STEP DESCENT METHOD STAYS NEAR SUBSPACE

...positive curvature of φ_{ABL} away from subspace

Algorithm---Retractive Minimization

SMALL STEP DESCENT METHOD STAYS NEAR SUBSPACE

...positive curvature of φ_{ABL} away from subspace

SET SPARSITY PENALTY $\lambda \lessapprox c / \sqrt{p\theta}$ WHEN $x_0 \sim c \cdot \mathcal{N}(0,1)$

...because λ acts like "soft-threshold of shifts"

Algorithm---Retractive Minimization

SMALL STEP DESCENT METHOD **STAYS NEAR** SUBSPACE

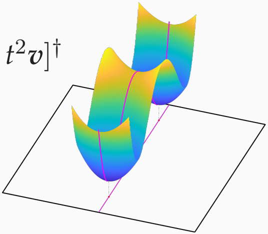
...positive curvature of φ_{ABL} away from subspace

SET **SPARSITY PENALTY** $\lambda \lessapprox c/\sqrt{p\theta}$ WHEN $\boldsymbol{x}_0 \sim c \cdot \mathcal{N}(0,1)$

...because λ acts like "soft-threshold of shifts"

In analysis: curvilinear $\boldsymbol{a}^{+} \leftarrow \mathcal{P}_{\mathbb{S}^{3p-1}}[\boldsymbol{a} - t\boldsymbol{g} - t^2\boldsymbol{v}]^{\dagger}$

In practice: alternating gradient‡



† $\mathcal{P}_{\mathbb{S}^{3p-1}}$: Riemannian retraction; $\boldsymbol{g}$: Riemannian gradient; $\boldsymbol{v}$: Riemannian curvature

‡For bilinear Lasso set $\boldsymbol{x}^{(0)}$ as minimizer given $\boldsymbol{a}^{(0)}$; small step gradients avoid saddles

Theory---Geometry & Algorithm

Thm1: Geometry of φ_{ABL} over subspaces

Given $\boldsymbol{a}_0 \in \mathbb{R}^{p_0}$, μ-shift coherent; $\boldsymbol{x}_0 \sim \mathrm{BG}(\theta)$ long and

$$\frac{1}{p_0} \lessapprox \theta \lessapprox \frac{1}{p_0\sqrt{\mu} + \sqrt{p_0}},$$

then local minima of φ_{ABL} over UoS are close to shifts.

Thm2: Provable algorithm for SaSD

A minimizing algorithm starts and stays near a subspace,

solves SaSD exactly up to a signed shift in poly time.

Analysis---Shift Space

WRITE $\boldsymbol{\alpha}$ AS COEFFICIENT OF SHIFTS SUPERPOSITION

FOR $\boldsymbol{a}$ NEAR $\mathcal{S}_{\boldsymbol{\tau}}$, $\boldsymbol{\tau} \subset \{-p, \ldots, p\}$

$$\boldsymbol{a} \;=\; \sum_{\ell \in \boldsymbol{\tau}} \boldsymbol{\alpha}_\ell s_\ell[\boldsymbol{a}_0] \;+\; \sum_{\ell \in \boldsymbol{\tau}^c} \boldsymbol{\alpha}_\ell s_\ell[\boldsymbol{a}_0] \;=\; \boldsymbol{C}_{\boldsymbol{a}_0}\boldsymbol{\alpha}^\dagger$$

Characterizes distance of $\boldsymbol{a}$ to subspace:

$$d(\boldsymbol{a}, \mathcal{S}_{\boldsymbol{\tau}}) \;=\; \inf\left\{ \|\boldsymbol{\alpha}_{\boldsymbol{\tau}^c}\|_2 : \sum_\ell \boldsymbol{\alpha}_\ell s_\ell[\boldsymbol{a}_0] = \boldsymbol{a} \right\}$$

Analysis---Shift Space

WRITE $\boldsymbol{\alpha}$ AS COEFFICIENT OF SHIFTS SUPERPOSITION FOR $\boldsymbol{a}$ NEAR $\mathcal{S}_{\boldsymbol{\tau}}$, $\boldsymbol{\tau} \subset \{-p, \dots, p\}$

$$\boldsymbol{a} = \sum_{\ell \in \boldsymbol{\tau}} \boldsymbol{\alpha}_\ell s_\ell[\boldsymbol{a}_0] + \sum_{\ell \in \boldsymbol{\tau}^c} \boldsymbol{\alpha}_\ell s_\ell[\boldsymbol{a}_0] = \boldsymbol{C}_{\boldsymbol{a}_0} \boldsymbol{\alpha}^\dagger$$

Characterizes distance of $\boldsymbol{a}$ to subspace:

$$d(\boldsymbol{a}, \mathcal{S}_{\boldsymbol{\tau}}) = \inf \left\{ \|\boldsymbol{\alpha}_{\boldsymbol{\tau}^c}\|_2 : \sum_\ell \boldsymbol{\alpha}_\ell s_\ell[\boldsymbol{a}_0] = \boldsymbol{a} \right\}$$

WRITE $\boldsymbol{\beta}$ AS COHERENCE WITH SHIFTS FOR $\boldsymbol{a}$ NEAR $\mathcal{S}_{\boldsymbol{\tau}}$

$$\boldsymbol{\beta}_\ell = \langle \boldsymbol{a}, s_\ell[\boldsymbol{a}_0] \rangle, \quad \boldsymbol{\beta} = \boldsymbol{C}^*_{\boldsymbol{a}_0} \boldsymbol{a}$$

Characterizes (geodesic) distance of $\boldsymbol{a}$ to each shifts:

$$d_s(\boldsymbol{a}, s_\ell[\boldsymbol{a}_0]) = \cos |\langle \boldsymbol{a}, s_\ell[\boldsymbol{a}_0] \rangle|$$

$^\dagger \boldsymbol{C}_{\boldsymbol{a}_0} \in \mathbb{R}^{n \times n}$ is circular convolution of zero padded $\boldsymbol{a}_0$

Analysis---Gradient & Hessian in Shift Space

SIMPLIFY OBJECTIVE (with $\rho = \ell^1$)

$$\varphi_{ABL}(\boldsymbol{a})$$

$$=_c \min_{\boldsymbol{x}} \lambda \|\boldsymbol{x}\|_1 + \tfrac{1}{2} \left\|\boldsymbol{x} - \check{\boldsymbol{y}} * \boldsymbol{a}\right\|_F^2$$

$$=_c \lambda \left\|\mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}]\right\|_1 + \tfrac{1}{2} \left\|\mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}\right\|_F^2 - \left\langle \mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}], \check{\boldsymbol{y}} * \boldsymbol{a} \right\rangle$$

$$=_c \lambda \left\|\mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}]\right\|_1 + \tfrac{1}{2} \left\|\mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}\right\|_F^2 - \left\langle \mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}], \mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}] + \lambda \boldsymbol{\sigma} \right\rangle$$

$$=_c -\tfrac{1}{2} \left\|\mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}]\right\|_F^2$$

Analysis---Gradient in Shift Space

$$\nabla \varphi_{ABL}(\boldsymbol{a}) = -\boldsymbol{\iota}^* \boldsymbol{y} * \mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}] = -\boldsymbol{\iota}^* \boldsymbol{a}_0 * \underbrace{\boldsymbol{x}_0 * \mathrm{soft}_\lambda[\check{\boldsymbol{x}}_0}_{\text{concentrate to } \chi} * \underbrace{\check{\boldsymbol{a}}_0 * \boldsymbol{a}}_{\boldsymbol{\beta}}]$$

Analysis---Gradient in Shift Space

$$\nabla \varphi_{ABL}(\boldsymbol{a}) = -\boldsymbol{\iota}^* \boldsymbol{y} * \mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}] = -\boldsymbol{\iota}^* \boldsymbol{a}_0 * \underbrace{\boldsymbol{x}_0 * \mathrm{soft}_\lambda[\check{\boldsymbol{x}}_0}_{\text{concentrate to } \chi} * \underbrace{\check{\boldsymbol{a}}_0 * \boldsymbol{a}}_{\boldsymbol{\beta}}]$$

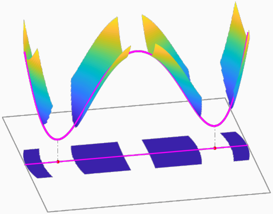
$$= -\boldsymbol{\iota}^* \boldsymbol{a}_0 * \boldsymbol{\chi}[\boldsymbol{\beta}] = -\sum_\ell \underbrace{\boldsymbol{\chi}[\boldsymbol{\beta}]_\ell}_{\approx \mathrm{soft}[\boldsymbol{\beta}]_\ell} s_\ell[\boldsymbol{a}_0]$$

Analysis---Gradient in Shift Space

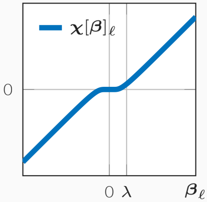
$$\nabla \varphi_{ABL}(\boldsymbol{a}) = -\boldsymbol{\iota}^* \boldsymbol{y} * \mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}] = -\boldsymbol{\iota}^* \boldsymbol{a}_0 * \underbrace{\boldsymbol{x}_0 * \mathrm{soft}_\lambda[\check{\boldsymbol{x}}_0 *}_{\text{concentrate to } \chi} \underbrace{\check{\boldsymbol{a}}_0 * \boldsymbol{a}]}_{\boldsymbol{\beta}}$$

$$= -\boldsymbol{\iota}^* \boldsymbol{a}_0 * \chi[\boldsymbol{\beta}] = -\sum_\ell \underbrace{\chi[\boldsymbol{\beta}]_\ell}_{\approx \mathrm{soft}[\boldsymbol{\beta}]_\ell} s_\ell[\boldsymbol{a}_0]$$

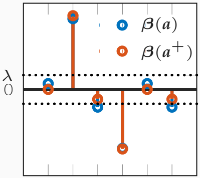
Large gradient region



gradient in shift space



gradient descent suppresses small $\boldsymbol{\beta}_i$

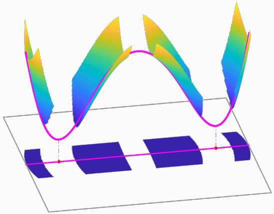


Analysis---Gradient in Shift Space

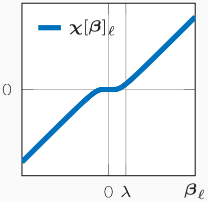
$$\nabla \varphi_{ABL}(\boldsymbol{a}) = -\boldsymbol{\iota}^* \boldsymbol{y} * \mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}] = -\boldsymbol{\iota}^* \boldsymbol{a}_0 * \underbrace{\boldsymbol{x}_0 * \mathrm{soft}_\lambda[\check{\boldsymbol{x}}_0}_{\text{concentrate to } \chi} * \underbrace{\check{\boldsymbol{a}}_0 * \boldsymbol{a}}_{\boldsymbol{\beta}}]$$

$$= -\boldsymbol{\iota}^* \boldsymbol{a}_0 * \chi[\boldsymbol{\beta}] = -\sum_\ell \underbrace{\chi[\boldsymbol{\beta}]_\ell}_{\approx \mathrm{soft}[\boldsymbol{\beta}]_\ell} s_\ell[\boldsymbol{a}_0]$$

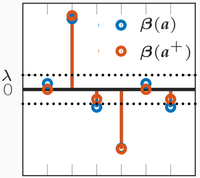
Large gradient region



gradient in shift space



gradient descent suppresses small $\boldsymbol{\beta}_i$



Riemannian gradient: $\mathcal{P}_{\boldsymbol{a}^\perp} \nabla \varphi_{ABL}(\boldsymbol{a})$:

- Gradient iterates is soft-thresholding power method on shifts
- Gradient vanishes at solution or in between shifts

Analysis---Hessian in Shift Space

$$\boldsymbol{v}^* \widetilde{\nabla}^2 \varphi(\boldsymbol{a}) \boldsymbol{v} = -\boldsymbol{v}^* \boldsymbol{a}_0 * \underbrace{\boldsymbol{x}_0 * \mathcal{P}_{\mathcal{I}}[\check{\boldsymbol{x}}_0}_{\approx c \cdot \mathbf{1}_{\{|\cdot|>\lambda\}}} * \check{\boldsymbol{a}}_0 * \boldsymbol{v}] \qquad (\mathcal{I} = \mathrm{supp}(\mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}]))$$

Analysis---Hessian in Shift Space

$$\boldsymbol{v}^* \widetilde{\nabla}^2 \varphi(\boldsymbol{a}) \boldsymbol{v} = -\boldsymbol{v}^* \boldsymbol{a}_0 * \underbrace{\boldsymbol{x}_0 * \mathcal{P}_{\mathcal{I}} [\check{\boldsymbol{x}}_0}_{\approx c \cdot \mathbf{1}_{\{|\cdot|>\lambda\}}} * \check{\boldsymbol{a}}_0 * \boldsymbol{v}] \qquad (\mathcal{I} = \operatorname{supp}(\operatorname{soft}_\lambda [\check{\boldsymbol{y}} * \boldsymbol{a}]))$$

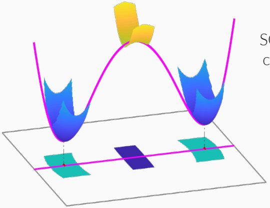
$$\approx_c -\left\langle (\check{\boldsymbol{a}}_0 * \boldsymbol{v})^{\circ 2}, \mathbf{1}_{\{|\check{\boldsymbol{a}}_0 * \boldsymbol{v}| > \lambda\}} \right\rangle = -\sum_\ell \underbrace{\boldsymbol{\beta}_\ell^2(\boldsymbol{v}) \mathbf{1}_{\{|\boldsymbol{\beta}_\ell(\boldsymbol{v})| > \lambda\}}}_{\text{logic function of } \boldsymbol{\beta}_\ell}$$

Analysis---Hessian in Shift Space

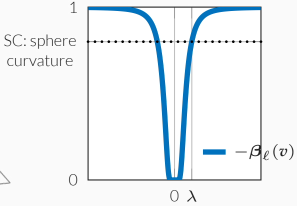
$$v^* \widetilde{\nabla}^2 \varphi(a) v = -v^* a_0 * \underbrace{x_0 * \mathcal{P}_{\mathcal{I}}[\check{x}_0}_{\approx c \cdot \mathbf{1}_{\{|\cdot|>\lambda\}}} * \check{a}_0 * v] \qquad (\mathcal{I} = \mathrm{supp}(\mathrm{soft}_\lambda[\check{y} * a]))$$

$$\approx_c -\left\langle (\check{a}_0 * v)^{\circ 2}, \mathbf{1}_{\{|\check{a}_0 * v|>\lambda\}} \right\rangle = -\sum_\ell \underbrace{\beta_\ell^2(v)\mathbf{1}_{\{|\beta_\ell(v)|>\lambda\}}}_{\text{logic function of } \beta_\ell}$$

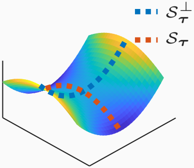
Negative curvature region
Strong convexity region



Hessian $\approx \mathbf{1}_{\{|\beta_\ell|>\lambda\}}$



positive curvature away $\mathcal{S}_\tau$
negative curvature in $\mathcal{S}_\tau$

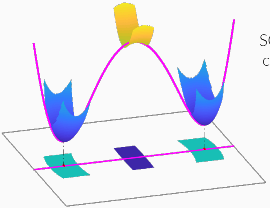


Analysis---Hessian in Shift Space

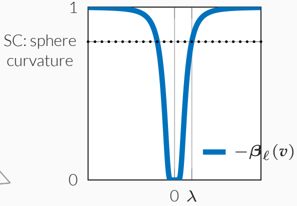
$$\boldsymbol{v}^*\widetilde{\nabla}^2\boldsymbol{\varphi}(\boldsymbol{a})\boldsymbol{v} = -\boldsymbol{v}^*\boldsymbol{a}_0 * \underbrace{\boldsymbol{x}_0 * \mathcal{P}_{\mathcal{I}}[\check{\boldsymbol{x}}_0}_{\approx c\cdot\mathbf{1}_{\{|\cdot|>\lambda\}}} * \check{\boldsymbol{a}}_0 * \boldsymbol{v}] \qquad (\mathcal{I}=\mathrm{supp}(\mathrm{soft}_\lambda[\check{\boldsymbol{y}} * \boldsymbol{a}]))$$

$$\approx_c -\left\langle (\check{\boldsymbol{a}}_0 * \boldsymbol{v})^{\circ 2}, \mathbf{1}_{\{|\check{\boldsymbol{a}}_0 * \boldsymbol{v}|>\lambda\}} \right\rangle = -\sum_\ell \underbrace{\boldsymbol{\beta}_\ell^2(\boldsymbol{v})\mathbf{1}_{\{|\boldsymbol{\beta}_\ell(\boldsymbol{v})|>\lambda\}}}_{\text{logic function of } \boldsymbol{\beta}_\ell}$$

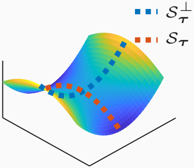
Negative curvature region
Strong convexity region



Hessian $\approx \mathbf{1}_{\{|\boldsymbol{\beta}_\ell|>\lambda\}}$



positive curvature away $\mathcal{S}_{\boldsymbol{\tau}}$
negative curvature in $\mathcal{S}_{\boldsymbol{\tau}}$

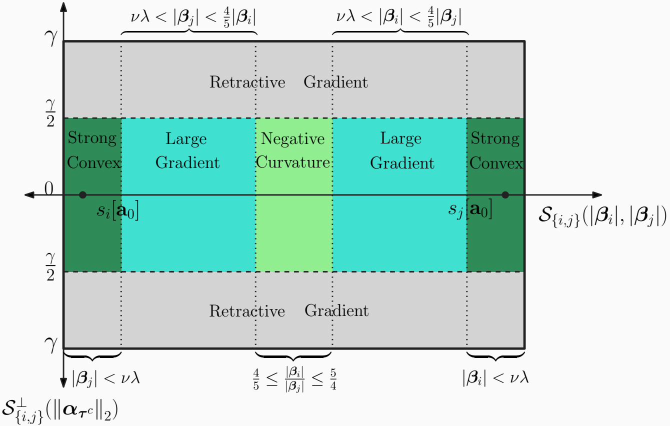


Riemannian Hessian: $\mathcal{P}_{\boldsymbol{a}^\perp}\Big(\underbrace{\widetilde{\nabla}^2\boldsymbol{\varphi}(\boldsymbol{a})}_{\boldsymbol{\varphi}\text{ curv. neg.}} + \underbrace{\langle -\nabla\boldsymbol{\varphi}(\boldsymbol{a}), \boldsymbol{a}\rangle}_{\text{sphere curv. pos.}} \Big)\mathcal{P}_{\boldsymbol{a}^\perp}$:

- $|\boldsymbol{\beta}_\ell| \uparrow$: Direction within subspace has negative curvature
- $|\boldsymbol{\beta}_\ell| \downarrow$: Direction away subspace has positive curvature

Analysis---Geometry Overview

FOUR SUBREGIONS:



$\boldsymbol{\alpha}_{\boldsymbol{\tau}^c}$: distance to subspace

$\boldsymbol{\beta}_i, \boldsymbol{\beta}_j$: distance to the shifts

Related Algorithmic Theory to SaSD-1

Works directly relevant to SaSD

Related Algorithmic Theory to SaSD-1

Works directly relevant to SaSD

[Zhang, Kuo, Wright '18]: SaSD via dictionary learning, ℓ^4 over sphere

- Better sparsty ($\boldsymbol{a}_0$ Gaussian, $\theta \le p^{-2/3}$, ours $\theta \le p^{-3/4}$)
- Only recover "truncated shifts", has addition condition requirements

[Zhang, Lau, Kuo, Wright '17]: SaSD with φ_{ABL}, highly sparse case

- Study only the dilute limit ($n \to \infty$) and highly sparse ($\theta \le 1/p$) case

Related Algorithmic Theory to SaSD-1

Works directly relevant to SaSD

[Zhang, Kuo, Wright '18]: SaSD via dictionary learning, ℓ^4 over sphere

- Better sparsty ($\boldsymbol{a}_0$ Gaussian, $\theta \leq p^{-2/3}$, ours $\theta \leq p^{-3/4}$)
- Only recover "truncated shifts", has addition condition requirements

[Zhang, Lau, Kuo, Wright '17]: SaSD with φ_{ABL}, highly sparse case

- Study only the dilute limit ($n \to \infty$) and highly sparse ($\theta \leq 1/p$) case

[Choudhary, Mitra '15] SaSD is unidentifiable

- If $\boldsymbol{x}_0$ has special support pattern, SaSD is unsolvable

Related Algorithmic Theory to SaSD-2

WORKS SOMEWHAT RELEVANT TO SASD

Related Algorithmic Theory to SaSD-2

WORKS SOMEWHAT RELEVANT TO SASD

[Ahmed, Recht, Romberg, '14] $\boldsymbol{a}_0, \boldsymbol{x}_0$ random subspace, SDP

[Chi '16] $\boldsymbol{a}_0$ random subspace, $\boldsymbol{x}_0$ sparse, atomic norm SDP

[Lee, Li, Junge, Bresler '16] random basis of sparsity, alt. min.

[Li, Ling, Strohmer, Wei, '16] random subspaces, nonconvex opt.

[Kech, Krahmer '17] random basis/subspace, optimal injectivity

...

- Random basis has no shift-symmetry, solvable with convex method
- Can be applied in communication, not SaSD cases

Related Algorithmic Theory to SaSD-2

WORKS SOMEWHAT RELEVANT TO SASD

[Ahmed, Recht, Romberg, '14] a_0, x_0 random subspace, SDP
[Chi '16] a_0 random subspace, x_0 sparse, atomic norm SDP
[Lee, Li, Junge, Bresler '16] random basis of sparsity, alt. min.
[Li, Ling, Strohmer, Wei, '16] random subspaces, nonconvex opt.
[Kech, Krahmer '17] random basis/subspace, optimal injectivity
...

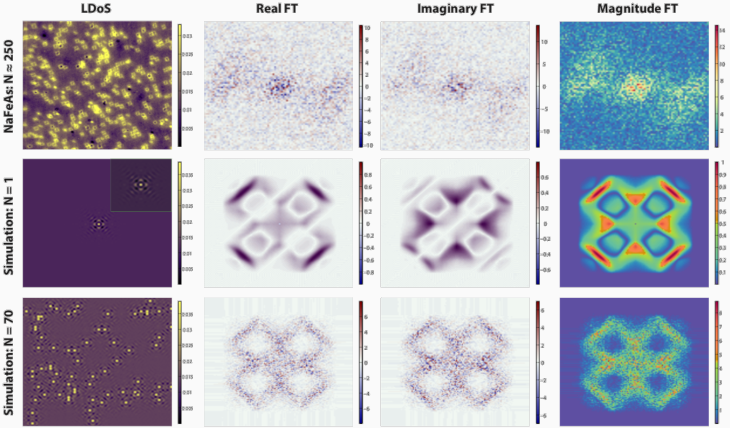
- Random basis has no shift-symmetry, solvable with convex method
- Can be applied in communication, not SaSD cases

[Wang, Chi '16] Multi-instance BD, dictionary learning
[Li, Bresler '18] Multi-instance BD, global geometry

- Multiple $y_1, \ldots y_m$, can be reduced from SaSD, not vise versa.
- Has good global geometry, more like dictionary learning

Performance of Bilinear Lasso-1

FOURIER TRANSFORM METHOD IN STM DATA



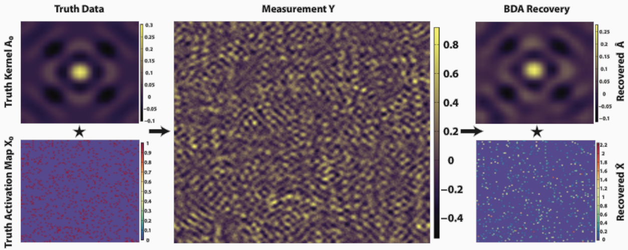
$$\hat{f}(\boldsymbol{\omega}) = \sum_{i=1}^{L} \exp\{-j(\omega_1 x_i + \omega_2 y_i)\} \times \hat{a}(\boldsymbol{\omega}).$$

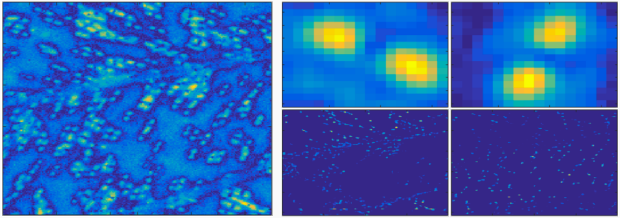
Frequency-variant "phase noise"

Defect signature (Fourier)

Performance of Bilinear Lasso-2

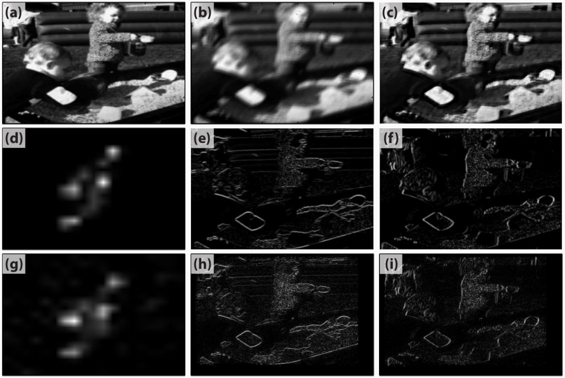
Recovery with Bilinear Lasso





Performance of Bilinear Lasso-3

Image Deblurring—recover sharp image



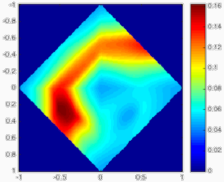
- $\boldsymbol{a}_0$ is blur kernel (d); $\boldsymbol{x}_0$ is sparse gradient (e,f)
- (a,d-f): original image, kernel, gradient x/y
- (c,g-i): recovered image, kernel, gradient x/y

Performance of Bilinear Lasso-4

COMMON METHOD IN DEBLURRING OPTIMIZE ON SIMPLEX

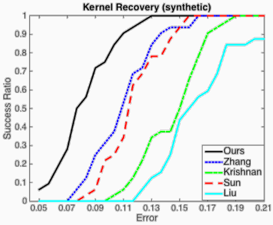
$$\min_{\boldsymbol{a},\boldsymbol{x}} \lambda \|\boldsymbol{x}\|_1 + \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{a} * \boldsymbol{x}\|_F^2, \quad s.t. \quad \|\boldsymbol{a}\|_1 = 1, \, \boldsymbol{a} \geq 0$$

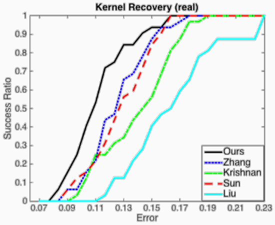
- It is a reasonable physically
- But has bad local minimizers at $\boldsymbol{a} = \boldsymbol{\delta}$
- Optimize over sphere has good geometry



Performance of Bilinear Lasso-5

COMPARISON WITH SOME OTHER METHODS





- Achieve relative good performance via simple method

Wrapping Up

Main theoretical results: **geometry of objective landscape**, and a **provable algorithm** for SaSD.

Optimizing φ_{ABL} is <u>not</u> recommended in practice.

Algorithmic ideas (sphere, initialization, etc.) are **useful for practical method** such as <u>bilinear Lasso</u>.

THANK YOU!

...AND COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK