Mathematics of Data: From Theory to Computation

Prof. Volkan Cevher
volkan.cevher@epfl.ch

Lecture 10: Source separation by convex optimization

Laboratory for Information and Inference Systems (LIONS)
École Polytechnique Fédérale de Lausanne (EPFL)

EE-556 (Fall 2015)

License Information for Mathematics of Data Slides

Outline

- Today
  1. Sourse separation
  2. Convex geometry of linear inverse problems
- Next week
  1. Primal-Dual methods

Recommended reading

- D. Amelunxen *et al.*, "Living on the edge: Phase transitions in convex programs with random data," 2014, arXiv:1303.6672v2 [cs.IT].
- M.B. McCoy *et al.*, "Convexity in source separation," *IEEE Sig. Process. Mag.*, vol. 31, pp. 87–95, 2014.
- V. Chandrasekaran *et al.*, "The convex geometry of linear inverse problems," *Found. Comput. Math.*, vol. 12, pp. 805–849, 2012.

Motivation

Motivation

This lecture illustrates how compressive sensing generalizes as a *source separation problem* in a unified framework.

It turns out that the formulations of convex estimators for both linear inverse problems and source separation problems, in general, require minimizing *nonsmooth* convex functions.

We introduce *constrained* optimization formulations as an alternative to regularization, and provide the corresponding statistics guarantees.

Source separation

Problem (Source separation)

Let $\mathbf{x}^\natural, \mathbf{y}^\natural \in \mathbb{R}^p$ *be two unknown vectors. How do we estimate* $\mathbf{x}^\natural$ *and* $\mathbf{y}^\natural$ *given* $\mathbf{z} := \mathbf{x}^\natural + \mathbf{y}^\natural$*?*

Source separation

Problem (Source separation)

Let $\mathbf{x}^\natural, \mathbf{y}^\natural \in \mathbb{R}^p$ *be two unknown vectors. How do we estimate* $\mathbf{x}^\natural$ *and* $\mathbf{y}^\natural$ *given* $\mathbf{z} := \mathbf{x}^\natural + \mathbf{y}^\natural$*?*

Observation

Source separation is impossible if we do not have any *additional information* about $\mathbf{x}^\natural$ and $\mathbf{y}^\natural$.

Example

Obviously, without any additional information, the equation $\mathbf{z} = \mathbf{x}^\natural + \mathbf{y}^\natural$ has infinitely many solutions for $(\mathbf{x}^\natural, \mathbf{y}^\natural)$.

Two important insights from nearly trivial examples

Insight # 1: To have a well-posed source separation problem, some information on the *signal structures* is needed. Here, simple representations (introduced in Lecture 7) turn out to be key.

Example

Let $\mathbf{z} = (2, 1)^T := \mathbf{x}^\natural + \mathbf{y}^\natural$. Without additional information it is impossible to perfectly recover $\mathbf{x}^\natural$ and $\mathbf{y}^\natural$.

However, suppose now we know $\mathbf{x}^\natural = (x^\natural, 0)^T$ and $\mathbf{y}^\natural = (0, y^\natural)^T$, then we can perfectly recover $\mathbf{x}^\natural = (2, 0)^T$ and $\mathbf{y}^\natural = (0, 1)^T$.

Insight # 2: The signal structures must be *incoherent* in some sense. That is, the superposed signals should not look alike so that we can separate them.

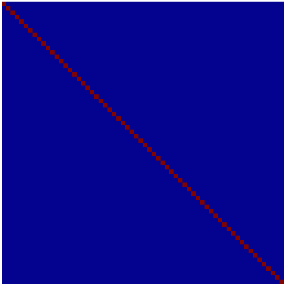
Example

Suppose now that we know $\mathbf{x}^\natural = (2, x^\natural)^T$ and $\mathbf{y}^\natural = (0, y^\natural)^T$, then it is still impossible to perfectly recover $\mathbf{x}^\natural$ and $\mathbf{y}^\natural$.

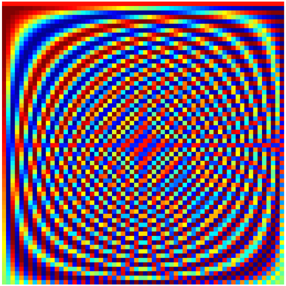
A classical well-posed source separation problem

Problem (Spikes and sines)

Let $\mathbf{x}^\natural, \mathbf{y}^\natural \in \mathbb{R}^p$ *be sparse, and let* $\mathbf{D}$ *denote the discrete cosine transform (DCT) matrix. How do we estimate* $\mathbf{x}^\natural$ *and* $\mathbf{y}^\natural$ *given* $\mathbf{z} := \mathbf{x}^\natural + \mathbf{D}\mathbf{y}^\natural$*?*



spikes



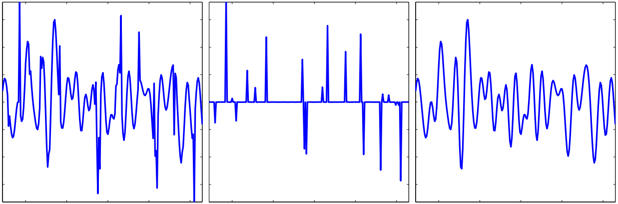
sines

A classical well-posed source separation problem

Problem (Spikes and sines)

Let $\mathbf{x}^\natural, \mathbf{y}^\natural \in \mathbb{R}^p$ be sparse, and let $\mathbf{D}$ denote the discrete cosine transform (DCT) matrix. How do we estimate $\mathbf{x}^\natural$ and $\mathbf{y}^\natural$ given $\mathbf{z} := \mathbf{x}^\natural + \mathbf{D}\mathbf{y}^\natural$?

$$\mathbf{z} \quad = \quad \mathbf{x}^\natural \quad + \quad \mathbf{D}\mathbf{y}^\natural$$



A classical well-posed source separation problem

Problem (Spikes and sines)

Let $\mathbf{x}^\natural, \mathbf{y}^\natural \in \mathbb{R}^p$ *be sparse, and let* $\mathbf{D}$ *denote the discrete cosine transform (DCT) matrix. How do we estimate* $\mathbf{x}^\natural$ *and* $\mathbf{y}^\natural$ *given* $\mathbf{z} := \mathbf{x}^\natural + \mathbf{D}\mathbf{y}^\natural$*?*

$\mathbf{z}$

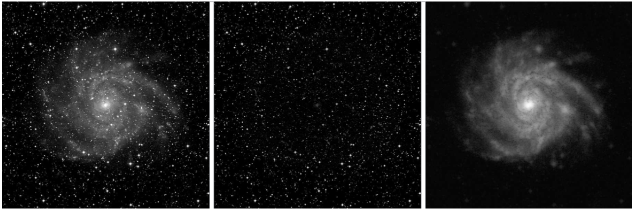
A classical well-posed source separation problem

Problem (Spikes and sines)

Let $\mathbf{x}^\natural, \mathbf{y}^\natural \in \mathbb{R}^p$ be sparse, and let $\mathbf{D}$ denote the discrete cosine transform (DCT) matrix. How do we estimate $\mathbf{x}^\natural$ and $\mathbf{y}^\natural$ given $\mathbf{z} := \mathbf{x}^\natural + \mathbf{D}\mathbf{y}^\natural$?

Observation: $\mathbf{x}^\natural$ and $\mathbf{y}^\natural$ are $\underbrace{\text{sparse}}_{\textit{signal structure}}$ $\underbrace{\text{in different bases}}_{\textit{incoherence}}$.

$$\mathbf{z} \quad = \quad \mathbf{x}^\natural \quad + \quad \mathbf{D}\mathbf{y}^\natural$$



Other applications of the source separation problem

Problem (Robust principal component analysis (PCA) [6])

Let $\mathbf{X}^\natural \in \mathbb{R}^{p\times p}$ *be sparse and* $\mathbf{Y}^\natural \in \mathbb{R}^{p\times p}$ *be low-rank. How do we estimate* $\mathbf{X}^\natural$ *and* $\mathbf{Y}^\natural$ *given* $\mathbf{Z} := \mathbf{X}^\natural + \mathbf{Y}^\natural$*?*

Applications: Background separation in videos taken with a stationary camera.

Figure: (**Left**) Original snapshot. **Center** "Low rank" background. **Right** "Sparse" foreground.

Other applications of the source separation problem

Problem (Robust principal component analysis (PCA) [6])

Let $\mathbf{X}^\natural \in \mathbb{R}^{p\times p}$ *be sparse and* $\mathbf{Y}^\natural \in \mathbb{R}^{p\times p}$ *be low-rank. How do we estimate* $\mathbf{X}^\natural$ *and* $\mathbf{Y}^\natural$ *given* $\mathbf{Z} := \mathbf{X}^\natural + \mathbf{Y}^\natural$*?*

Applications: Face illumination removal [2]: the set of all images of a convex Lambertian scene under changing illumination is close to a 9-dimensional subspace.

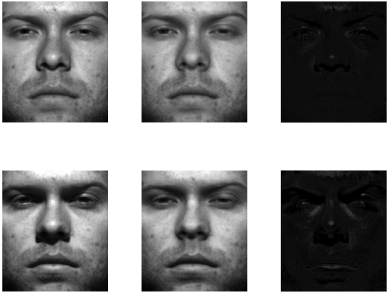


Figure: (**Left**) Faces with varying illumination. **Center** “Low rank” part. **Right** “Sparse” part.

There are many other applications

Problem (Signal denoising [16])

Let $\mathbf{x}^\natural \in \mathbb{R}^p$ and let $\mathbf{w}^\natural \in \mathbb{R}^p$ denote some unknown noise. How do we estimate $\mathbf{x}^\natural$ (and thus also $\mathbf{w}^\natural$) given $\mathbf{b} = \mathbf{x}^\natural + \mathbf{w}^\natural$?

Applications: Wireless communications with narrowband interferences, signal processing with impulse noises, etc.

Problem (Morphological component analysis [7])

Let $\mathbf{x}^\natural, \mathbf{y}^\natural \in \mathbb{R}^p$ be sparse, and $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{n\times p}$. How do we estimate $\mathbf{x}^\natural$ and $\mathbf{y}^\natural$ given $\mathbf{z} := \mathbf{U}\mathbf{x}^\natural + \mathbf{V}\mathbf{y}^\natural$?

Applications: Spikes and Sines, texture separation, image inpainting, etc.

Problem (Covariance denoising [15])

Consider the standard linear array model, where we have narrowband signals $\mathbf{s}(t) \in \mathbb{R}^r$ impinging on an array of $p \gg r$ sensors at bearings $\boldsymbol{\theta} \in \mathbb{R}^r$. The array observations $\mathbf{b}(t) \in \mathbb{R}^p$ can be written as a linear superposition of the source signals and noise $\mathbf{w} \in \mathbb{R}^p$ via a linear manifold matrix $\mathbf{A}(\boldsymbol{\theta})$: $\mathbf{b}(t) = \mathbf{A}(\theta)\mathbf{s}(t) + \mathbf{w}(t)$.

If we assume that the noise is white Gaussian with unknown variance σ^2, then the covariance of the observations $\mathbf{Z} = \mathbb{E}[\mathbf{b}\mathbf{b}^T]$ have a low-rank and diagonal decomposition: $\mathbf{Z} = \mathbf{X}^\natural + \mathbf{Y}^\natural$, where $\mathbf{X}^\natural = \mathbf{A}(\theta)^T\boldsymbol{\Sigma}_s\mathbf{A}(\theta)$ and $\mathbf{Y}^\natural = \sigma^2\mathbb{I}$, and $\boldsymbol{\Sigma}_s \in \mathbb{R}^{r\times r}$ is the source covariance. How do we estimate $\mathbf{X}^\natural$ and $\mathbf{Y}^\natural$ given $\mathbf{Z}$?

Applications: Direction-of-arrival estimation, radar, mixture of factor analyzers, etc.

Computational issue

Consider the general estimator of $(\mathbf{x}^\natural, \mathbf{y}^\natural)$ given $\mathbf{z} := \mathbf{U}\mathbf{x}^\natural + \mathbf{V}\mathbf{y}^\natural$ for *sparse* vectors $\mathbf{x}^\natural$ and $\mathbf{y}^\natural$ and corresponding linear transformations $\mathbf{U}$ and $\mathbf{V}$.

ℓ_0-"norm" approach

$$(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in \arg \min_{\mathbf{x},\mathbf{y}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_0 + \rho \|\mathbf{y}\|_0 : \mathbf{z} = \mathbf{U}\mathbf{x} + \mathbf{V}\mathbf{y} \right\}.$$

with some $\rho > 0$ that trades the relative sparsity of $\mathbf{x}$ and $\mathbf{y}$.

Observation: Since $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{U}\mathbf{x} + \mathbf{V}\mathbf{y}$ is a linear mapping, there exists a matrix $\mathbf{A}$ such that $\mathbf{z} = \mathbf{A}\tilde{\mathbf{x}}^\natural$, where $\tilde{\mathbf{x}}^\natural := ((\mathbf{x}^\natural)^T, (\mathbf{y}^\natural)^T)^T$. In fact $\mathbf{A} := \left[\begin{array}{cc} \mathbf{U} & \mathbf{V} \end{array} \right]$.

Tractability

Choosing $\rho = 1$, we have

$$\hat{\tilde{\mathbf{x}}} \in \arg \min_{\tilde{\mathbf{x}}\in\mathbb{R}^{2p}} \left\{ \|\tilde{\mathbf{x}}\|_0 : \mathbf{z} = \mathbf{A}\tilde{\mathbf{x}} \right\}.$$

In general, this procedure is *NP-hard*.

Source separation with the ℓ_1-norm

Recall the following definition for linear inverse problems.

Definition (Lasso)

Let $\mathbf{x}^\natural \in \mathbb{R}^p$, $\mathbf{A} \in \mathbb{R}^{n \times p}$, and $\mathbf{b} := \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$. The Lasso estimator for $\mathbf{x}^\natural$ is given by

$$\hat{\mathbf{x}}_{\text{Lasso}} \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \rho \|\mathbf{x}\|_1 \right\}.$$

for some $\rho \geq 0$.

For sparse source separation with $\mathbf{z} = \mathbf{U}\mathbf{x}^\natural + \mathbf{V}\mathbf{y}^\natural$, it is natural to consider the following *convex optimization* analogy.

ℓ_1-norm approach

$$(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in \arg\min_{\mathbf{x},\mathbf{y}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_1 + \rho \|\mathbf{y}\|_1 : \mathbf{z} = \mathbf{U}\mathbf{x} + \mathbf{V}\mathbf{y} \right\}$$

with some $\rho > 0$.

Atomic norms

Definition (Atomic sets & atoms)

An *atomic set* $\mathcal{A}$ is a set of vectors in $\mathbb{R}^p$. An *atom* is an element in an atomic set.

Definition (Gauge function)

Let $\mathcal{C}$ be a convex set in $\mathbb{R}^p$, the **gauge function** associated with $\mathcal{C}$ is given by

$$g_{\mathcal{C}}(\mathbf{x}) := \inf \{t : \mathbf{x} = t\mathbf{c} \text{ with some } \mathbf{c} \in \mathcal{C}, t > 0\}, \quad \forall \mathbf{x} \in \mathbb{R}^p.$$

Definition (Atomic norm)

Let $\mathcal{A}$ be an *atomic set* in $\mathbb{R}^p$, the **atomic norm** associated with $\mathcal{A}$ is given by

$$\|\mathbf{x}\|_{\mathcal{A}} := g_{\text{conv}(\mathcal{A})}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^p,$$

where $\text{conv}(\mathcal{A})$ denotes the *convex hull* of $\mathcal{A}$.

Source separation with the ℓ_1-norm

Definition (Lasso)

Let $\mathbf{x}^\natural \in \mathbb{R}^p$, $\mathbf{A} \in \mathbb{R}^{n\times p}$, and $\mathbf{b} := \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$. The Lasso estimator for $\mathbf{x}^\natural$ is given by

$$\hat{\mathbf{x}}_{\text{Lasso}} \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \rho\,\|\mathbf{x}\|_1 \right\}.$$

for some $\rho \geq 0$.

ℓ_1-norm approach

$$(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in \arg\min_{\mathbf{x},\mathbf{y}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_1 + \rho\,\|\mathbf{y}\|_1 : \mathbf{z} = \mathbf{U}\mathbf{x} + \mathbf{V}\mathbf{y} \right\}$$

with some $\rho > 0$.

Another way of looking at things

Define atomic sets $\mathcal{A}_{\mathbf{x}}$ using the set of columns of $\mathbf{U}$ and $\mathcal{A}_{\mathbf{y}}$ using the set of columns of $\mathbf{V}$. Let $\tilde{\mathbf{x}}^\natural = \mathbf{U}\mathbf{x}^\natural$ and $\tilde{\mathbf{y}}^\natural = \mathbf{V}\mathbf{y}^\natural$. With some $\rho > 0$, we equivalently have

$$(\hat{\tilde{\mathbf{x}}}, \hat{\tilde{\mathbf{y}}}) \in \arg\min_{\tilde{\mathbf{x}},\tilde{\mathbf{y}}\in\mathbb{R}^p} \left\{ \|\tilde{\mathbf{x}}\|_{\mathcal{A}_{\mathbf{x}}} + \rho\,\|\tilde{\mathbf{y}}\|_{\mathcal{A}_{\mathbf{y}}} : \mathbf{z} = \tilde{\mathbf{x}} + \tilde{\mathbf{y}} \right\}$$

General recipe for source separation

Problem (Source separation)

Let $\mathcal{A}_{\mathbf{x}}$ and $\mathcal{A}_{\mathbf{y}}$ be two atomic sets in $\mathbb{R}^p$, and let $\mathbf{x}^\natural \in \mathbb{R}^p$ and $\mathbf{y}^\natural \in \mathbb{R}^p$ be simple with respect to $\mathcal{A}_{\mathbf{x}}$ and $\mathcal{A}_{\mathbf{y}}$ respectively. How do we estimate $\mathbf{x}^\natural$ and $\mathbf{y}^\natural$ given $\mathbf{z} := \mathbf{x}^\natural + \mathbf{y}^\natural$?

A general recipe

$$(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in \arg\min_{\mathbf{x},\mathbf{y}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_{\mathcal{A}_{\mathbf{x}}} + \rho \|\mathbf{y}\|_{\mathcal{A}_{\mathbf{y}}} : \mathbf{z} = \mathbf{x} + \mathbf{y} \right\}$$

with some $\rho > 0$. In the sequel, we consider how to choose ρ.

Alternative formulations

Other variants are possible. For instance, consider the constrained variant

$$(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in \arg\min_{\mathbf{x},\mathbf{y}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_{\mathcal{A}_{\mathbf{x}}} : \mathbf{z} = \mathbf{x} + \mathbf{y}, \|\mathbf{y}\|_{\mathcal{A}_{\mathbf{y}}} \leq \kappa \right\}.$$

When $\kappa = \left\|\mathbf{y}^\natural\right\|_{\mathcal{A}_{\mathbf{y}}}$, the true vectors are feasible. As compared to the regularized version, the difficulty of choosing ρ shifts to the difficulty of choosing κ.

Example: Robust PCA

Problem (Robust principal component analysis (PCA) [2])

Let $\mathbf{X} \in \mathbb{R}^{p \times p}$ be sparse and $\mathbf{Y} \in \mathbb{R}^{p \times p}$ be low-rank. How do we estimate $\mathbf{X}$ and $\mathbf{Y}$ given $\mathbf{Z} := \mathbf{X} + \mathbf{Y}$?

Observation:

- $\mathbf{X}$ is *simple* with respect to the atomic set $\mathcal{A}_{\mathbf{X}} := \left\{ \mathbf{A}_{\mathbf{X}} : \|\mathrm{vec}(\mathbf{A}_{\mathbf{X}})\|_0 = 1, \|\mathbf{A}_{\mathbf{X}}\|_F = 1 \right\}$, and
- $\mathbf{Y}$ is *simple* with respect to the atomic set $\mathcal{A}_{\mathbf{Y}} := \left\{ \mathbf{A}_{\mathbf{Y}} : \mathrm{rank}(\mathbf{A}_{\mathbf{Y}}) = 1, \|\mathbf{A}_{\mathbf{Y}}\|_F = 1 \right\}$.

Atomic norm approach

$$(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) \in \arg \min_{\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{p \times p}} \left\{ \|\mathbf{X}\|_{\mathcal{A}_{\mathbf{X}}} + \rho \|\mathbf{Y}\|_{\mathcal{A}_{\mathbf{Y}}} \right\}$$

with some $\rho > 0$. Theory states that $\rho = 1/\sqrt{p}$ is nearly optimal.

Recall that $\|\mathbf{X}\|_{\mathcal{A}_{\mathbf{X}}} = \|\mathrm{vec}(\mathbf{X})\|_1$ and $\|\mathbf{Y}\|_{\mathcal{A}_{\mathbf{Y}}} = \|\mathbf{Y}\|_{S_1}$.

Basis pursuit with atomic norms

The convex optimization tools used for source separation ($\mathbf{z} = \mathbf{x} + \mathbf{y}$) and linear inverse problems ($\mathbf{b} = \mathbf{A}\mathbf{x}$) are similar. For the rest of the lecture, we will focus on the latter.

Linear model with *simple* parameter

Let $\mathcal{A}$ be an atomic set in $\mathbb{R}^p$. Let $\mathbf{x}^\natural \in \mathbb{R}^p$ be *simple* with respect to $\mathcal{A}$, and let $\mathbf{A} \in \mathbb{R}^{n \times p}$. The samples are given by $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$, where $\mathbf{w}$ denotes the unknown noise.

We consider the following *constrained* estimator.

Basis pursuit denoising with atomic norms

$$\hat{\mathbf{x}}_{\text{BPDN}} := \arg\min_{\mathbf{x} \in \mathbb{R}^p} \left\{ \|\mathbf{x}\|_{\mathcal{A}} : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \kappa \right\}$$

with some $\kappa \geq 0$.

- In general, this problem cannot be solved in polynomial time even if it is convex.
- When we can solve it, this heuristic formulation provides surprisingly good results.

Performance guarantee of basis pursuit denoising

Theorem

[5] Recall

$$\hat{\mathbf{x}}_{BPDN} := \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_{\mathcal{A}} : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \kappa \right\}$$

If $\|\mathbf{w}\|_2 := \left\|\mathbf{b} - \mathbf{A}\mathbf{x}^\natural\right\|_2 \leq \kappa$, *it is possible to have*

$$\left\|\hat{\mathbf{x}}_{BPDN} - \mathbf{x}^\natural\right\|_2 \leq \frac{2\kappa}{\sqrt{\mu}},$$

given that

$$n \geq \frac{w^2 + \frac{3}{2}}{\left(1 - \sqrt{\mu}\right)^2},$$

with some $\mu(\mathbf{A}) > 0$, *where* w *is some function of the atomic set* $\mathcal{A}$ *and* $\mathbf{x}^\natural$.

- The quantity w^2 characterizes the *degrees-of-freedom* of $\mathbf{x}^\natural$.
- The parameter $\mu(\mathbf{A})$ characterizes the *well-posedness* of the estimation problem.

We formally define w and prove the theorem in the following slides.

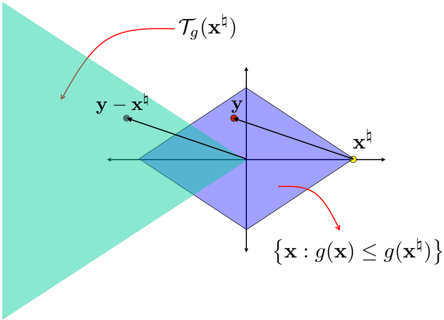
First we need the notion of *tangent cones*.

Tangent cone

Definition (Tangent cone)

Let $g : \mathbb{R}^p \to \mathbb{R} \cup \{-\infty, +\infty\}$ be a proper lower semi-continuous convex function. The tangent cone $\mathcal{T}_g(\mathbf{x})$ of the function g at a point $\mathbf{x} \in \mathbb{R}^p$ is defined as

$$\mathcal{T}_g(\mathbf{x}) := \text{cone}\left\{\mathbf{y} - \mathbf{x} : g(\mathbf{y}) \leq g(\mathbf{x}), \mathbf{y} \in \mathbb{R}^p\right\}.$$



Condition for exact recovery in the *noiseless* case

We consider estimating $\mathbf{x}^\natural \in \mathbb{R}^p$, which is simple with respect to an atomic set $\mathcal{A}$, given samples $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural$ and $\mathbf{A} \in \mathbb{R}^{n\times p}$, $n \leq p$, by

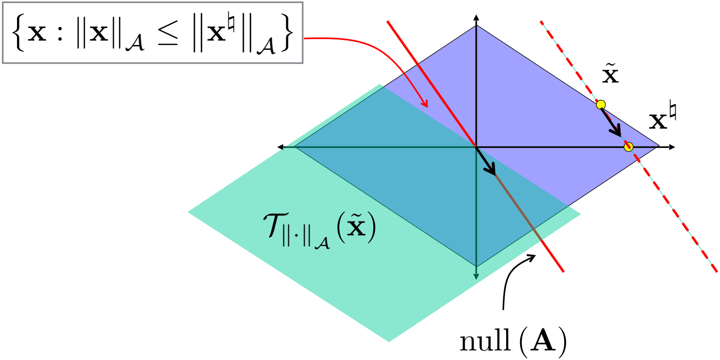
$$\hat{\mathbf{x}}_{\mathsf{BPDN}} \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_{\mathcal{A}} : \mathbf{b} = \mathbf{A}\mathbf{x} \right\}.$$

Condition for exact recovery in the *noiseless* case

Proposition

Let $g : \mathbf{x} \mapsto \|\mathbf{x}\|_{\mathcal{A}}$. *Recall* $\hat{\mathbf{x}}_{BPDN} \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{\|\mathbf{x}\|_{\mathcal{A}} : \mathbf{b} = \mathbf{A}\mathbf{x}\right\}$.

We have $\hat{\mathbf{x}}_{BPDN} = \mathbf{x}^\natural$ *if and only if* $\mathcal{T}_g\left(\mathbf{x}^\natural\right) \cap \operatorname{null}(\mathbf{A}) = \{\mathbf{0}\}$.

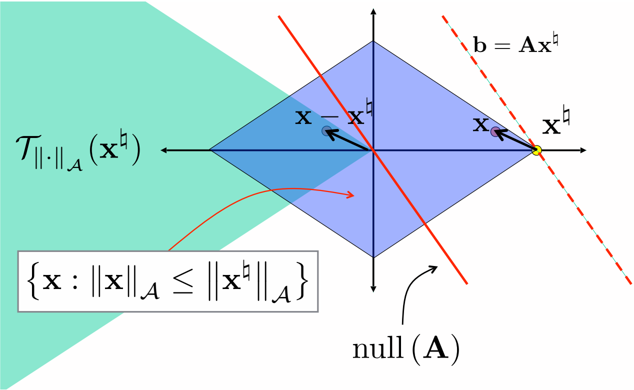


Condition for exact recovery in the *noiseless* case

Proposition

Let $g : \mathbf{x} \mapsto \|\mathbf{x}\|_{\mathcal{A}}$*. Recall* $\hat{\mathbf{x}}_{BPDN} \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{\|\mathbf{x}\|_{\mathcal{A}} : \mathbf{b} = \mathbf{A}\mathbf{x}\right\}$.

We have $\hat{\mathbf{x}}_{BPDN} = \mathbf{x}^\natural$ *if and only if* $\mathcal{T}_g\left(\mathbf{x}^\natural\right) \cap \operatorname{null}(\mathbf{A}) = \{\mathbf{0}\}$.



Condition for exact recovery in the *noisy* case

We consider estimating $\mathbf{x}^\natural \in \mathbb{R}^p$, which is simple with respect to an atomic set $\mathcal{A}$, given samples $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$ and $\mathbf{A} \in \mathbb{R}^{n \times p}$, $n \leq p$, where $\mathbf{w}$ denotes the unknown noise, by

$$\hat{\mathbf{x}}_{\mathsf{BPDN}} \in \arg\min_{\mathbf{x} \in \mathbb{R}^p} \left\{ \|\mathbf{x}\|_{\mathcal{A}} : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \kappa \right\}.$$

Condition for good recovery in the *noisy* case

Definition (Restricted strong convexity)

The restricted strong convexity condition holds if $\|\mathbf{A}\mathbf{z}\|_2^2 \geq \mu \|\mathbf{z}\|_2^2$ for all $\mathbf{z} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right)$ with some $\mu > 0$.

Proposition

Let $g : \mathbf{x} \mapsto \|\mathbf{x}\|_{\mathcal{A}}$. *Recall* $\hat{\mathbf{x}}_{BPDN} := \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{\|\mathbf{x}\|_{\mathcal{A}} : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \kappa\right\}$.

We have $\left\|\hat{\mathbf{x}}_{BPDN} - \mathbf{x}^\natural\right\|_2 \leq \frac{2\kappa}{\sqrt{\mu}}$ *if* $\|\mathbf{w}\|_2 \leq \kappa$ *and the restricted strong convexity condition holds with some* $\mu > 0$.

Condition for good recovery in the *noisy* case

Definition (Restricted strong convexity)

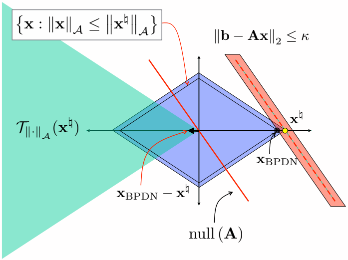
The restricted strong convexity condition holds if $\|\mathbf{A}\mathbf{z}\|_2^2 \geq \mu \|\mathbf{z}\|_2^2$ for all $\mathbf{z} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right)$ with some $\mu > 0$.

Proposition

Let $g : \mathbf{x} \mapsto \|\mathbf{x}\|_{\mathcal{A}}$. *Recall* $\hat{\mathbf{x}}_{BPDN} := \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_{\mathcal{A}} : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \kappa \right\}$.

We have $\left\|\hat{\mathbf{x}}_{BPDN} - \mathbf{x}^\natural\right\|_2 \leq \frac{2\kappa}{\sqrt{\mu}}$ *if* $\|\mathbf{w}\|_2 \leq \kappa$ *and the restricted strong convexity condition holds with some* $\mu > 0$.

Key observation: $\hat{\mathbf{x}}_{\mathrm{BPDN}} - \mathbf{x}^\natural \in \mathcal{T}_g\left(\mathbf{x}^\natural\right)$ (since $\hat{\mathbf{x}}_{\mathrm{BPDN}}$ minimizes $\|\mathbf{x}\|_{\mathcal{A}}$ subject to $\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \kappa$, and $\mathbf{x}^\natural$ satisfies this constraint by assumption)



Condition for good recovery in the *noisy* case

Definition (Restricted strong convexity)

The restricted strong convexity condition holds if $\|\mathbf{A}\mathbf{z}\|_2^2 \geq \mu \|\mathbf{z}\|_2^2$ for all $\mathbf{z} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right)$ with some $\mu > 0$.

Proposition

Let $g : \mathbf{x} \mapsto \|\mathbf{x}\|_{\mathcal{A}}$. *Recall* $\hat{\mathbf{x}}_{BPDN} := \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{\|\mathbf{x}\|_{\mathcal{A}} : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \kappa\right\}$.

We have $\left\|\hat{\mathbf{x}}_{BPDN} - \mathbf{x}^\natural\right\|_2 \leq \frac{2\kappa}{\sqrt{\mu}}$ *if* $\|\mathbf{w}\|_2 \leq \kappa$ *and the restricted strong convexity condition holds with some* $\mu > 0$.

Proof.

By definition $\hat{\mathbf{x}}_{\text{BPDN}} - \mathbf{x}^\natural \in \mathcal{T}_g\left(\mathbf{x}^\natural\right)$; thus

$$\left\|\mathbf{A}\left(\hat{\mathbf{x}}_{\text{BPDN}} - \mathbf{x}^\natural\right)\right\|_2 \geq \sqrt{\mu}\left\|\hat{\mathbf{x}}_{\text{BPDN}} - \mathbf{x}^\natural\right\|_2.$$

By the triangle inequality,

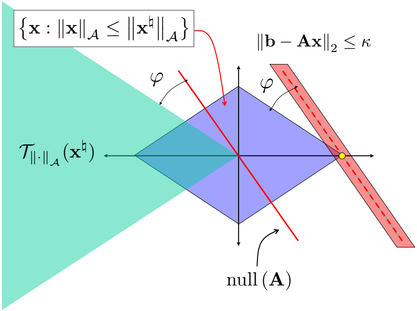
$$\left\|\mathbf{A}\left(\hat{\mathbf{x}}_{\text{BPDN}} - \mathbf{x}^\natural\right)\right\|_2 \leq \|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}_{\text{BPDN}}\|_2 + \left\|\mathbf{b} - \mathbf{A}\mathbf{x}^\natural\right\|_2 \leq 2\kappa.$$

□

Condition for good recovery in the *noisy* case

Definition (Restricted strong convexity)

The restricted strong convexity condition holds if $\|\mathbf{A}\mathbf{z}\|_2^2 \geq \mu \|\mathbf{z}\|_2^2$ for all $\mathbf{z} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right)$ with some $\mu > 0$.



- In the figure, μ is proportional to $\sin^2(\varphi)$, where the proportionality depends on the norm of the rows of $\mathbf{A}$.

Interpretation of the *restricted strong convexity* condition

Definition (Restricted strong convexity)

The restricted strong convexity condition holds if $\|\mathbf{A}\mathbf{z}\|_2^2 \geq \mu \|\mathbf{z}\|_2^2$ for all $\mathbf{z} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right)$ with some $\mu > 0$.

Proposition

The restricted strong convexity condition holds if and only if the function $f : \mathbf{h} \mapsto \frac{1}{2}\left\|\mathbf{b} - \mathbf{A}\left(\mathbf{x}^\natural + \mathbf{h}\right)\right\|_2^2$ *satisfies*

$$f(\mathbf{x}^\natural + \mathbf{h}) \geq f(\mathbf{x}^\natural) + \left\langle \nabla f(\mathbf{x}^\natural), \mathbf{h} \right\rangle + \frac{\mu}{2}\|\mathbf{h}\|_2^2, \quad \textit{for all } \mathbf{h} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right),$$

or, $f(\mathbf{h})$ *behaves as a strongly convex function for* $\mathbf{h} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right)$.

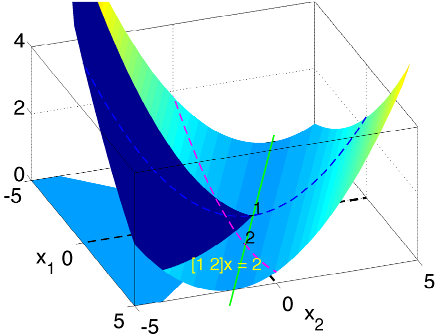
Interpretation of the *restricted strong convexity* condition

Proposition

The restricted strong convexity condition holds if and only if the function $f : \mathbf{h} \mapsto \frac{1}{2} \left\| \mathbf{b} - \mathbf{A}\left(\mathbf{x}^\natural + \mathbf{h}\right) \right\|_2^2$ *satisfies*

$$f(\mathbf{x}^\natural + \mathbf{h}) \geq f(\mathbf{x}^\natural) + \left\langle \nabla f(\mathbf{x}^\natural), \mathbf{h} \right\rangle + \frac{\mu}{2} \|\mathbf{h}\|_2^2, \quad \text{for all } \mathbf{h} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right),$$

or, $f(\mathbf{h})$ *behaves as a strongly convex function for* $\mathbf{h} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right)$.



Interpretation of the *restricted strong convexity* condition

Definition (Restricted strong convexity)

The restricted strong convexity condition holds if $\|\mathbf{A}\mathbf{z}\|_2^2 \geq \mu \|\mathbf{z}\|_2^2$ for all $\mathbf{z} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right)$ with some $\mu > 0$.

Proposition

The restricted strong convexity condition holds if and only if the function $f : \mathbf{h} \mapsto \frac{1}{2}\left\|\mathbf{b} - \mathbf{A}\left(\mathbf{x}^\natural + \mathbf{h}\right)\right\|_2^2$ *satisfies*

$$f(\mathbf{x}^\natural + \mathbf{h}) \geq f(\mathbf{x}^\natural) + \left\langle \nabla f(\mathbf{x}^\natural), \mathbf{h} \right\rangle + \frac{\mu}{2}\|\mathbf{h}\|_2^2, \quad \text{for all } \mathbf{h} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right),$$

or, $f(\mathbf{h})$ *behaves as a strongly convex function for* $\mathbf{h} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right)$.

Observation: Note that $\hat{\mathbf{x}}_{\text{BPDN}} = \mathbf{x}^\natural + \mathbf{h}$ with some $\mathbf{h} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right)$ by definition. Thus the restricted strong convexity condition implies that the function $\frac{1}{2}\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2$ behaves as if $\mathbf{A}$ had full column rank for all possible values of $\hat{\mathbf{x}}_{\text{BPDN}}$.

- There are some variants of this restricted strong convexity condition based on similar ideas [1, 12].

Verifying the conditions

Now we have performance guarantees for $\hat{\mathbf{x}}_{\mathrm{BPDN}}$.

Proposition (Noiseless)

Let $g : \mathbf{x} \mapsto \|\mathbf{x}\|_{\mathcal{A}}$. *We have* $\hat{\mathbf{x}}_{BPDN} = \mathbf{x}^\natural$ *if and only if* $\mathcal{T}_g\left(\mathbf{x}^\natural\right) \cap \operatorname{null}(\mathbf{A}) = \{\mathbf{0}\}$.

Proposition (Noisy)

Let $g : \mathbf{x} \mapsto \|\mathbf{x}\|_{\mathcal{A}}$. *We have* $\left\|\hat{\mathbf{x}}_{BPDN} - \mathbf{x}^\natural\right\|_2 \leq \frac{2\kappa}{\sqrt{\mu}}$ *if* $\|\mathbf{w}\|_2 \leq \kappa$ *and* $\|\mathbf{A}\mathbf{z}\|_2^2 \geq \mu \|\mathbf{z}\|_2^2$ *for all* $\mathbf{z} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right)$ *with some* $\mu > 0$.

How do we verify *these conditions*, especially when we do not know $\mathbf{x}^\natural$ and thus $\mathcal{T}_g\left(\mathbf{x}^\natural\right)$?

No good answers currently.

The probabilistic approach

Suppose now that $\mathbf{A}$ is *random*.

Show that no matter what $\mathbf{x}^\natural$ is, under *some other verifiable conditions*, we have

$$\mathcal{T}_g\left(\mathbf{x}^\natural\right) \cap \mathrm{null}\left(\mathbf{A}\right) = \{\mathbf{0}\}\,, \text{ or}$$
$$\|\mathbf{A}\mathbf{z}\|_2^2 \geq \mu \|\mathbf{z}\|_2^2\,, \quad \forall \mathbf{z} \in \mathcal{T}_g\left(\mathbf{x}^\natural\right) \text{ with some } \mu > 0,$$

with probability bounded away from 0.

A key quantity characterizing the degrees of freedom of the tangent cone is the *Gaussian width*, and the key technical tool is the *escape-through-the-mesh theorem*.

Gaussian width

Definition (Gaussian width)

The Gaussian width $w(\Omega)$ of a set $\Omega \subset \mathbb{R}^n$ is given by

$$w(\Omega) := \mathbb{E}\left[\max_{\mathbf{x}\in\Omega} \langle \mathbf{g}, \mathbf{x} \rangle\right],$$

where $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Example

Let V be a d-dimensional subspace of $\mathbb{R}^p$, and let Ω be the intersection of V and the unit ℓ_2-norm sphere. Then $w(\Omega) = \sqrt{d}$.

This supports our claim that $[w(\Omega)]^2$ characterizes the degree of freedom of a set.

Proposition

1. *The Gaussian width is invariant under translation and unitary transforms (rotations).*
2. *Let $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \mathbb{R}^n$. Then $w(\mathcal{C}_1) \leq w(\mathcal{C}_2)$.*

Examples

Let Ω always denote the intersection of $\mathcal{T}_{\|\cdot\|_{\mathcal{A}}}\left(\mathbf{x}^\natural\right)$ and the unit ℓ_2-norm sphere.

Example ([5])

1. Let $\mathcal{A} = \{\mathbf{e}_1, \ldots, \mathbf{e}_p\}$, and let $\mathbf{x}^\natural \in \mathbb{R}^p$ with at most s non-zero entries. Then $\|\cdot\|_{\mathcal{A}}$ is the ℓ_1-norm, and $w(\Omega)^2 \leq 2s\log\left(\frac{p}{s}\right) + \frac{5}{4}s$.
2. Let $\mathcal{A} = \{-1, +1\}^p$, and let $\mathbf{x}^\natural \in \mathbb{R}^p$ be a convex combination of k vectors in $\mathcal{A}$. Then $\|\cdot\|_{\mathcal{A}}$ is the ℓ_∞-norm, and $w(\Omega)^2 \leq \frac{p+k}{2}$.
3. Let $\mathcal{A} = \left\{\mathbf{X} : \mathrm{rank}\ (\mathbf{X}) = 1, \|\mathbf{X}\|_F = 1, \mathbf{X} \in \mathbb{R}^{p\times p}\right\}$, and let $\mathbf{X}^\natural \in \mathbb{R}^{p\times p}$ with rank r. Then $\|\cdot\|_{\mathcal{A}}$ is the nuclear norm, and $w(\Omega)^2 \leq 3r(2p - r)$.

Some applications follow directly.

*Escape-through-the-mesh theorem

Theorem (Escape-through-the-mesh theorem [5, 10, 14])

Let $\mathbf{A} \in \mathbb{R}^{n\times p}$ *be a matrix of i.i.d. Gaussian random variables with zero means and variances* $1/n$*. Let* Ω *be a given set on the unit* ℓ_2*-norm sphere. Then*

$$\mathbb{P}\left(\left\{\|\mathbf{Ax}\|_2 \geq \sqrt{\mu}, \forall \mathbf{x} \in \Omega\right\}\right) \geq 1 - \exp\left\{-\frac{1}{2}\left[a_n - w(\Omega) - \sqrt{n\mu}\right]^2\right\}$$

given that $a_n - w(\Omega) - \sqrt{n\mu} \geq 0$*, where* $a_n := \sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n}{2}\right)$, Γ *being the gamma function, and*

$$w(\Omega) := \mathbb{E}\left[\max_{\mathbf{x}\in\Omega} \langle \mathbf{g}, \mathbf{x}\rangle\right],$$

$\mathbf{g}$ *being a vector of i.i.d. standard Gaussian random variables.*

Observation:

- The event $\left\{\|\mathbf{Ax}\|_2^2 \geq \mu, \forall \mathbf{x} \in \Omega\right\}$ implies the event that $\mathrm{null}\,(\mathbf{A})$ does not intersect with the mesh Ω.
- One can prove that $\frac{n}{\sqrt{n+1}} \leq a_n \leq \sqrt{n}$, which implies $a_n \approx \sqrt{n}$.

Probabilistic results for the *noiseless* case

Let $\mathbf{A} \in \mathbb{R}^{n \times p}$ be a *matrix of i.i.d. Gaussian random variables* with zero means and variances $1/n$.

Let Ω be the intersection of $\mathcal{T}_{\|\cdot\|_{\mathcal{A}}}\left(\mathbf{x}^\natural\right)$ and the unit ℓ_2-norm sphere.

Theorem (Noiseless)

We have $\hat{\mathbf{x}}_{BPDN} = \mathbf{x}^\natural$ with probability at least $1 - \exp\left\{-\frac{1}{2}\left[a_n - w(\Omega)\right]^2\right\}$, provided that $n \geq w(\Omega)^2 + 1$.

Proof.

Replace Ω by the intersection of $\mathcal{T}_{\|\cdot\|_{\mathcal{A}}}\left(\mathbf{x}^\natural\right)$ and the unit ℓ_2-norm sphere in the escape-through-the-mesh theorem. Note that the escape-through-the-mesh theorem is only meaningful when $a_n \geq w(\Omega)$; this condition leads to the constraint $n \geq w(\Omega)^2 + 1$. □

Probabilistic results for the *noisy* case

Assume that $\mathbf{A} \in \mathbb{R}^{n \times p}$ be a *matrix of i.i.d. Gaussian random variables* with zero means and variances $1/n$.

Let Ω be the intersection of $\mathcal{T}_{\|\cdot\|_{\mathcal{A}}}\left(\mathbf{x}^\natural\right)$ and the unit ℓ_2-norm sphere.

Theorem (Noisy)

For any $\mu \in (0,1)$*, we have* $\left\|\hat{\mathbf{x}}_{BPDN} - \mathbf{x}^\natural\right\|_2 \leq \frac{2\delta}{\sqrt{\mu}}$ *with probability at least* $1 - \exp\left\{-\frac{1}{2}\left[a_n - w(\Omega) - \sqrt{\mu n}\right]^2\right\}$ *provided that* $\|\mathbf{w}\|_2 \leq \delta$ *and* $n \geq \frac{w(\Omega)^2 + \frac{3}{2}}{(1-\sqrt{\mu})^2}$.

Proof.

Replace Ω by the intersection of $\mathcal{T}_{\|\cdot\|_{\mathcal{A}}}\left(\mathbf{x}^\natural\right)$ and the unit ℓ_2-norm sphere in the escape-through-the-mesh theorem. Note that the escape-through-the-mesh theorem is only meaningful when $a_n \geq w(\Omega) + \sqrt{\mu n}$; this condition leads to the constraint $n \geq \frac{w(\Omega)^2 + \frac{3}{2}}{(1-\sqrt{\mu})^2}$, assuming $\mu \in (0,1)$. □

Interpretation of the results

Recall the result in the previous slide.

Theorem (Noisy)

For any $\mu \in (0,1)$, *we have* $\left\|\hat{\mathbf{x}}_{BPDN} - \mathbf{x}^\natural\right\|_2 \leq \frac{2\kappa}{\sqrt{\mu}}$ *with probability at least* $1 - \exp\left\{-\frac{1}{2}\left[a_n - w(\Omega) - \sqrt{\mu n}\right]^2\right\}$ *provided that* $\|\mathbf{w}\|_2 \leq \kappa$ *and* $n \geq \frac{w(\Omega)^2 + \frac{3}{2}}{(1-\sqrt{\mu})^2}$.

We have an equivalent formulation assuming $\kappa = \|\mathbf{w}\|_2$.

Theorem

For any $\mu \in (0,1)$, *we have*

$$\left\|\hat{\mathbf{x}}_{BPDN} - \mathbf{x}^\natural\right\|_2 \leq \frac{2\sqrt{n}}{a_n - w(\Omega) - t}\|\mathbf{w}\|_2 \leq \frac{2\sqrt{n}}{\sqrt{n} - w(\Omega) - t}\|\mathbf{w}\|_2$$

with probability at least $1 - \exp\left(-\frac{1}{2}t^2\right)$ *provided* $n \geq \frac{w(\Omega)^2 + \frac{3}{2}}{(1-\sqrt{\mu})^2}$.

Observation: The quantity $w(\Omega)^2$ characterizes the degree of freedom of $\mathbf{x}^\natural$.

Remark: We will discuss an improvement of this guarantee.

Application 1: Compressive sensing

Problem formulation [4, 9]

Let $\mathbf{x}^\natural \in \mathbb{R}^p$ with at most s non-zero entries, and let $\mathbf{A} \in \mathbb{R}^{n\times p}$. How do we estimate $\mathbf{x}^\natural$ given $\mathbf{A}$ and $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$, where $\mathbf{w}$ denotes unknown noise?

Example

Let $\mathcal{A} = \{\pm\mathbf{e}_1, \ldots, \pm\mathbf{e}_p\}$, and let $\mathbf{x}^\natural \in \mathbb{R}^p$ with at most s non-zero entries. Then $\|\cdot\|_{\mathcal{A}}$ is the ℓ_1-norm, and $w(\Omega)^2 \leq 2s\log\left(\frac{p}{s}\right) + \frac{5}{4}s$.

Choose $\mathbf{A}$ to be a matrix of i.i.d. Gaussian random variables with zero means and variances $1/n$. Then by

$$\hat{\mathbf{x}}_{\text{BPDN}} \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_1 : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \kappa \right\}$$

with $\kappa = \|\mathbf{w}\|_2$, we have

$$\left\| \hat{\mathbf{x}}_{\text{BPDN}} - \mathbf{x}^\natural \right\|_2 \lesssim \frac{2\sqrt{n}}{\sqrt{n} - \sqrt{2s\log\left(\frac{p}{s}\right) + \frac{5}{4}s}} \|\mathbf{w}\|_2 .$$

Application 2: Multi-knapsack feasibility problem

Problem formulation [11]

Let $\mathbf{x}^\natural \in \mathbb{R}^p$ which is a convex combination of k vectors in $\mathcal{A} := \{-1, +1\}^p$, and let $\mathbf{A} \in \mathbb{R}^{n\times p}$. How large should n be such that we can recover $\mathbf{x}^\natural$ given $\mathbf{A}$ and $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural$ via

$$\hat{\mathbf{x}}_{\text{BPDN}} \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_\infty : \mathbf{b} = \mathbf{A}\mathbf{x} \right\}?$$

Example

Let $\mathcal{A} = \{-1, +1\}^p$, and let $\mathbf{x}^\natural \in \mathbb{R}^p$ be a convex combination of k vectors in $\mathcal{A}$. Then $\|\cdot\|_{\mathcal{A}}$ is the ℓ_∞-norm, and $w(\Omega)^2 \leq \frac{p+k}{2}$.

Choose $\mathbf{A}$ to be a matrix of i.i.d. Gaussian random variables with zero means and variances $1/n$. Then we have

$$\mathbb{P}\left(\left\{\hat{\mathbf{x}}_{\text{BPDN}} = \mathbf{x}^\natural\right\}\right) \gtrsim 1 - \exp\left\{ -\frac{1}{2}\left[\sqrt{n} - \sqrt{\frac{p+k}{2}}\right]^2 \right\}.$$

Application 3: Matrix completion

Problem formulation [3, 8]

Let $\mathbf{X}^\natural \in \mathbb{R}^{p\times p}$ with $\mathrm{rank}(\mathbf{X}^\natural) = r$, and let $\mathbf{A}_1, \ldots, \mathbf{A}_n$ be matrices in $\mathbb{R}^{p\times p}$. How do we estimate $\mathbf{X}^\natural$ given $\mathbf{A}_1, \ldots, \mathbf{A}_n$ and $b_i = \mathrm{Tr}\left(\mathbf{A}_i \mathbf{X}^\natural\right) + w_i$, $i = 1, \ldots, n$, where $\mathbf{w} := (w_1, \ldots, w_n)^T$ denotes unknown noise?

Example

Let $\mathcal{A} = \left\{\mathbf{X} : \mathrm{rank}\ (\mathbf{X}) = 1, \|\mathbf{X}\|_F = 1, \mathbf{X} \in \mathbb{R}^{p\times p}\right\}$, and let $\mathbf{X}^\natural \in \mathbb{R}^{p\times p}$ with rank r. Then $\|\cdot\|_{\mathcal{A}}$ is the nuclear norm, and $w(\Omega)^2 \leq 3r(2p - r)$.

Choose each A_i to be a matrix of i.i.d. Gaussian random variables with zero means and variances $1/n$. Then by

$$\hat{\mathbf{X}}_{\text{BPDN}} \in \arg \min_{\mathbf{X} \in \mathbb{R}^{p\times p}} \left\{ \|\mathbf{X}\|_* : \sum_{i=1}^{n} \left(b_i - \mathrm{Tr}\left(\mathbf{A}_i \mathbf{X}\right)\right)^2 \leq \kappa^2 \right\}$$

with $\kappa = \|\mathbf{w}\|_2$, we have

$$\left\|\hat{\mathbf{X}}_{\text{BPDN}} - \mathbf{X}^\natural\right\|_2 \lesssim \frac{2\sqrt{n}}{\sqrt{n} - \sqrt{3r(2p - r)}} \|\mathbf{w}\|_2 .$$

Sharper bounds with oracle information

Suppose that we are able to set

$$\hat{\mathbf{x}}_{\text{BPDN,oracle}} \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_{\mathcal{A}} : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{w}\|_2 \right\}.$$

Theorem ([13])

With probability at least $1 - 6\exp\left(-t^2/26\right)$, *we have*

$$\left\|\hat{\mathbf{x}}_{BPDN,oracle} - \mathbf{x}^\natural\right\|_2 \leq \left[\frac{w(\Omega)+t}{a_{n-1}}\right] \left[\frac{2\sqrt{n}}{a_n - w(\Omega) - t}\right] \|\mathbf{w}\|_2$$

for any $t > 0$, *where* Ω *denotes the intersection of* $\mathcal{T}_{\|\cdot\|_{\mathcal{A}}}\left(\mathbf{x}^\natural\right)$ *and the unit* ℓ_2*-norm sphere.*

Observation: Recall that our analysis gives that with probability at least $1 - \exp\left(-t^2/2\right)$,

$$\left\|\hat{\mathbf{x}}_{\text{BPDN,oracle}} - \mathbf{x}^\natural\right\|_2 \lesssim \left[\frac{2\sqrt{n}}{a_n - w(\Omega) - t}\right] \|w\|_2.$$

An improvement by the factor $\frac{w(\Omega)+t}{a_{n-1}} \leq 1$ appears assuming access of the oracle information $\|\mathbf{w}\|_2$.

References I

[1] Peter Bickel, Ya'acov Ritov, and Alexandre B. Tsybakov.
Simultaneous analysis of Lasso and Dantzig selector.
Ann. Stat., 37(4):1705–1732, 2009.

[2] Emmanuel Candès, Xiaodong Li, Yi Ma, and John Wright.
Robust principal component analysis?
J. ACM, 58(3), may 2011.

[3] Emmanuel Candès and Benjamin Recht.
Exact matrix completion via convex optimization.
Found. Comput. Math., 9:717–772, 2009.

[4] Emmanuel J. Candès, Justin Romberg, and Terence Tao.
Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.
IEEE Trans. Inf. Theory, 52(2):489–509, February 2006.

[5] Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan S. Willsky.
The convex geometry of linear inverse problems.
Found. Comput. Math., 12:805–849, 2012.

[6] Venkat Chandrasekaran, Sujay Sanghavi, Pablo A. Parrilo, and Alan S. WIllsky.
Rank-sparsity incoherence for matrix decomposition.
SIAM J. Optim., 21(2):572–596, 2011.

References II

[7] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho.
Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA).
Appl. Comput. Harmon. Anal., 19:340–358, 2005.

[8] Steven T. Flammia, David Gross, Yi-Kai Liu, and Jens Eisert.
Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators.
New J. Phys., 14, 2012.

[9] Simon Foucart and Holger Rauhut.
A Mathematical Introduction to Compressive Sensing.
Birkhäuser, Basel, 2013.

[10] Y. Gordon.
On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$.
In Joram Lindenstrauss and Vitali D. Milman, editors, *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 1986–87*, Berlin, 1988. Springer-Verl.

[11] O. L. Mangasarian and Benjamin Recht.
Probability of unique integer solution to a system of linear equations.
Eur. J. Oper. Res., 214:27–30, 2011.

References III

[12] Sahand N. Negahban, Pradeep Ravikumar, Martin J. Wainwright, and Bin Yu.
A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers.
Stat. Sci., 27(4):538–557, 2012.

[13] Samet Oymak, Christos Thrampoulidis, and Babak Hassibi.
Simple bounds for noisy linear inverse problems with exact side information.
2013.
arXiv:1312.0641v2 [cs.IT].

[14] Mark Rudelson and Roman Vershynin.
On sparse reconstruction from Fourier and Gaussian measurements.
Commun. Pure Appl. Math., LXI:1025–1045, 2008.

[15] James Saunderson, Venkat Chandrasekaran, Pablo A Parrilo, and Alan S Willsky.
Diagonal and low-rank matrix decompositions, correlation matrices, and ellipsoid fitting.
SIAM Journal on Matrix Analysis and Applications, 33(4):1395–1416, 2012.

[16] Ghristoph Studer, Patrick Kuppinger, Graeme Pope, and Helmut Bölcskei.
Recovery of sparsely corrupted signals.
IEEE Trans. Inf. Theory, 58(5):3115–3130, May 2012.