Compressed sensing off-the-grid:
The Fisher metric, support stability and optimal sampling bounds

Clarice Poon
University of Bath

Joint work with:

Nicolas Keriven and **Gabriel Peyré**
École Normale Supérieure

February 6, 2019

Outline

Compressed sensing [Candès, Romberg & Tao '06; Donoho '06]

Task: Recover $a \in \mathbb{C}^N$ from $y = \Phi a$ where $\Phi \in \mathbb{C}^{m \times N}$ with $m \ll N$ and a is s-sparse.

Typical compressed sensing statement:

For certain random matrices $\Phi \in \mathbb{C}^{m \times N}$, with high probability, a can be uniquely recovered from $m = \mathcal{O}\left(s \log\left(N\right)\right)$ measurements by solving

$$\min_{z \in \mathbb{C}^N} \|z\|_1 \text{ subject to } \Phi z = y$$

or in the noisy case of $y = \Phi a + w$, the minimizer $\hat{a}$ of

$$\min_{z \in \mathbb{C}^N} \lambda \|z\|_1 + \frac{1}{2} \|\Phi z - y\|_2^2$$

with $\lambda \sim \delta/\sqrt{s}$ and $\|w\| \leqslant \delta$ satisfies $\|a - \hat{a}\|_1 \lesssim \sigma_s(x)_1 + \sqrt{s}\delta.$

Compressed sensing [Candès, Romberg & Tao '06; Donoho '06]

Task: Recover $a \in \mathbb{C}^N$ from $y = \Phi a$ where $\Phi \in \mathbb{C}^{m\times N}$ with $m \ll N$ and a is s-sparse.

Typical compressed sensing statement:

For certain random matrices $\Phi \in \mathbb{C}^{m\times N}$, with high probability, a can be uniquely recovered from $m = \mathcal{O}(s \log(N))$ measurements by solving

$$\min_{z\in\mathbb{C}^N} \|z\|_1 \text{ subject to } \Phi z = y$$

or in the noisy case of $y = \Phi a + w$, the minimizer $\hat{a}$ of

$$\min_{z\in\mathbb{C}^N} \lambda \|z\|_1 + \frac{1}{2} \|\Phi z - y\|_2^2$$

with $\lambda \sim \delta/\sqrt{s}$ and $\|w\| \leqslant \delta$ satisfies $\|a - \hat{a}\|_1 \lesssim \sigma_s(x)_1 + \sqrt{s}\delta$.

In the case where U is unitary, the above statement holds with $\Phi = P_\Omega U$ where Ω are

$$m = \mathcal{O}(N \cdot \mu(U)^2 \cdot s \cdot \log(N))$$

uniformly drawn indices, $\mu(U) = \max_{i,j} |U_{ij}|$ is the so called *coherence*.

In the case of U being the DFT, we have $\mu(U)^2 = 1/N$.

Compressed sensing off the grid

Aim: Recover $\mu_0 \in \mathcal{M}(\mathcal{X})$, $\mathcal{X} \subseteq \mathbb{R}^d$, from m observations, $y = \Phi\mu_0 + w$

- Let (Ω, Λ) be a probability space. For $\omega \in \Omega$, we have random features $\varphi_\omega \in \mathcal{C}(\mathcal{X})$.
- For $k = 1, \ldots, m$, let $\omega_k \overset{iid}{\sim} \Lambda$. The measurement operator is

$$\Phi : \mathcal{M}(\mathcal{X}) \to \mathbb{C}^m, \qquad \Phi\mu \stackrel{\text{def.}}{=} \frac{1}{\sqrt{m}} \left(\int \varphi_{\omega_k}(x) \mathrm{d}\mu(x) \right)_{k=1}^m$$

Typically, the measure of interest is $\mu_0 = \sum_{j=1}^s a_j \delta_{x_j}$ where $a\delta_x$ denotes the Dirac at $x \in \mathcal{X}$ with amplitude $a \in \mathbb{C}$ (also called a "spike").

Imaging

Sampling the Fourier transform (e.g. astronomy)

Recover $\mu \in \mathcal{M}(\mathbb{T}^d)$ from $(\mathcal{F}\mu(\omega_k))_{k=1}^m$ where $\mathcal{F}$ is the Fourier transform and ω_k are drawn iid from $([\![-f_c, f_c]\!]^d, \mathrm{Unif})$.

Here, $\varphi_\omega(x) = \exp\left(-\mathrm{i}2\pi x^\top \omega\right)$ and

$$\Phi\mu_0 = \frac{1}{\sqrt{m}}\left(\sum_{j=1}^{s} a_j \exp\left(-\mathrm{i}2\pi x_j^\top \omega_k\right)\right)_{k=1}^{m}$$

Sampling the Laplace transform (e.g. fluorescence microscopy)

Recover $\mu \in \mathcal{M}(\mathbb{R}_+^d)$ from $(\mathcal{L}\mu(\omega_k))_{k=1}^m$ where $\mathcal{L}$ is the Laplace transform and ω_k are drawn iid from $(\mathbb{R}_+^d, \Lambda_\alpha)$ where $\Lambda_\alpha(\omega) \propto \exp\left(-2\alpha^\top \omega\right)$.

Here, $\varphi_\omega(x) = \exp\left(-x^\top \omega\right)$ and

$$\Phi\mu_0 = \frac{1}{\sqrt{m}}\left(\sum_{j=1}^{s} a_j \exp\left(-x_j^\top \omega_k\right)\right)_{k=1}^{m}$$

Two layer neural network [Bach, 2015]

Let $\Omega \subseteq \mathbb{R}^d$, and $\omega_1, \ldots, \omega_m$ are the training samples drawn from (Ω, Λ), with corresponding values $y_1, \ldots, y_m \in \mathbb{R}$. Find a function of the form

$$f(\omega) = \sum_{j=1}^{s} a_j \max\left(\langle x_j, \omega \rangle, 0\right)$$

where $a_j \in \mathbb{R}$ and $x_j \in \mathbb{R}^d$ such that $f(\omega_j) \approx y_j$ for $j = 1, \ldots, m$. We can then use the function f to predict y given $\omega \in \Omega$.

Two layer neural network [Bach, 2015]

Let $\Omega \subseteq \mathbb{R}^d$, and $\omega_1, \ldots, \omega_m$ are the training samples drawn from (Ω, Λ), with corresponding values $y_1, \ldots, y_m \in \mathbb{R}$. Find a function of the form

$$f(\omega) = \sum_{j=1}^{s} a_j \max\left(\langle x_j, \omega \rangle, 0\right)$$

where $a_j \in \mathbb{R}$ and $x_j \in \mathbb{R}^d$ such that $f(\omega_j) \approx y_j$ for $j = 1, \ldots, m$. We can then use the function f to predict y given $\omega \in \Omega$.

This is precisely our sparse spikes problem where we let $\varphi_\omega(x) = \max\left(\langle x, \omega \rangle, 0\right)$ and

$$\Phi\mu_0 = \left(\sum_{j=1}^{s} a_j \max\left(\langle x_j, \omega_k \rangle, 0\right)\right)_{k=1}^{m}$$

where $\mu_0 = \sum_{j=1}^{s} a_j \delta_{x_j}$.

Density estimation

Task: Given data on $\mathcal{T}$, estimate parameters $(a_i) \in \mathbb{R}_+^N$ and $(x_i)_{i=1}^s \in \mathcal{X}^s$ of a mixture

$$\xi(t) = \sum_{j=1}^{s} a_j \xi_{x_j}(t) = \int_{\mathcal{X}} \xi_x(t) \mathrm{d}\mu_0(x)$$

where $\mu_0 = \sum_j a_j \delta_{x_j}$ where $(\xi_x)_{x \in \mathcal{X}}$ is a family of template distributions. E.g. $x = (m, \sigma) \in \mathcal{X} = \mathbb{R} \times \mathbb{R}_+$ and $\xi_x = \mathcal{N}(m, \sigma^2)$.

Density estimation

Task: Given data on $\mathcal{T}$, estimate parameters $(a_i) \in \mathbb{R}_+^N$ and $(x_i)_{i=1}^s \in \mathcal{X}^s$ of a mixture

$$\xi(t) = \sum_{j=1}^{s} a_j \xi_{x_j}(t) = \int_{\mathcal{X}} \xi_x(t) \mathrm{d}\mu_0(x)$$

where $\mu_0 = \sum_j a_j \delta_{x_j}$ where $(\xi_x)_{x \in \mathcal{X}}$ is a family of template distributions. E.g. $x = (m, \sigma) \in \mathcal{X} = \mathbb{R} \times \mathbb{R}_+$ and $\xi_x = \mathcal{N}(m, \sigma^2)$.

Sketching [Gribonval, Blanchard, Keriven & Traonmilin, 2017]

- No direct access to ξ but n iid samples $(t_1, \ldots, t_n) \in \mathcal{T}^n$ drawn from ξ.
- You do not record this (possibly huge) set of data, but compute online a small set $y \in \mathbb{C}^m$ of m sketches against sketching functions $\theta_\omega(t)$:

$$y_k \overset{\text{def.}}{=} \frac{1}{n} \sum_{j=1}^{n} \theta_{\omega_k}(t_j) \approx \int_{\mathcal{T}} \theta_{\omega_k}(t) \xi(t) \mathrm{d}t = \int_{\mathcal{X}} \int_{\mathcal{T}} \theta_{\omega_k}(t) \xi_x(t) \mathrm{d}t \mathrm{d}\mu_0(x).$$

- So, $\varphi_\omega(x) \overset{\text{def.}}{=} \int_{\mathcal{T}} \theta_{\omega_k}(t) \xi_x(t) \mathrm{d}t$. E.g. $\theta_\omega(t) = e^{\mathrm{i}\langle \omega,\, t \rangle}$ and $\varphi_\cdot(x)$ is the characterisatic function of ξ_x.

The Beurling LASSO

The BLASSO was initially proposed by [De Castro & Gamboa, 2012] and [Bredies & Pikkarainnen, 2013]. Solve

$$\min_{\mu \in \mathcal{M}(\mathcal{X})} \frac{1}{2} \|\Phi\mu - y\|^2 + \lambda |\mu| (\mathcal{X}) \qquad (\hat{\mathcal{P}}_\lambda(y))$$

where $|\mu|(\mathcal{X}) \overset{\text{def.}}{=} \sup \left\{ \mathrm{Re}\left(\langle f, \mu \rangle\right) \; ; \; f \in \mathcal{C}(\mathcal{X}), \; \|f\|_\infty \leqslant 1 \right\}$.

Noiseless problem: for $y_0 = \Phi\mu_0$,

$$\min_{\mu \in \mathcal{M}(\mathcal{X})} |\mu| (\mathcal{X}) \text{ subject to } \Phi\mu = y_0 \qquad (\hat{\mathcal{P}}_0(y_0))$$

NB: If $\mu = \sum_j a_j \delta_{x_j}$, then $|\mu|(\mathcal{X}) = \|a\|_1$.

The Beurling LASSO

The BLASSO was initially proposed by [De Castro & Gamboa, 2012] and [Bredies & Pikkarainnen, 2013]. Solve

$$\min_{\mu \in \mathcal{M}(\mathcal{X})} \frac{1}{2} \|\Phi\mu - y\|^2 + \lambda |\mu|(\mathcal{X}) \qquad (\hat{\mathcal{P}}_\lambda(y))$$

where $|\mu|(\mathcal{X}) \stackrel{\text{def.}}{=} \sup\left\{\mathrm{Re}\left(\langle f, \mu\rangle\right) \ ; \ f \in \mathcal{C}(\mathcal{X}), \ \|f\|_\infty \leqslant 1\right\}$.

Noiseless problem: for $y_0 = \Phi\mu_0$,

$$\min_{\mu \in \mathcal{M}(\mathcal{X})} |\mu|(\mathcal{X}) \text{ subject to } \Phi\mu = y_0 \qquad (\hat{\mathcal{P}}_0(y_0))$$

NB: If $\mu = \sum_j a_j \delta_{x_j}$, then $|\mu|(\mathcal{X}) = \|a\|_1$.

Goal: *A CS-type theory.*
Under what conditions can we recover $\mu_0 = \sum_{j=1}^s a_j \delta_{x_j}$ exactly (stably) from

$$m = \mathcal{O}(s \times \text{ log factors})$$

(noisy) randomised linear measurements?

Remarks

- Other approaches include **Prony-type methods** (1795): MUSIC [Schmidt, 1986], ESPRIT [Roy, 1987], Finite Rate of Innovation [Vetterli, 2002] ...
  - Nonvariational approaches which encodes the spikes positions as the zeros of some polynomial, whose coefficients are derived from the measurements.
  - Generally restricted to Fourier type measurements.
  - Extension to multivariate setting is nontrivial.
- There are efficient algorithms for solving this infinite dimensional problem, e.g. **SDP approaches** [Candès & Fernandez-Granda, 2012; De Castro, Gamboa, Henrion & Lasserre 2015] and **Frank-Wolfe approaches** [Bredies & Pikkarainnen 2013; Boyd, Schiebinger & Recht '15; Denoyelle, Duval & Peyré '18] .

Background on the BLASSO

Recovery of spikes of arbitrary signs require a minimum separation condition:

- [Candès & Fernandez-Granda '12]: Given $\left\{\mathcal{F}\mu_0(k) \; ; \; k \in \mathbb{Z}^d, \; \|k\|_\infty \leqslant f_c\right\}$, μ_0 can be recovered uniquely if $\Delta = \min_{i \neq j} \|x_i - x_j\|_\infty \geqslant \frac{C_d}{f_c}$.
- Many extensions to other measurement operators, minimum separation is *fundamental* (for BLASSO) and often imposed via ad hoc metrics [Bendory et al '15, Tang '15].

Background on the BLASSO

Recovery of spikes of arbitrary signs require a minimum separation condition:

- [Candès & Fernandez-Granda '12]: Given $\{\mathcal{F}\mu_0(k)\ ;\ k \in \mathbb{Z}^d,\ \|k\|_\infty \leqslant f_c\}$, μ_0 can be recovered uniquely if $\Delta = \min_{i\neq j} \|x_i - x_j\|_\infty \geqslant \frac{C_d}{f_c}$.
- Many extensions to other measurement operators, minimum separation is *fundamental* (for BLASSO) and often imposed via ad hoc metrics [Bendory et al '15, Tang '15].

Stability for the recovered measure $\hat{\mu}$:

- Integral type stability estimates [Candès & Fernandez-Granda '13]: $\|K_{\mathrm{hi}} \star (\hat{\mu} - \mu_0)\|_{L_1}$.
- Support concentration [Fernandez-Granda '13; Asaïs, De Castro & Gamboa '12]: Bounds on $\left|\hat{\mu}(\mathcal{X}_j^{\mathrm{near}}) - a_j\right|$ and $|\hat{\mu}|\,(\mathcal{X}^{\mathrm{far}})$.
- Support stability [Duval and Peyré '15]: in the small noise regime where $\|w\|$ and λ are sufficiently small, $\hat{\mu}$ consists of *exactly* s spikes, and the recovered amplitudes and positions vary continuously with respect to λ and w.

Background on the BLASSO

Recovery of spikes of arbitrary signs require a minimum separation condition:

- [Candès & Fernandez-Granda '12]: Given $\left\{\mathcal{F}\mu_0(k)\ ;\ k \in \mathbb{Z}^d,\ \|k\|_\infty \leqslant f_c\right\}$, μ_0 can be recovered uniquely if $\Delta = \min_{i\neq j} \|x_i - x_j\|_\infty \geqslant \frac{C_d}{f_c}$.
- Many extensions to other measurement operators, minimum separation is *fundamental* (for BLASSO) and often imposed via ad hoc metrics [Bendory et al '15, Tang '15].

Stability for the recovered measure $\hat{\mu}$:

- Integral type stability estimates [Candès & Fernandez-Granda '13]: $\|K_{\text{hi}} \star (\hat{\mu} - \mu_0)\|_{L_1}$.
- Support concentration [Fernandez-Granda '13; Asaïs, De Castro & Gamboa '12]: Bounds on $\left|\hat{\mu}(\mathcal{X}_j^{\text{near}}) - a_j\right|$ and $|\hat{\mu}|\,(\mathcal{X}^{\text{far}})$.
- Support stability [Duval and Peyré '15]: in the small noise regime where $\|w\|$ and λ are sufficiently small, $\hat{\mu}$ consists of *exactly* s spikes, and the recovered amplitudes and positions vary continuously with respect to λ and w.

Subsampling in the Fourier setting:

- [Tang et al '13]: If $\text{sign}(a_j)_{j=1}^s$ is a Steinhaus sequence and $\Delta \geqslant \frac{C}{f_c}$, then exact recovery is guaranteed with $\mathcal{O}(s\log(f_c)\log(s))$ number of noiseless random Fourier coefficients.
- Extended to two dimensional setting by [Chi & Chen '15]. So far, removal of the random signs assumption results in $\mathcal{O}(s^2)$ measurements [Li & Chi '17].

Background on the BLASSO

Recovery of spikes of arbitrary signs require a minimum separation cond.:

- [Candès & Fernandez-Granda '12]: Given $\left\{\mathcal{F}\mu_0(k)\ ;\ k \in \mathbb{Z}^d,\ \|k\|_\infty \leqslant f_c\right\}$, μ_0 can be recovered uniquely if $\Delta = \min_{i\neq j} \|x_i - x_j\|_\infty \geqslant \frac{C_d}{f_c}$.
- Many extensions to other measurement operators, minimum separation is *fundamental* (for BLASSO) and often imposed via **ad-hoc** metrics [Bendory et al '15, Tang '15].

Stability for the recovered measure $\hat{\mu}$:

- Integral type stability estimates [Candès & Fernandez-Granda '13]: $\|K_{\mathrm{hi}} \star (\hat{\mu} - \mu_0)\|_{L_1}$.
- Support concentration [Fernandez-Granda '13; Asaïs, De Castro & Gamboa '12]: Bounds on $\left|\hat{\mu}(\mathcal{X}_j^{\mathrm{near}}) - a_j\right|$ and $|\hat{\mu}|(\mathcal{X}^{\mathrm{far}})$.
- Support stability [Duval and Peyré '15]: in the small noise regime where $\|w\|$ and λ are sufficiently small, $\hat{\mu}$ consists of *exactly* s spikes, and the recovered amplitudes and positions vary continuously with respect to λ and w.

Subsampling in the Fourier setting:

- [Tang et al '13]: If $\mathrm{sign}(a_j)_{j=1}^s$ is a Steinhaus sequence and $\Delta \geqslant \frac{C}{f_c}$, then exact recovery is guaranteed with $\mathcal{O}(s\log(f_c)\log(s))$ number of noiseless random Fourier coefficients.
- Extended to two dimensional setting by [Chi & Chen '15]. So far, removal of the random signs assumption results in $\mathcal{O}(s^2)$ measurements [Li & Chi '17].

Background on the BLASSO

Recovery of spikes of arbitrary signs require a minimum separation cond.:

- [Candès & Fernandez-Granda '12]: Given $\{\mathcal{F}\mu_0(k)\ ;\ k \in \mathbb{Z}^d,\ \|k\|_\infty \leqslant f_c\}$, μ_0 can be recovered uniquely if $\Delta = \min_{i \neq j} \|x_i - x_j\|_\infty \geqslant \frac{C_d}{f_c}$.
- Many extensions to other measurement operators, minimum separation is *fundamental* (for BLASSO) and often imposed via **ad-hoc** metrics [Bendory et al '15, Tang '15].

Stability for the recovered measure $\hat{\mu}$:

- Integral type stability estimates [Candès & Fernandez-Granda '13]: $\|K_{\text{hi}} \star (\hat{\mu} - \mu_0)\|_{L_1}$.
- Support concentration [Fernandez-Granda '13; Asaïs, De Castro & Gamboa '12]: Bounds on $\left|\hat{\mu}(\mathcal{X}_j^{\text{near}}) - a_j\right|$ and $|\hat{\mu}|\,(\mathcal{X}^{\text{far}})$.
- **Support stability** [Duval and Peyré '15]: in the small noise regime where $\|w\|$ and λ are sufficiently small, $\hat{\mu}$ consists of *exactly* s spikes, and the recovered amplitudes and positions vary continuously with respect to λ and w.

Subsampling in the Fourier setting:

- [Tang et al '13]: If $\text{sign}(a_j)_{j=1}^s$ is a Steinhaus sequence and $\Delta \geqslant \frac{C}{f_c}$, then exact recovery is guaranteed with $\mathcal{O}(s \log(f_c) \log(s))$ number of **noiseless** random Fourier coefficients.
- Extended to two dimensional setting by [Chi & Chen '15]. So far, removal of the random signs assumption results in $\mathcal{O}(s^2)$ measurements [Li & Chi '17].

Background on the BLASSO

Recovery of spikes of arbitrary signs require a minimum separation cond.:

- [Candès & Fernandez-Granda '12]: Given $\{\mathcal{F}\mu_0(k)\ ;\ k \in \mathbb{Z}^d,\ \|k\|_\infty \leqslant f_c\}$, μ_0 can be recovered uniquely if $\Delta = \min_{i\neq j} \|x_i - x_j\|_\infty \geqslant \frac{C_d}{f_c}$.
- Many extensions to other measurement operators, minimum separation is *fundamental* (for BLASSO) and often imposed via **ad-hoc** metrics [Bendory et al '15, Tang '15].

Stability for the recovered measure $\hat{\mu}$:

- Integral type stability estimates [Candès & Fernandez-Granda '13]: $\|K_{\mathrm{hi}} \star (\hat{\mu} - \mu_0)\|_{L_1}$.
- **Support concentration** [Fernandez-Granda '13; Asaïs, De Castro & Gamboa '12]: Bounds on $\left|\hat{\mu}(\mathcal{X}_j^{\mathrm{near}}) - a_j\right|$ and $|\hat{\mu}|\,(\mathcal{X}^{\mathrm{far}})$.
- **Support stability** [Duval and Peyré '15]: in the small noise regime where $\|w\|$ and λ are sufficiently small, $\hat{\mu}$ consists of *exactly* s spikes, and the recovered amplitudes and positions vary continuously with respect to λ and w.

Subsampling in the Fourier setting:

- [Tang et al '13]: If $\mathrm{sign}(a_j)_{j=1}^s$ is a **Steinhaus sequence** and $\Delta \geqslant \frac{C}{f_c}$, then exact recovery is guaranteed with $\mathcal{O}(s \log(f_c) \log(s))$ number of **noiseless** random Fourier coefficients.
- Extended to two dimensional setting by [Chi & Chen '15]. So far, removal of the random signs assumption results in $\mathcal{O}(s^2)$ measurements [Li & Chi '17].

Outline

The covariance kernel

The covariance kernel

Define the covariance kernel: $\hat{K}(x,x') \overset{\text{def.}}{=} \frac{1}{m}\sum_{k=1}^m \overline{\varphi_{\omega_k}(x)}\varphi_{\omega_k}(x')$, and the limit covariance kernel as $K(x,x') \overset{\text{def.}}{=} \mathbb{E}[\hat{K}(x,x')] = \int \overline{\varphi_\omega(x)}\varphi_\omega(x')\mathrm{d}\Lambda(\omega)$.

Denote $\hat{f} \overset{\text{def.}}{=} \Phi^* y = \int \hat{K}(x,x')\mathrm{d}\mu_0(x') + \Phi^* w \in \mathcal{C}(\mathcal{X})$. The BLASSO can be rewritten as

$$\min_{\mu\in\mathcal{M}(\mathcal{X})} \frac{1}{2}\int \hat{K}(x,x')\mathrm{d}\overline{\mu}(x)\mathrm{d}\mu(x') - \mathrm{Re}\langle \hat{f},\, \mu\rangle + \lambda\, |\mu|\,(\mathcal{X}) \tag{$\hat{\mathcal{P}}_\lambda(y)$}$$

and

$$\min_{\mu\in\mathcal{M}(\mathcal{X})} |\mu|\,(\mathcal{X}) \text{ subject to } \int \hat{K}(x,x')\mathrm{d}\overline{(\mu-\mu_0)}(x)\mathrm{d}(\mu-\mu_0)(x') = 0. \tag{$\hat{\mathcal{P}}_0(y_0)$}$$

The covariance kernel

The covariance kernel

Define the covariance kernel: $\hat{K}(x,x') \stackrel{\text{def.}}{=} \frac{1}{m}\sum_{k=1}^m \overline{\varphi_{\omega_k}(x)}\varphi_{\omega_k}(x')$, and the limit covariance kernel as $K(x,x') \stackrel{\text{def.}}{=} \mathbb{E}[\hat{K}(x,x')] = \int \overline{\varphi_\omega(x)}\varphi_\omega(x')\mathrm{d}\Lambda(\omega)$.

Denote $\hat{f} \stackrel{\text{def.}}{=} \Phi^* y = \int \hat{K}(x,x')\mathrm{d}\mu_0(x') + \Phi^* w \in \mathcal{C}(\mathcal{X})$. The BLASSO can be rewritten as

$$\min_{\mu\in\mathcal{M}(\mathcal{X})} \frac{1}{2}\int \hat{K}(x,x')\mathrm{d}\overline{\mu}(x)\mathrm{d}\mu(x') - \mathrm{Re}\langle \hat{f},\, \mu\rangle + \lambda\,|\mu|\,(\mathcal{X}) \qquad (\hat{\mathcal{P}}_\lambda(y))$$

and

$$\min_{\mu\in\mathcal{M}(\mathcal{X})} |\mu|\,(\mathcal{X}) \text{ subject to } \int \hat{K}(x,x')\mathrm{d}\overline{(\mu-\mu_0)}(x)\mathrm{d}(\mu-\mu_0)(x') = 0. \qquad (\hat{\mathcal{P}}_0(y_0))$$

Before discussing the role of subsampling, let's look at the limit problem associated to K.

What separation conditions should we impose to guarantee recovery of $\mu_0 = \sum_{j=1}^s a_j\delta_{x_j}$?

The covariance kernel

The covariance kernel

Define the covariance kernel: $\hat{K}(x,x') \stackrel{\text{def.}}{=} \frac{1}{m}\sum_{k=1}^m \overline{\varphi_{\omega_k}(x)}\varphi_{\omega_k}(x')$, and the limit covariance kernel as $K(x,x') \stackrel{\text{def.}}{=} \mathbb{E}[\hat{K}(x,x')] = \int \overline{\varphi_\omega(x)}\varphi_\omega(x')\mathrm{d}\Lambda(\omega)$.

Denote $f \stackrel{\text{def.}}{=} \int K(x,x')\mathrm{d}\mu_0(x') + \varepsilon \in \mathcal{C}(\mathcal{X})$. The BLASSO can be rewritten as

$$\min_{\mu\in\mathcal{M}(\mathcal{X})} \frac{1}{2}\int K(x,x')\mathrm{d}\overline{\mu}(x)\mathrm{d}\mu(x') - \mathrm{Re}\langle f,\, \mu\rangle + \lambda\,|\mu|\,(\mathcal{X}) \qquad (\mathcal{P}_\lambda(f))$$

and

$$\min_{\mu\in\mathcal{M}(\mathcal{X})} |\mu|\,(\mathcal{X}) \text{ subject to } \int K(x,x')\mathrm{d}\overline{(\mu-\mu_0)}(x)\mathrm{d}(\mu-\mu_0)(x') = 0. \qquad (\mathcal{P}_0(f_0))$$

Before discussing the role of subsampling, let's look at the limit problem associated to K.

What separation conditions should we impose to guarantee recovery of $\mu_0 = \sum_{j=1}^s a_j\delta_{x_j}$?

The Fisher metric

Assume that for all $x \in \mathcal{X}$, $\mathbb{E}_\omega[|\varphi_\omega(x)|^2] = 1$. Let $\mathbf{H}_x \stackrel{\text{def.}}{=} \nabla_1 \nabla_2 K(x,x) \in \mathbb{C}^{d \times d}$ and assume that $\mathbf{H}_x$ is positive definite for all $x \in \mathcal{X}$.

The Fisher metric

Assume that for all $x \in \mathcal{X}$, $\mathbb{E}_\omega[|\varphi_\omega(x)|^2] = 1$. Let $\mathbf{H}_x \stackrel{\text{def.}}{=} \nabla_1 \nabla_2 K(x, x) \in \mathbb{C}^{d \times d}$ and assume that $\mathbf{H}_x$ is positive definite for all $x \in \mathcal{X}$.

- $f(x, \omega) \stackrel{\text{def.}}{=} |\varphi_\omega(x)|^2$ can be interpreted as a probability density function for the random variable ω conditional on parameter $x \in \mathcal{X}$ and its Fisher information matrix is:

$$\int \nabla \left(\log f(x, \omega)\right) \nabla \left(\log f(x, \omega)\right)^\top f(x, \omega) \mathrm{d}\Lambda(\omega) = 4\mathbf{H}_x.$$

The Fisher metric

Assume that for all $x \in \mathcal{X}$, $\mathbb{E}_\omega[|\varphi_\omega(x)|^2] = 1$. Let $\mathbf{H}_x \overset{\text{def.}}{=} \nabla_1 \nabla_2 K(x,x) \in \mathbb{C}^{d\times d}$ and assume that $\mathbf{H}_x$ is positive definite for all $x \in \mathcal{X}$.

- $f(x,\omega) \overset{\text{def.}}{=} |\varphi_\omega(x)|^2$ can be interpreted as a probability density function for the random variable ω conditional on parameter $x \in \mathcal{X}$ and its Fisher information matrix is:

$$\int \nabla \left(\log f(x,\omega)\right) \nabla \left(\log f(x,\omega)\right)^\top f(x,\omega) \mathrm{d}\Lambda(\omega) = 4\mathbf{H}_x.$$

- $\mathbf{H}$ naturally induces a distance between points on $\mathcal{X}$. Given a curve $\gamma : [0,1] \to \mathcal{X}$, $\ell_{\mathbf{H}}[\gamma] \overset{\text{def.}}{=} \int_0^1 \sqrt{\langle \mathbf{H}_{\gamma(t)} \gamma'(t),\, \gamma'(t)\rangle} \mathrm{d}t$ and given $x, x' \in \mathcal{X}$,

$$d_{\mathbf{H}}(x,x') \overset{\text{def.}}{=} \inf \left\{ \ell_{\mathbf{H}}[\gamma]\ ;\ \gamma : [0,1] \to \mathcal{X}, \gamma(0) = x, \gamma(1) = x' \right\}.$$

Also called the "Fisher-Rao" geodesic distance, this is used extensively in information geometry for estimation and learning problems on parametric families of distributions (Amari and Nagaoka, 2007).

The Fisher metric

Assume that for all $x \in \mathcal{X}$, $\mathbb{E}_\omega[|\varphi_\omega(x)|^2] = 1$. Let $\mathbf{H}_x \stackrel{\text{def.}}{=} \nabla_1 \nabla_2 K(x,x) \in \mathbb{C}^{d\times d}$ and assume that $\mathbf{H}_x$ is positive definite for all $x \in \mathcal{X}$.

- $f(x,\omega) \stackrel{\text{def.}}{=} |\varphi_\omega(x)|^2$ can be interpreted as a probability density function for the random variable ω conditional on parameter $x \in \mathcal{X}$ and its Fisher information matrix is:

$$\int \nabla\left(\log f(x,\omega)\right) \nabla\left(\log f(x,\omega)\right)^\top f(x,\omega) \mathrm{d}\Lambda(\omega) = 4\mathbf{H}_x.$$

- $\mathbf{H}$ naturally induces a distance between points on $\mathcal{X}$. Given a curve $\gamma : [0,1] \to \mathcal{X}$, $\ell_{\mathbf{H}}[\gamma] \stackrel{\text{def.}}{=} \int_0^1 \sqrt{\langle \mathbf{H}_{\gamma(t)}\gamma'(t), \gamma'(t)\rangle}\mathrm{d}t$ and given $x, x' \in \mathcal{X}$,

$$d_{\mathbf{H}}(x,x') \stackrel{\text{def.}}{=} \inf\left\{\ell_{\mathbf{H}}[\gamma]\ ;\ \gamma : [0,1] \to \mathcal{X}, \gamma(0) = x, \gamma(1) = x'\right\}.$$

Also called the "Fisher-Rao" geodesic distance, this is used extensively in information geometry for estimation and learning problems on parametric families of distributions (Amari and Nagaoka, 2007).

Theorem

Under some generic conditions on K and Δ, if $\min_{j\neq k} d_{\mathbf{H}}(x_j, x_k) \geqslant \Delta$ and $s \leqslant s_{\max}$, then μ_0 can be exactly (stably) recovered as a solution to $\mathcal{P}_0(f)$ (to $\mathcal{P}_\lambda(f)$).

Notation for derivatives

We can interpret the r^{th} derivative as a multilinear map $\nabla^r f : (\mathbb{C}^d)^r \to \mathbb{C}$, given $Q = \{q_\ell\}_{\ell=1}^r \in (\mathbb{C}^d)^r$,

$$\nabla^r f[Q] = \sum_{i_1,\ldots,i_r} \partial_{i_1} \cdots \partial_{i_r} f(x) q_{i_1} \cdots q_{i_r}.$$

The normalised r^{th} derivative is

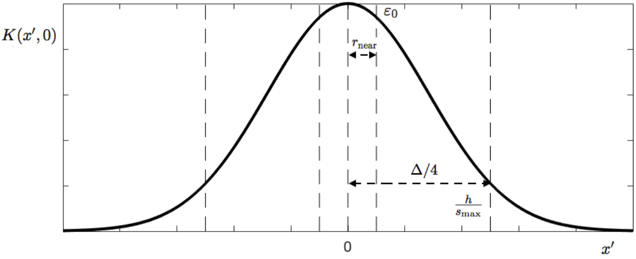
$$\mathrm{D}_r[f](x)[Q] = \nabla^r f(x)[\{\mathbf{H}_x^{-\frac{1}{2}} q_i\}_{i=1}^r].$$

and $K^{ij}(x, x') : (\mathbb{C}^d)^i \times (\mathbb{C}^d)^j \to \mathbb{C}$ is defined by

$$K^{(ij)}[Q, V] \stackrel{\text{def.}}{=} \mathbb{E}\left(\overline{\mathrm{D}_i[\varphi_\omega][Q]} \cdot \mathrm{D}_j[\varphi_\omega][V]\right).$$

Admissible kernels *

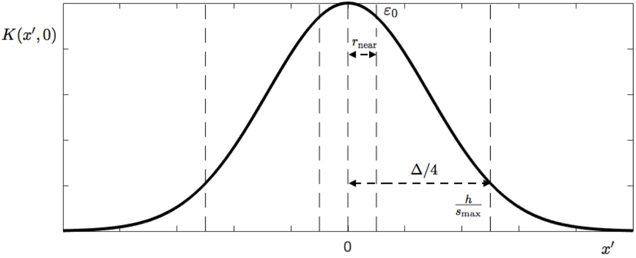
A kernel K will be said *admissible* with respect to $\{r_{\text{near}}, \Delta, \varepsilon_i, B_{ij}, s_{\max}\}$, if



*For simplicity, assume that K is real-valued.

Admissible kernels

A kernel K will be said *admissible* with respect to $\{r_{\text{near}}, \Delta, \varepsilon_i, B_{ij}, s_{\max}\}$, if

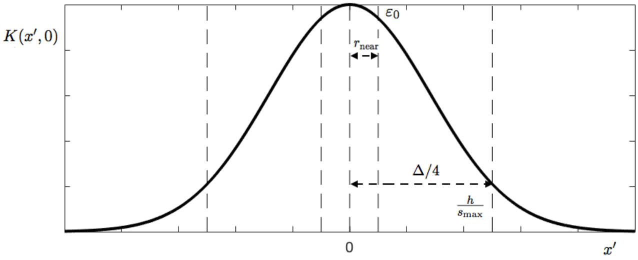


Sufficient peak:

- For $d_{\mathbf{H}}(x,x') \geqslant r_{\text{near}}$, $|K(x,x')| \leqslant 1-\varepsilon_0$.
- For $d_{\mathbf{H}}(x,x') \leqslant r_{\text{near}}$, $K^{(02)}(x,x') \preccurlyeq -\varepsilon_2 \mathrm{Id}$

Admissible kernels

A kernel K will be said *admissible* with respect to $\{r_{\text{near}}, \Delta, \varepsilon_i, B_{ij}, s_{\max}\}$, if



Sufficient peak:

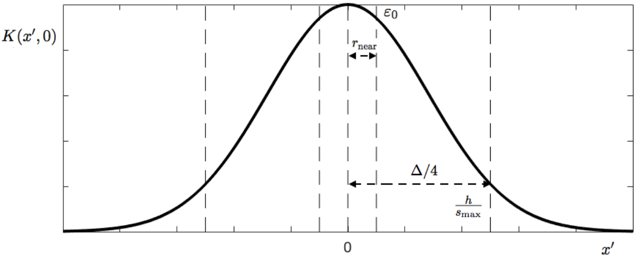
- For $d_{\mathbf{H}}(x, x') \geqslant r_{\text{near}}$, $|K(x, x')| \leqslant 1 - \varepsilon_0$.
- For $d_{\mathbf{H}}(x, x') \leqslant r_{\text{near}}$, $K^{(02)}(x, x') \preccurlyeq -\varepsilon_2 \text{Id}$

Sufficient decay:

- For $d_{\mathbf{H}}(x, x') \geqslant \Delta/4$, $\left\| K^{(ij)}(x, x') \right\| \leqslant \frac{h}{s_{\max}}$, where $i, j \in \{0, \ldots, 2\}$ with $i + j \leqslant 3$, $h \overset{\text{def.}}{=} \min_{i \in \{0,2\}} \left(\frac{\varepsilon_i}{32 B_{1i} + 32} \right)$.

Admissible kernels

A kernel K will be said *admissible* with respect to $\{r_{\text{near}}, \Delta, \varepsilon_i, B_{ij}, s_{\max}\}$, if



Sufficient peak:

- For $d_{\mathbf{H}}(x,x') \geqslant r_{\text{near}}$, $|K(x,x')| \leqslant 1-\varepsilon_0$.
- For $d_{\mathbf{H}}(x,x') \leqslant r_{\text{near}}$, $K^{(02)}(x,x') \preccurlyeq -\varepsilon_2 \mathrm{Id}$

Sufficient decay:

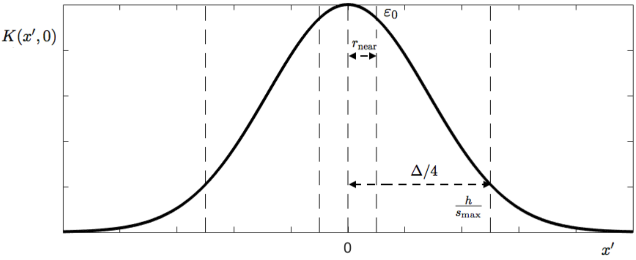
- For $d_{\mathbf{H}}(x,x') \geqslant \Delta/4$, $\left\|K^{(ij)}(x,x')\right\| \leqslant \frac{h}{s_{\max}}$, where $i,j \in \{0,\ldots,2\}$ with $i+j \leqslant 3$, $h \overset{\text{def.}}{=} \min_{i\in\{0,2\}} \left(\frac{\varepsilon_i}{32B_{1i}+32}\right)$.

Uniform bounds: $\sup_{x,x'\in\mathcal{X}} \left\|K^{(ij)}(x,x')\right\| \leqslant B_{ij}$ for $i,j \in \{0,1,2\}$.

Additionally, for $d_{\mathbf{H}}(x,x') \leqslant r_{\text{near}}$: $\left\|\mathrm{Id} - \mathbf{H}_{x'}^{-\frac{1}{2}} \mathbf{H}_x^{\frac{1}{2}}\right\| \leqslant C_{\mathbf{H}} d_{\mathbf{H}}(x,x')$.

Admissible kernels

A kernel K will be said *admissible* with respect to $\{r_{\text{near}}, \Delta, \varepsilon_i, B_{ij}, s_{\max}\}$, if

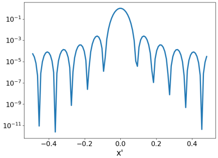
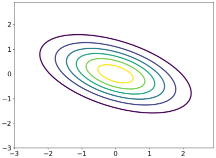
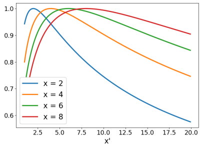


Theorem

Suppose that K is admissible, and $\mu_0 = \sum_{j=1}^{s} a_j \delta_{x_j}$ with $\min_{j \neq k} d_{\mathbf{H}}(x_j, x_k) \geqslant \Delta$ and $s \leqslant s_{\max}$. Then, μ_0 can be exactly (stably) recovered as a solution to $\mathcal{P}_0(f)$ (to $\mathcal{P}_\lambda(f)$).

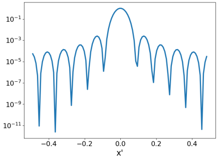
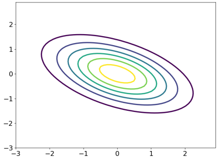
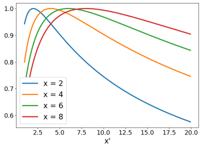
NB: in general, ε_i, r_{near}, B_{ij}, $C_{\mathbf{H}}$ are just constants (possibly dependent on d), but independent of problem parameters.

Examples

Discrete Fourier	*Continuous Fourier*	*Microscopy (Laplace trans.)*
Random features		
$\varphi_\omega(x) = e^{\mathrm{i}2\pi\omega^\top x}$ $\Lambda \propto \prod_j g(\omega_j)$ on $[\![-f_c, f_c]\!]^d$	$\varphi_\omega(x) = e^{\mathrm{i}\omega^\top x}$ $\Lambda = \mathcal{N}(0, \Sigma)$ on $\mathbb{R}^d$	$\varphi_\omega(x) = \prod_{i=1}^d \sqrt{\frac{2(x_i+\alpha_i)}{\alpha_i}} e^{-\omega^\top x}$ $\Lambda \propto e^{-\alpha^\top \omega}$ on $\mathbb{R}_+^d$
Kernel		
Jackson $\prod_i \kappa(x_i - x_i')$ 	Gaussian † $e^{-\|x-x'\|_\Sigma}$ 	$\prod_i \kappa(x_i + \alpha_i, x_i' + \alpha_i)$,  $\kappa(x, x') = \frac{\sqrt{4xx'}}{x+x'}$
Metric and separation		
$\mathbf{H}_x = C_{f_c} \mathrm{Id}$ ‡ $d_{\mathbf{H}}(x, x') = C_{f_c}^{\frac{1}{2}} \|x - x'\|_2$ $\Delta = \sqrt{d\sqrt{s}}$	$\mathbf{H}_x = \Sigma$ $d_{\mathbf{H}}(x, x') = \|x - x'\|_\Sigma$ $\Delta = \sqrt{\log(s)}$	$\mathbf{H}_x = \mathrm{diag}\left(\frac{1}{4(x_i+\alpha_i)^2}\right)$ $d_{\mathbf{H}}(x, x') = \sqrt{\sum_i \left|\log\left(\frac{x_i+\alpha_i}{x_i'+\alpha_i}\right)\right|^2}$ $\Delta = d + \log(ds)$

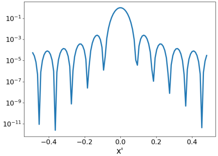
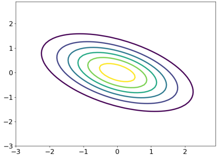
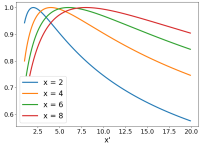
‡ $C_{f_c} = \frac{\pi^2}{3} f_c(f_c + 4) \sim f_c^2$

Examples

Discrete Fourier	*Continuous Fourier*	*Microscopy (Laplace trans.)*
Random features		
$\varphi_\omega(x) = e^{\mathrm{i}2\pi\omega^\top x}$ $\Lambda \propto \prod_j g(\omega_j)$ on $[\![-f_c, f_c]\!]^d$	$\varphi_\omega(x) = e^{\mathrm{i}\omega^\top x}$ $\Lambda = \mathcal{N}(0, \Sigma)$ on $\mathbb{R}^d$	$\varphi_\omega(x) = \prod_{i=1}^d \sqrt{\frac{2(x_i+\alpha_i)}{\alpha_i}} e^{-\omega^\top x}$ $\Lambda \propto e^{-\alpha^\top \omega}$ on $\mathbb{R}_+^d$
Kernel		
Jackson $\prod_i \kappa(x_i - x_i')$ 	Gaussian † $e^{-\|x-x'\|_\Sigma}$ 	$\prod_i \kappa(x_i + \alpha_i, x_i' + \alpha_i)$,  $\kappa(x, x') = \frac{\sqrt{4xx'}}{x+x'}$
Metric and separation		
$\mathbf{H}_x = C_{f_c} \mathrm{Id}$ ‡ $d_{\mathbf{H}}(x, x') = C_{f_c}^{\frac{1}{2}} \|x - x'\|_2$ $\Delta = \sqrt{d\sqrt{s}}$	$\mathbf{H}_x = \Sigma$ $d_{\mathbf{H}}(x, x') = \|x - x'\|_\Sigma$ $\Delta = \sqrt{\log(s)}$	$\mathbf{H}_x = \mathrm{diag}\left(\frac{1}{4(x_i+\alpha_i)^2}\right)$ $d_{\mathbf{H}}(x, x') = \sqrt{\sum_i \left\|\log\left(\frac{x_i+\alpha_i}{x_i'+\alpha_i}\right)\right\|^2}$ $\Delta = d + \log(ds)$

Our result: $\|x_i - x_j\| \gtrsim \frac{\sqrt{d}\sqrt[4]{s}}{f_c}$, Candès & Fernandez-Granda: $\|x_i - x_j\| \gtrsim \frac{C_d}{f_c}$

Examples

<table>
<tr><td>Discrete Fourier</td><td>Continuous Fourier</td><td>Microscopy (Laplace trans.)</td></tr>
<tr><td colspan="3">Random features</td></tr>
<tr><td>$\varphi_\omega(x) = e^{\mathrm{i}2\pi\omega^\top x}$
$\Lambda \propto \prod_j g(\omega_j)$ on $[\![-f_c, f_c]\!]^d$</td><td>$\varphi_\omega(x) = e^{\mathrm{i}\omega^\top x}$
$\Lambda = \mathcal{N}(0, \Sigma)$ on $\mathbb{R}^d$</td><td>$\varphi_\omega(x) = \prod_{i=1}^d \sqrt{\frac{2(x_i+\alpha_i)}{\alpha_i}} e^{-\omega^\top x}$
$\Lambda \propto e^{-\alpha^\top \omega}$ on $\mathbb{R}^d_+$</td></tr>
<tr><td colspan="3">Kernel</td></tr>
<tr><td>Jackson $\prod_i \kappa(x_i - x_i')$

</td><td>Gaussian † $e^{-\|x-x'\|_\Sigma}$

</td><td>$\prod_i \kappa(x_i + \alpha_i, x_i' + \alpha_i)$,


$\kappa(x, x') = \frac{\sqrt{4xx'}}{x+x'}$</td></tr>
<tr><td colspan="3">Metric and separation</td></tr>
<tr><td>$\mathbf{H}_x = C_{f_c}\mathrm{Id}$ ‡
$d_{\mathbf{H}}(x, x') = C_{f_c}^{\frac{1}{2}} \|x - x'\|_2$
$\Delta = \sqrt{d\sqrt{s}}$</td><td>$\mathbf{H}_x = \Sigma$
$d_{\mathbf{H}}(x, x') = \|x - x'\|_\Sigma$
$\Delta = \sqrt{\log(s)}$</td><td>$\mathbf{H}_x = \mathrm{diag}\left(\frac{1}{4(x_i+\alpha_i)^2}\right)$
$d_{\mathbf{H}}(x, x') = \sqrt{\sum_i \left|\log\left(\frac{x_i+\alpha_i}{x_i'+\alpha_i}\right)\right|^2}$
$\Delta = d + \log(ds)$</td></tr>
</table>

† $\|x\|_\Sigma = \langle \Sigma x,\ x\rangle$

Outline

The subsampled setting

Assumption 1

- K is admissible, $\mu_0 = \sum_{j=1}^{s} a_j \delta_{x_j}$ with $\min_{j \neq k} d_{\mathbf{H}}(x_j, x_k) \geqslant \Delta$ and $s \leqslant s_{\max}$.
- $\mathcal{X}$ is a compact domain wth $R_{\mathcal{X}} \overset{\text{def.}}{=} \sup_{x, x' \in \mathcal{X}} d_{\mathbf{H}}(x, x')$,

The subsampled setting

Assumption 1

- K is admissible, $\mu_0 = \sum_{j=1}^{s} a_j \delta_{x_j}$ with $\min_{j \neq k} d_{\mathbf{H}}(x_j, x_k) \geqslant \Delta$ and $s \leqslant s_{\max}$.
- $\mathcal{X}$ is a compact domain wth $R_{\mathcal{X}} \stackrel{\text{def.}}{=} \sup_{x,x' \in \mathcal{X}} d_{\mathbf{H}}(x, x')$,

To analyse the subsampled case, we need to control the deviation of $\hat{K}$ from K.

Ideally, $L_r(\omega) = \sup_{x \in \mathcal{X}} \|\mathrm{D}_r[\varphi_\omega](x)\|$ are uniformly bounded. But...

$$\varphi_\omega(x) = \exp(\mathrm{i}\omega^\top x) \implies \|\mathrm{D}_r[\varphi_\omega](x)\| \propto \|\omega\|_{\Sigma^{-1}}^r \,,$$

on the other hand $\mathbb{P}(\|\omega\|_{\Sigma^{-1}} > x) \leqslant 2^{d/2} e^{-x/4}$.

The subsampled setting

Assumption 1

- K is admissible, $\mu_0 = \sum_{j=1}^{s} a_j \delta_{x_j}$ with $\min_{j \neq k} d_{\mathbf{H}}(x_j, x_k) \geqslant \Delta$ and $s \leqslant s_{\max}$.
- $\mathcal{X}$ is a compact domain wth $R_{\mathcal{X}} \stackrel{\text{def.}}{=} \sup_{x,x' \in \mathcal{X}} d_{\mathbf{H}}(x, x')$,

Let $L_r(\omega) \stackrel{\text{def.}}{=} \sup_{x \in \mathcal{X}} \|\mathrm{D}_r[\varphi_\omega](x)\|$

Assumption 2

With high probability, $L_r(\omega_k) \leqslant \bar{L}_r$ for $r = 0, 1, 2, 3$ and $k = 1, \ldots, m$.
and either one of the following hold:

- $\mathrm{sign}(a)$ is a Steinhaus sequence and $m \gtrsim C \cdot s \cdot \log\left(\frac{N^d}{\rho}\right) \log\left(\frac{s}{\rho}\right)$,
- $\mathrm{sign}(a)$ is an arbitrary sign sequence and $m \gtrsim C \cdot s^{3/2} \cdot \log\left(\frac{N^d}{\rho}\right)$,

where $C \stackrel{\text{def.}}{=} \varepsilon^{-2}(\mathbb{L}_2^2 B_{11} + \mathbb{L}_1^2 B_{22} + \mathbb{L}_1^2 B)$, $N \stackrel{\text{def.}}{=} \frac{d \cdot R_{\mathcal{X}} \cdot \mathbb{L}_3}{r_{\text{near}}\varepsilon}$.

$B = B_{00} + B_{02} + B_{10} + B_{12}$, $\varepsilon = \min\{\varepsilon_0, \varepsilon_2\}$, $\mathbb{L}_r = \max_{i \leqslant r} \bar{L}_i$

The subsampled setting

Assumption 1

- K is admissible, $\mu_0 = \sum_{j=1}^{s} a_j \delta_{x_j}$ with $\min_{j \neq k} d_{\mathbf{H}}(x_j, x_k) \geqslant \Delta$ and $s \leqslant s_{\max}$.
- $\mathcal{X}$ is a compact domain wth $R_{\mathcal{X}} \overset{\text{def.}}{=} \sup_{x,x' \in \mathcal{X}} d_{\mathbf{H}}(x, x')$,

Let $L_r(\omega) \overset{\text{def.}}{=} \sup_{x \in \mathcal{X}} \|\mathrm{D}_r[\varphi_\omega](x)\|$ and let F_r be such that $\mathbb{P}_\omega(L_r(\omega) > t) \leqslant F_r(t)$.

Assumption 2

For $\rho > 0$ (probability of failure) choose $m \in \mathbb{N}$ (number of measurements), and $\{\bar{L}_i\}_{i=0}^{3}$ such that

$$\sum_{j=0}^{3} F_j(\bar{L}_j) \leqslant \frac{\rho}{m} \quad \text{and} \quad \bar{L}_j^2 \sum_{i=0}^{3} F_i(\bar{L}_i) + 2 \int_{\bar{L}_j}^{\infty} t F_j(t) \mathrm{d}t \leqslant \frac{\varepsilon}{m}.$$

and either one of the following hold:

- $\mathrm{sign}(a)$ is a Steinhaus sequence and $m \gtrsim C \cdot s \cdot \log\left(\frac{N^d}{\rho}\right) \log\left(\frac{s}{\rho}\right)$,
- $\mathrm{sign}(a)$ is an arbitrary sign sequence and $m \gtrsim C \cdot s^{3/2} \cdot \log\left(\frac{N^d}{\rho}\right)$,

where $C \overset{\text{def.}}{=} \varepsilon^{-2}(\mathbb{L}_2^2 B_{11} + \mathbb{L}_1^2 B_{22} + \mathbb{L}_1^2 B)$, $N \overset{\text{def.}}{=} \frac{d \cdot R_{\mathcal{X}} \cdot \mathbb{L}_3}{r_{\text{near}} \varepsilon}$.

$B = B_{00} + B_{02} + B_{10} + B_{12}$, $\varepsilon = \min\{\varepsilon_0, \varepsilon_2\}$, $\mathbb{L}_r = \max_{i \leqslant r} \bar{L}_i$

Support stability statement

Theorem

Let $\mathcal{D}_{\lambda_0,c_0} \stackrel{\text{def.}}{=} \left\{(\lambda, w) \in \mathbb{R}_+ \times \mathbb{C}^m \ ; \ \lambda \leqslant \frac{D}{s}, \ \|w\| \leqslant c_0\lambda\right\}$ *where*

$$D \sim \underline{a} \min\left(r_{\text{near}}\sqrt{s}, \ \frac{\varepsilon\sqrt{s}}{\mathbb{L}_2^2\|a\|}, \ \frac{\varepsilon}{C_{\mathbf{H}}(B+\mathbb{L}_2^2)}\right) \quad \textit{and} \quad c_0 \sim \min\left(\frac{\varepsilon_0}{\bar{L}_0}, \ \frac{\varepsilon_2}{\bar{L}_2}\right) \tag{3.1}$$

and $\underline{a} = \min\{|a_i|, |a_i|^{-1}\}$.

Then, with probability at least $1-\rho$,

(i) *for all* $v \stackrel{\text{def.}}{=} (\lambda, w) \in \mathcal{D}_{\lambda_0,c_0}$, $(\hat{P}_\lambda(y))$ *has a unique solution which consists of exactly* s *spikes.*

(ii) *The mapping* $v \in \mathcal{D}_{\lambda_0,c_0} \mapsto (\hat{a}^v, \{\hat{x}_j^v\}_{j=1}^s)$ *is continuously differentiable and we have the error bound*

$$\|\hat{a}^v - a\| + \sqrt{\sum_j d_{\mathbf{H}}^2(\hat{x}_j^v, x_{0,j})} \leqslant \frac{\sqrt{s}(\lambda+\|w\|)}{\min_i |a_i|} \tag{3.2}$$

Examples

Discrete Fourier	*Continuous Fourier*	*Laplace transform*
Random features		
$\varphi_\omega(x) = e^{\mathrm{i}2\pi\omega^\top x}$ $\Lambda \propto \prod_{j=1}^d g_j(\omega_j)$	$\varphi_\omega(x) = e^{\mathrm{i}\omega^\top x}$ $\Lambda = \mathcal{N}(0, \Sigma)$	$\varphi_\omega(x) = \prod_{i=1}^d \sqrt{\frac{2(x_i+\alpha_i)}{\alpha_i}} e^{-\omega^\top x}$ $\Lambda \propto e^{-\alpha^\top \omega}$
No. samples (up to $\log$ **factors),** $p = 1$ **for random signs,** $p = 3/2$ **in general**		
Rand. sgn.: $\mathcal{O}(sd^3)$ General: $\mathcal{O}(s^{3/2}d^3)$	Rand. sgn.: $\mathcal{O}(sd^3)$ General: $\mathcal{O}(s^{3/2}d^3)$	Rand. sgn.: $\mathcal{O}(sd^7)$ General: $\mathcal{O}(s^{3/2}d^7)$
Stability regions		
$\lambda = \mathcal{O}(s^{-1}d^{-2})$ $\lVert w \rVert = \mathcal{O}(s^{-1}d^{-3})$	$\lambda = \mathcal{O}(s^{-1}d^{-2})$ $\lVert w \rVert = \mathcal{O}(s^{-1}d^{-3})$	$\lambda = \mathcal{O}(s^{-1}d^{-3})$ $\lVert w \rVert = \mathcal{O}(s^{-1}d^{-5})$

Examples

Discrete Fourier	*Continuous Fourier*	*Laplace transform*
Random features		
$\varphi_\omega(x) = e^{\mathrm{i}2\pi\omega^\top x}$ $\Lambda \propto \prod_{j=1}^d g_j(\omega_j)$	$\varphi_\omega(x) = e^{\mathrm{i}\omega^\top x}$ $\Lambda = \mathcal{N}(0, \Sigma)$	$\varphi_\omega(x) = \prod_{i=1}^d \sqrt{\frac{2(x_i+\alpha_i)}{\alpha_i}} e^{-\omega^\top x}$ $\Lambda \propto e^{-\alpha^\top \omega}$
No. samples (up to $\log$ **factors),** $p = 1$ **for random signs,** $p = 3/2$ **in general**		
Rand. sgn.: $\mathcal{O}(sd^3)$ General: $\mathcal{O}(s^{3/2}d^3)$	Rand. sgn.: $\mathcal{O}(sd^3)$ General: $\mathcal{O}(s^{3/2}d^3)$	Rand. sgn.: $\mathcal{O}(sd^7)$ General: $\mathcal{O}(s^{3/2}d^7)$
Stability regions		
$\lambda = \mathcal{O}(s^{-1}d^{-2})$ $\\|w\\| = \mathcal{O}(s^{-1}d^{-3})$	$\lambda = \mathcal{O}(s^{-1}d^{-2})$ $\\|w\\| = \mathcal{O}(s^{-1}d^{-3})$	$\lambda = \mathcal{O}(s^{-1}d^{-3})$ $\\|w\\| = \mathcal{O}(s^{-1}d^{-5})$

- Linear in sparsity when we have random signs.
- Improvement from s^2 to $s^{3/2}$ in the arbitrary signs case.
- Dependency on d is still in progress.

Gaussian mixture estimation (1D)

Task: Suppose we have data $\{t_1, \ldots, t_n\}$ drawn from

$$\xi = \sum_{j=1}^{s} a_j \mathcal{N}(\mathfrak{m}_j, \mathfrak{s}_j^2), \quad \text{where} \quad a_j > 0 \quad \text{and} \quad \sum_j a_j = 1$$

Find $a_j \in \mathbb{R}_+$, $x_j \stackrel{\text{def.}}{=} (\mathfrak{m}_j, \mathfrak{s}_j) \in \mathbb{R} \times \mathbb{R}_+$, $j = 1, \ldots, s$.

Gaussian mixture estimation (1D)

Task: Suppose we have data $\{t_1, \ldots, t_n\}$ drawn from

$$\xi = \sum_{j=1}^{s} a_j \mathcal{N}(\mathfrak{m}_j, \mathfrak{s}_j^2), \quad \text{where} \quad a_j > 0 \quad \text{and} \quad \sum_j a_j = 1$$

Find $a_j \in \mathbb{R}_+$, $x_j \stackrel{\text{def.}}{=} (\mathfrak{m}_j, \mathfrak{s}_j) \in \mathbb{R} \times \mathbb{R}_+$, $j = 1, \ldots, s$.

Observe: $y \in \mathbb{C}^m$ of m sketches against sketching functions $\theta_\omega(t)$:

$$y_k \stackrel{\text{def.}}{=} \frac{1}{n} \sum_{j=1}^{n} \theta_{\omega_k}(t_j) \approx \int_{\mathcal{T}} \theta_{\omega_k}(t) \xi(t) \mathrm{d}t = \int_{\mathcal{X}} \int_{\mathcal{T}} \theta_{\omega_k}(t) \xi_x(t) \mathrm{d}t \mathrm{d}\mu_0(x),$$

where $\xi_x = \mathcal{N}(\mathfrak{m}, \mathfrak{s}^2)$.

i.e. our sparse spikes problem with $\mu_0 \stackrel{\text{def.}}{=} \sum_{i=1}^{s} a_i \delta_{(\mathfrak{m}_i, \mathfrak{s}_i)}$ and $\varphi_\omega(x) \stackrel{\text{def.}}{=} \int_{\mathcal{T}} \theta_\omega(t) \xi_x(t) \mathrm{d}t$.

Gaussian mixture estimation (1D)

Task: Suppose we have data $\{t_1, \ldots, t_n\}$ drawn from

$$\xi = \sum_{j=1}^{s} a_j \mathcal{N}(\mathfrak{m}_j, \mathfrak{s}_j^2), \quad \text{where} \quad a_j > 0 \quad \text{and} \quad \sum_j a_j = 1$$

Find $a_j \in \mathbb{R}_+$, $x_j \stackrel{\text{def.}}{=} (\mathfrak{m}_j, \mathfrak{s}_j) \in \mathbb{R} \times \mathbb{R}_+$, $j = 1, \ldots, s$.

Observe: $y \in \mathbb{C}^m$ of m sketches against sketching functions $\theta_\omega(t)$:

$$y_k \stackrel{\text{def.}}{=} \frac{1}{n} \sum_{j=1}^{n} \theta_{\omega_k}(t_j) \approx \int_{\mathcal{T}} \theta_{\omega_k}(t) \xi(t) \mathrm{d}t = \int_{\mathcal{X}} \int_{\mathcal{T}} \theta_{\omega_k}(t) \xi_x(t) \mathrm{d}t \mathrm{d}\mu_0(x),$$

where $\xi_x = \mathcal{N}(\mathfrak{m}, \mathfrak{s}^2)$.

i.e. our sparse spikes problem with $\mu_0 \stackrel{\text{def.}}{=} \sum_{i=1}^{s} a_i \delta_{(\mathfrak{m}_i, \mathfrak{s}_i)}$ and $\varphi_\omega(x) \stackrel{\text{def.}}{=} \int_{\mathcal{T}} \theta_\omega(t) \xi_x(t) \mathrm{d}t$.

Fourier sketching

Suppose that $\theta_{\omega_k}(t) = \exp(-\mathrm{i}\omega_k t)$, where $\omega_k \sim \mathcal{N}(0, \sigma^2)$. Then,

- Random features: $\varphi_\omega(\mathfrak{m}, \mathfrak{s}) = \sqrt[4]{2\mathfrak{s}^2\sigma^2 + 1} \exp\left(-\mathrm{i}\mathfrak{m}\omega - \frac{(\mathfrak{s}\omega)^2}{2}\right)$
- Noise: $\|w\|_2 = \mathcal{O}\left(\sqrt{\frac{\log(\rho^{-1})}{n}}\right)$ w.p. $1 - \rho$.

Support stability for Gaussian mixture estimation (1D)

Kernel	$K((\mathfrak{m},\mathfrak{s}),(\mathfrak{n},\mathfrak{t})) = \sqrt{\frac{2\mathfrak{s}_\sigma \mathfrak{t}_\sigma}{\mathfrak{s}_\sigma^2+\mathfrak{t}_\sigma^2}} \exp\left(-\frac{(\mathfrak{m}-\mathfrak{n})^2}{2(\mathfrak{s}_\sigma^2+\mathfrak{t}_\sigma^2)}\right)$ where $\mathfrak{s}_\sigma^2 = \frac{1}{2\sigma^2} + \mathfrak{s}^2$
Metric and separation	$\mathbf{H}_{(\mathfrak{m},\mathfrak{s})} = \begin{pmatrix} 1/(2\mathfrak{s}_\sigma^2) & 0 \\ 0 & 1/(2\mathfrak{s}_\sigma^2) \end{pmatrix}$ $d_{\mathbf{H}}((\mathfrak{m},\mathfrak{s}),(\mathfrak{n},\mathfrak{t})) = 2\mathrm{arcsinh}\left(\frac{1}{2}\sqrt{\frac{(\mathfrak{m}-\mathfrak{n})^2+(\mathfrak{s}_\sigma-\mathfrak{t}_\sigma)^2}{\mathfrak{s}\mathfrak{t}}}\right)$ $\Delta = \mathcal{O}(\log(s_{\max}))$.
No. samples	Suppose $\mathcal{X} \subset \mathbb{R}\times(0,A]$ and $\sigma \propto \frac{1}{A\sqrt{\log(m/\rho)+1}}$. $m = \mathcal{O}(s^{3/2})$ (up to log factors)
Stability region	$\lambda = \mathcal{O}(\min \lvert a_i\rvert / (\sqrt{s}\,\lVert a\rVert_2))$, $n = \mathcal{O}(s^2/\min_i \lvert a_i\rvert^2)$

Support stability for Gaussian mixture estimation (1D)

Kernel	$K((\mathfrak{m},\mathfrak{s}),(\mathfrak{n},\mathfrak{t})) = \sqrt{\frac{2\mathfrak{s}_\sigma \mathfrak{t}_\sigma}{\mathfrak{s}_\sigma^2+\mathfrak{t}_\sigma^2}} \exp\left(-\frac{(\mathfrak{m}-\mathfrak{n})^2}{2(\mathfrak{s}_\sigma^2+\mathfrak{t}_\sigma^2)}\right)$ where $\mathfrak{s}_\sigma^2 = \frac{1}{2\sigma^2}+\mathfrak{s}^2$
Metric and separation	$\mathbf{H}_{(\mathfrak{m},\mathfrak{s})} = \begin{pmatrix} 1/(2\mathfrak{s}_\sigma^2) & 0 \\ 0 & 1/(2\mathfrak{s}_\sigma^2) \end{pmatrix}$ $d_{\mathbf{H}}((\mathfrak{m},\mathfrak{s}),(\mathfrak{n},\mathfrak{t})) = 2\mathrm{arcsinh}\left(\frac{1}{2}\sqrt{\frac{(\mathfrak{m}-\mathfrak{n})^2+(\mathfrak{s}_\sigma-\mathfrak{t}_\sigma)^2}{\mathfrak{s}\mathfrak{t}}}\right)$ $\Delta = \mathcal{O}(\log(s_{\max}))$.
No. samples	Suppose $\mathcal{X} \subset \mathbb{R}\times(0,A]$ and $\sigma \propto \frac{1}{A\sqrt{\log(m/\rho)+1}}$. $m = \mathcal{O}(s^{3/2})$ (up to log factors)
Stability region	$\lambda = \mathcal{O}(\min\|a_i\|/(\sqrt{s}\,\\|a\\|_2))$, $n = \mathcal{O}(s^2/\min_i\|a_i\|^2)$

- No closed form expression for $d_{\mathbf{H}}$ in higher dimensions.
- If $\mu_0 = \sum_i a_i \mathcal{N}(x_{0,i},\Sigma)$ and $\Sigma \in \mathbb{R}^{d\times d}$ is known, then $\omega_k \sim \mathcal{N}(0,\Sigma^{-1}/d)$ implies the associated kernel is $\exp\left(-\|x-x'\|_{\Sigma^{-1}/(2+d)}\right)$, support stability guaranteed if
  - $\|x_j - x_\ell\|_{\Sigma^{-1}} \gtrsim \sqrt{d\log(s)}$
  - $m = \mathcal{O}(s^{3/2}d^3)$, $n = \mathcal{O}(s^2 d^6/\min_i|a_i|^2)$ and $\lambda = \mathcal{O}(\min|a_i|/(\sqrt{s}d^2\|a\|_2))$.

Outline

Fenchel Duals

The Fenchel dual of $\mathcal{P}_\lambda(y)$ is

$$\sup_{p \in \mathbb{C}^m, \|\Phi^* p\|_\infty \leqslant 1} \mathrm{Re}\langle p,\, y\rangle - \lambda \|p\|_2^2 \tag{4.1}$$

Note that for $\lambda > 0$, there is a unique dual solution p_λ, since this is equivalent to $\min_{\|\Phi^* p\|_\infty \leqslant 1} \|p - y/\lambda\|$ which a projection of y/λ onto a closed convex set.

Fenchel Duals

The Fenchel dual of $\mathcal{P}_\lambda(y)$ is

$$\sup_{p\in\mathbb{C}^m, \|\Phi^* p\|_\infty \leqslant 1} \mathrm{Re}\langle p,\, y\rangle - \lambda \|p\|_2^2 \tag{4.1}$$

Note that for $\lambda > 0$, there is a unique dual solution p_λ, since this is equivalent to $\min_{\|\Phi^* p\|_\infty \leqslant 1} \|p - y/\lambda\|$ which a projection of y/λ onto a closed convex set.

Primal dual relations: The dual solution p_λ is related to any primal solution μ_λ by

$$\Phi^* p_\lambda \in \partial\, |\mu_\lambda| \quad \text{and} \quad p_\lambda = \frac{1}{\lambda}(y - \Phi\mu_\lambda)$$

and for $\lambda = 0$, $\Phi^* p_0 \in \partial\, |\mu_0|$ and $y = \Phi\mu_0$.

Fenchel Duals

The Fenchel dual of $\mathcal{P}_\lambda(y)$ is

$$\sup_{p\in\mathbb{C}^m, \|\Phi^* p\|_\infty \leqslant 1} \mathrm{Re}\langle p,\, y\rangle - \lambda \|p\|_2^2 \tag{4.1}$$

Note that for $\lambda > 0$, there is a unique dual solution p_λ, since this is equivalent to $\min_{\|\Phi^* p\|_\infty \leqslant 1} \|p - y/\lambda\|$ which a projection of y/λ onto a closed convex set.

Primal dual relations: The dual solution p_λ is related to any primal solution μ_λ by

$$\Phi^* p_\lambda \in \partial\,|\mu_\lambda| \quad \text{and} \quad p_\lambda = \frac{1}{\lambda}(y - \Phi\mu_\lambda)$$

and for $\lambda = 0$, $\Phi^* p_0 \in \partial\,|\mu_0|$ and $y = \Phi\mu_0$.

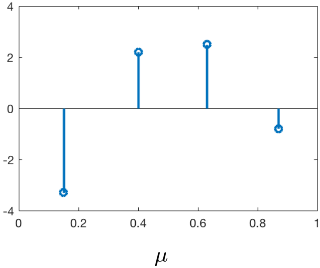
We have $\partial\,|\mu| = \left\{ f \in \mathcal{C}(\mathcal{X})\ ;\ \|f\|_\infty \leqslant 1,\ \langle f,\, \mu\rangle = |\mu|\,(\mathcal{X}) \right\}$, and

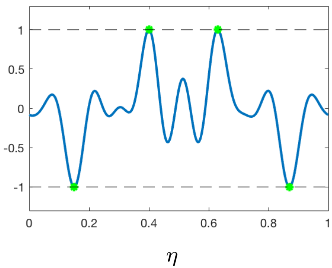
$$\mathrm{Supp}(\mu_\lambda) \subseteq \{x\ ;\ |\eta_\lambda(x)| = 1\}\,, \quad \text{where} \quad \eta_\lambda = \Phi^* p_\lambda.$$

η_λ are often called dual certificates.

Dual certificate guarantees for sparse measures

Let $\mu_0 = \sum_j a_j \delta_{x_j}$. Then $\partial |\mu_0| = \left\{ f \in \mathcal{C}(\mathcal{X}) \; ; \; \|f\|_\infty \leqslant 1, \; f(x_j) = \mathrm{sign}(a_j) \right\}$



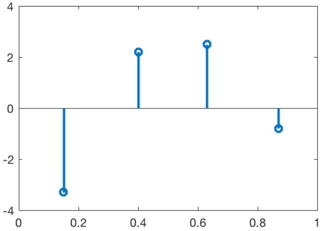


Uniqueness: μ_0 is the unique solution if

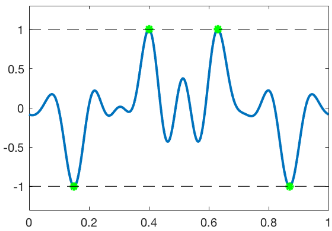
- $\exists \eta$ such that $\eta(x_j) = \mathrm{sign}(a_j)$, $|\eta(x)| < 1$ for all $x \notin X$
- $\Phi_X : b \in \mathbb{C}^s \to \sum_j b_j \varphi(x_j)$ is injective.

Dual certificate guarantees for sparse measures

Let $\mu_0 = \sum_j a_j \delta_{x_j}$. Then $\partial |\mu_0| = \left\{ f \in \mathcal{C}(\mathcal{X}) \; ; \; \|f\|_\infty \leqslant 1, \; f(x_j) = \mathrm{sign}(a_j) \right\}$



μ



η

Uniqueness: μ_0 is the unique solution if

- $\exists \eta$ such that $\eta(x_j) = \mathrm{sign}(a_j)$, $|\eta(x)| < 1$ for all $x \notin X$
- $\Phi_X : b \in \mathbb{C}^s \to \sum_j b_j \varphi(x_j)$ is injective.

Stability is guaranteed if η is nondegenerate:

$$\forall j \;\; \mathrm{sign}(a_j) \nabla^2 \eta(x_j) \prec 0 \quad \text{and} \quad \forall x \notin \{x_j\}_{j=1}^s, \; |\eta(x)| < 1$$

Stability

Clustering stability [Candès & Fernandez-Granda '14 and Azäis, De Castro & Gamboa '13]
Suppose η is nondegenerate with $\varepsilon_0, \varepsilon_2 > 0$, $\mathcal{X}_j^{\text{near}} \ni x_j$ such that

- $|\eta(x)| \leqslant 1 - \varepsilon_0$ for all $x \in \mathcal{X}^{\text{far}}$ where $\mathcal{X}^{\text{far}} \overset{\text{def.}}{=} \mathcal{X} \setminus \bigcup_{j=1}^{s} \mathcal{X}_j^{\text{near}}$,
- $\forall i$, $\forall x \in \mathcal{X}_i^{\text{near}}$, $|\eta(x)| \leqslant 1 - \varepsilon_2 d_{\mathbf{H}}(x, x_i)^2$.

Then, for $\lambda \sim \delta / \|p\|$,

$$\varepsilon_0 \left|\hat{\mu}\right| (\mathcal{X}^{\text{far}}) + \varepsilon_2 \sum_{j=1}^{s} \int_{\mathcal{X}_i^{\text{near}}} d_{\mathbf{H}}(x, x_i)^2 \mathrm{d} \left|\hat{\mu}\right|(x) \lesssim \delta(1 + \|p\|).$$

Stability

Clustering stability [Candès & Fernandez-Granda '14 and Azäis, De Castro & Gamboa '13]
Suppose η is nondegenerate with $\varepsilon_0, \varepsilon_2 > 0$, $\mathcal{X}_j^{\text{near}} \ni x_j$ such that

- $|\eta(x)| \leqslant 1 - \varepsilon_0$ for all $x \in \mathcal{X}^{\text{far}}$ where $\mathcal{X}^{\text{far}} \overset{\text{def.}}{=} \mathcal{X} \setminus \bigcup_{j=1}^{s} \mathcal{X}_j^{\text{near}}$,
- $\forall i$, $\forall x \in \mathcal{X}_i^{\text{near}}$, $|\eta(x)| \leqslant 1 - \varepsilon_2 d_{\mathbf{H}}(x, x_i)^2$.

Then, for $\lambda \sim \delta / \|p\|$, defining $P_X(|\hat{\mu}|) \overset{\text{def.}}{=} \sum_{j=1}^{s} |\hat{\mu}| (\mathcal{X}_j^{\text{near}}) \delta_{x_j}$, we have

$$\mathcal{T}_{\mathbf{H}}^2(|\hat{\mu}|, P_X(|\hat{\mu}|)) \lesssim \frac{\delta \|p\|}{\min\{\varepsilon_0, \varepsilon_2\}}.$$

where $\mathcal{T}_{\mathbf{H}}^2 \overset{\text{def.}}{=} \inf_{\mu,\nu} W_{\mathbf{H}}^2(\hat{\mu}, \hat{\nu}) + |\mu - \hat{\mu}| (\mathcal{X}) + |\nu - \hat{\nu}| (\mathcal{X})$.

Stability

Clustering stability [Candès & Fernandez-Granda '14 and Azäis, De Castro & Gamboa '13]
Suppose η is nondegenerate with $\varepsilon_0, \varepsilon_2 > 0$, $\mathcal{X}_j^{\text{near}} \ni x_j$ such that

- $|\eta(x)| \leqslant 1 - \varepsilon_0$ for all $x \in \mathcal{X}^{\text{far}}$ where $\mathcal{X}^{\text{far}} \stackrel{\text{def.}}{=} \mathcal{X} \setminus \bigcup_{j=1}^s \mathcal{X}_j^{\text{near}}$,
- $\forall i, \forall x \in \mathcal{X}_i^{\text{near}}, |\eta(x)| \leqslant 1 - \varepsilon_2 d_{\mathbf{H}}(x, x_i)^2$.

Then, for $\lambda \sim \delta / \|p\|$, defining $P_X(|\hat{\mu}|) \stackrel{\text{def.}}{=} \sum_{j=1}^s |\hat{\mu}| (\mathcal{X}_j^{\text{near}}) \delta_{x_j}$, we have

$$\mathcal{T}_{\mathbf{H}}^2(|\hat{\mu}|, P_X(|\hat{\mu}|)) \lesssim \frac{\delta \|p\|}{\min\{\varepsilon_0, \varepsilon_2\}}.$$

where $\mathcal{T}_{\mathbf{H}}^2 \stackrel{\text{def.}}{=} \inf_{\mu,\nu} W_{\mathbf{H}}^2(\hat{\mu}, \hat{\nu}) + |\mu - \hat{\mu}| (\mathcal{X}) + |\nu - \hat{\nu}| (\mathcal{X})$.

Support stability [Duval & Peyré '15] We have $p_\lambda \to p_0$ where

$$p_0 \stackrel{\text{def.}}{=} \operatorname{argmin} \{\|p\| \; ; \; \Phi^* p \in \operatorname{argmax}(\mathcal{D}_0(y))\}$$

If the **minimal norm certificate** $\eta_0 \stackrel{\text{def.}}{=} \Phi^* p_0$ is nondegenerate and μ_0 is identifiable, then for λ and $\frac{\|w\|}{\lambda}$ sufficiently small, $\mathcal{P}_\lambda(\Phi\mu_0 + w)$ has unique solution $\mu_{\lambda,w}$ which consists of exactly s spikes and the recovered positions and amplitudes follow a $\mathcal{C}^1$ path as λ and w converge to 0.

What is natural candidate for a nondegenerate solution to $\mathcal{D}_0(y)$?

What is natural candidate for a nondegenerate solution to $\mathcal{D}_0(y)$?

In CS, for $\Phi \in \mathbb{C}^{m\times N}$, we need to find $v \in \mathrm{Im}(\Phi^*)$ such that $|v_j| < 1$ for $j \notin T$ and $v_j = \mathrm{sign}(a_j)$ for $j \in T$.

What is natural candidate for a nondegenerate solution to $\mathcal{D}_0(y)$?

In CS, for $\Phi \in \mathbb{C}^{m\times N}$, we need to find $v \in \mathrm{Im}(\Phi^*)$ such that $|v_j| < 1$ for $j \notin T$ and $v_j = \mathrm{sign}(a_j)$ for $j \in T$.
In the case $\mathbb{E}[\Phi^*\Phi] = \mathrm{Id}$, the Fuchs certificate is an appropriate choice:

$$v = \Phi^*\Phi_T(\Phi_T^*\Phi_T)^{-1}\,\mathrm{sign}(a_T).$$

By definition, $v_T = \mathrm{sign}(a_T)$.

What is natural candidate for a nondegenerate solution to $\mathcal{D}_0(y)$?

In CS, for $\Phi \in \mathbb{C}^{m\times N}$, we need to find $v \in \mathrm{Im}(\Phi^*)$ such that $|v_j| < 1$ for $j \notin T$ and $v_j = \mathrm{sign}(a_j)$ for $j \in T$.
In the case $\mathbb{E}[\Phi^*\Phi] = \mathrm{Id}$, the Fuchs certificate is an appropriate choice:

$$v = \Phi^*\Phi_T(\Phi_T^*\Phi_T)^{-1}\,\mathrm{sign}(a_T).$$

By definition, $v_T = \mathrm{sign}(a_T)$.

Vanishing derivatives precertificate [Duval & Peyré '15]

In our case, for $\alpha \in \mathbb{C}^s$ and $\beta \in \mathbb{C}^{sd}$, define $\Gamma_X : \mathbb{C}^{s(d+1)} \to \mathbb{C}^m$ by

$$\Gamma_X \binom{\alpha}{\beta} = \Phi_X\alpha + \Phi_X^{(1)}\beta, \quad \text{where} \quad \Phi_X\alpha = \sum_j \alpha_j\varphi(x_j),\ \Phi_X^{(1)}\beta = \sum_{j=1}^s \beta_j{}^\top\nabla\varphi(x_j)$$

Consider

$$\eta_V = \Phi^*\Gamma_X(\Gamma_X^*\Gamma_X)^{-1}\binom{\mathrm{sign}(a)}{0_{sd}}$$

What is natural candidate for a nondegenerate solution to $\mathcal{D}_0(y)$?

In CS, for $\Phi \in \mathbb{C}^{m\times N}$, we need to find $v \in \mathrm{Im}(\Phi^*)$ such that $|v_j| < 1$ for $j \notin T$ and $v_j = \mathrm{sign}(a_j)$ for $j \in T$.
In the case $\mathbb{E}[\Phi^*\Phi] = \mathrm{Id}$, the Fuchs certificate is an appropriate choice:

$$v = \Phi^*\Phi_T(\Phi_T^*\Phi_T)^{-1}\,\mathrm{sign}(a_T).$$

By definition, $v_T = \mathrm{sign}(a_T)$.

Vanishing derivatives precertificate [Duval & Peyré '15]

In our case, for $\alpha \in \mathbb{C}^s$ and $\beta \in \mathbb{C}^{sd}$, define $\Gamma_X : \mathbb{C}^{s(d+1)} \to \mathbb{C}^m$ by

$$\Gamma_X\binom{\alpha}{\beta} = \Phi_X\alpha + \Phi_X^{(1)}\beta, \quad \text{where} \quad \Phi_X\alpha = \sum_j \alpha_j\varphi(x_j),\ \Phi_X^{(1)}\beta = \sum_{j=1}^s \beta_j{}^\top\nabla\varphi(x_j)$$

Consider

$$\eta_V = \Phi^*\Gamma_X(\Gamma_X^*\Gamma_X)^{-1}\binom{\mathrm{sign}(a)}{0_{sd}}$$

- by definition $\eta_V(x_j) = \mathrm{sign}(a_j)$ and $\nabla\eta_V(x_j) = 0$.

What is natural candidate for a nondegenerate solution to $\mathcal{D}_0(y)$?

In CS, for $\Phi \in \mathbb{C}^{m\times N}$, we need to find $v \in \mathrm{Im}(\Phi^*)$ such that $|v_j| < 1$ for $j \notin T$ and $v_j = \mathrm{sign}(a_j)$ for $j \in T$.
In the case $\mathbb{E}[\Phi^*\Phi] = \mathrm{Id}$, the Fuchs certificate is an appropriate choice:

$$v = \Phi^*\Phi_T(\Phi_T^*\Phi_T)^{-1}\,\mathrm{sign}(a_T).$$

By definition, $v_T = \mathrm{sign}(a_T)$.

Vanishing derivatives precertificate [Duval & Peyré '15]

In our case, for $\alpha \in \mathbb{C}^s$ and $\beta \in \mathbb{C}^{sd}$, define $\Gamma_X : \mathbb{C}^{s(d+1)} \to \mathbb{C}^m$ by

$$\Gamma_X \binom{\alpha}{\beta} = \Phi_X\alpha + \Phi_X^{(1)}\beta, \quad \text{where} \quad \Phi_X\alpha = \sum_j \alpha_j\varphi(x_j),\ \Phi_X^{(1)}\beta = \sum_{j=1}^s \beta_j{}^\top\nabla\varphi(x_j)$$

Consider

$$\eta_V = \Phi^*\Gamma_X(\Gamma_X^*\Gamma_X)^{-1}\binom{\mathrm{sign}(a)}{0_{sd}}$$

- by definition $\eta_V(x_j) = \mathrm{sign}(a_j)$ and $\nabla\eta_V(x_j) = 0$.
- In fact, $\eta_V = \Phi^* p_V$ where

$$p_V = \mathrm{argmin}\left\{\|p\|_2\ ;\ (\Phi^*p)(x_j) = \mathrm{sign}(a_j),\ \nabla(\Phi^*p)(x_j) = 0\right\}.$$

What is natural candidate for a nondegenerate solution to $\mathcal{D}_0(y)$?

In CS, for $\Phi \in \mathbb{C}^{m\times N}$, we need to find $v \in \mathrm{Im}(\Phi^*)$ such that $|v_j| < 1$ for $j \notin T$ and $v_j = \mathrm{sign}(a_j)$ for $j \in T$.
In the case $\mathbb{E}[\Phi^*\Phi] = \mathrm{Id}$, the Fuchs certificate is an appropriate choice:

$$v = \Phi^*\Phi_T(\Phi_T^*\Phi_T)^{-1}\,\mathrm{sign}(a_T).$$

By definition, $v_T = \mathrm{sign}(a_T)$.

Vanishing derivatives precertificate [Duval & Peyré '15]

In our case, for $\alpha \in \mathbb{C}^s$ and $\beta \in \mathbb{C}^{sd}$, define $\Gamma_X : \mathbb{C}^{s(d+1)} \to \mathbb{C}^m$ by

$$\Gamma_X\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \Phi_X\alpha + \Phi_X^{(1)}\beta, \quad \text{where} \quad \Phi_X\alpha = \sum_j \alpha_j\varphi(x_j),\ \Phi_X^{(1)}\beta = \sum_{j=1}^s {\beta_j}^\top\nabla\varphi(x_j)$$

Consider

$$\eta_V = \Phi^*\Gamma_X(\Gamma_X^*\Gamma_X)^{-1}\begin{pmatrix}\mathrm{sign}(a)\\ 0_{sd}\end{pmatrix}$$

- by definition $\eta_V(x_j) = \mathrm{sign}(a_j)$ and $\nabla\eta_V(x_j) = 0$.
- In fact, $\eta_V = \Phi^* p_V$ where
$$p_V = \mathrm{argmin}\left\{\|p\|_2\ ;\ (\Phi^*p)(x_j) = \mathrm{sign}(a_j),\ \nabla(\Phi^*p)(x_j) = 0\right\}.$$
- If $\|\eta_V\|_\infty \leqslant 1$, then we have $\eta_V = \eta_0$, and nondegeneracy guarantees support stability.

Key ideas of proof

We can also write

$$\eta_V(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x) + \sum_{i=1}^{N} \beta_i K^{(10)}(x_i, x), \qquad \binom{\alpha}{\beta} = D_{K,X}^{-1} \binom{\operatorname{sign}(a)}{0_N}$$

with covariance kernel $K(x, x') = \langle \varphi(x), \varphi(x') \rangle$, $D_{K,X} \overset{\text{def.}}{=} \begin{pmatrix} M_0, & M_1 \\ M_1^T & M_2 \end{pmatrix}$,

$$\text{where} \quad M_0 = (K(x_i, x_j))_{i,j}, \quad M_1 = (K^{(01)}(x_i, x_j))_{i,j}, \quad M_2 = (K^{(11)}(x_i, x_j))_{i,j}.$$

Key ideas of proof

We can also write

$$\eta_V(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x) + \sum_{i=1}^{N} \beta_i K^{(10)}(x_i, x), \qquad \binom{\alpha}{\beta} = D_{K,X}^{-1} \binom{\mathrm{sign}(a)}{0_N}$$

with covariance kernel $K(x, x') = \langle \varphi(x), \varphi(x') \rangle$, $D_{K,X} \overset{\text{def.}}{=} \begin{pmatrix} M_0, & M_1 \\ M_1^T & M_2 \end{pmatrix}$,

$$\text{where} \quad M_0 = (K(x_i, x_j))_{i,j}, \quad M_1 = (K^{(01)}(x_i, x_j))_{i,j}, \quad M_2 = (K^{(11)}(x_i, x_j))_{i,j}.$$

The idea of the proof is

- η_V associated to the limit kernel $K = \mathbb{E}[\hat{K}]$ is nondegenerate.

Key ideas of proof

We can also write

$$\eta_V(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x) + \sum_{i=1}^{N} \beta_i K^{(10)}(x_i, x), \qquad \begin{pmatrix}\alpha\\ \beta\end{pmatrix} = D_{K,X}^{-1} \begin{pmatrix}\mathrm{sign}(a)\\ 0_N\end{pmatrix}$$

with covariance kernel $K(x, x') = \langle \varphi(x), \varphi(x') \rangle$, $D_{K,X} \stackrel{\text{def.}}{=} \begin{pmatrix} M_0, & M_1 \\ M_1^T & M_2 \end{pmatrix}$,

$$\text{where} \quad M_0 = (K(x_i, x_j))_{i,j}, \quad M_1 = (K^{(01)}(x_i, x_j))_{i,j}, \quad M_2 = (K^{(11)}(x_i, x_j))_{i,j}.$$

The idea of the proof is

- η_V associated to the limit kernel $K = \mathbb{E}[\hat{K}]$ is nondegenerate.
- We therefore simply need to show that $\hat{\eta}_V$ associated to $\hat{K}$ is close to η_V:
  - On $\mathcal{X}^{\text{far}}$ $\hat{\eta}_V \approx \eta_V$ is bounded away from 1 in absolute value.
  - On $\mathcal{X}_j^{\text{near}}$, $\mathrm{sign}(a_j)\nabla^2 \hat{\eta}_V \approx \mathrm{sign}(a_j)\nabla^2 \eta_V$ is negative definite.

Key ideas of proof

We can also write

$$\eta_V(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x) + \sum_{i=1}^{N} \beta_i K^{(10)}(x_i, x), \qquad \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = D_{K,X}^{-1} \begin{pmatrix} \mathrm{sign}(a) \\ 0_N \end{pmatrix}$$

with covariance kernel $K(x, x') = \langle \varphi(x), \varphi(x') \rangle$, $D_{K,X} \overset{\text{def.}}{=} \begin{pmatrix} M_0, & M_1 \\ M_1^T & M_2 \end{pmatrix}$,

$$\text{where} \quad M_0 = (K(x_i, x_j))_{i,j}, \quad M_1 = (K^{(01)}(x_i, x_j))_{i,j}, \quad M_2 = (K^{(11)}(x_i, x_j))_{i,j}.$$

The idea of the proof is

- η_V associated to the limit kernel $K = \mathbb{E}[\hat{K}]$ is nondegenerate.
- We therefore simply need to show that $\hat{\eta}_V$ associated to $\hat{K}$ is close to η_V:
  - On $\mathcal{X}^{\mathrm{far}}$ $\hat{\eta}_V \approx \eta_V$ is bounded away from 1 in absolute value.
  - On $\mathcal{X}_j^{\mathrm{near}}$, $\mathrm{sign}(a_j)\nabla^2 \hat{\eta}_V \approx \mathrm{sign}(a_j)\nabla^2 \eta_V$ is negative definite.
- Our proof still requires random signs and is a direct extension of the work of Tang et al (to the higher dimensional and general operator setting), key difference is incorporation of the Fisher metric.

Comment on our $s^{1.5}$ bound

To explain the random signs requirement, consider the Fuchs certificate in the finite dimensional case,

$$v = \Phi^* \Phi_T (\Phi_T^* \Phi_T)^{-1} \operatorname{sign}(a_T) = (\langle \operatorname{sign}(a_T),\, u_j \rangle)_{j=1}^N$$

where $u_j = (\Phi_T^* \Phi_T)^{-1} \Phi_T^* \Phi_{\{j\}}$, and we need to show $|v_j| < 1$ for $j \notin T$:

Comment on our $s^{1.5}$ bound

To explain the random signs requirement, consider the Fuchs certificate in the finite dimensional case,

$$v = \Phi^* \Phi_T (\Phi_T^* \Phi_T)^{-1} \operatorname{sign}(a_T) = (\langle \operatorname{sign}(a_T), u_j \rangle)_{j=1}^N$$

where $u_j = (\Phi_T^* \Phi_T)^{-1} \Phi_T^* \Phi_{\{j\}}$, and we need to show $|v_j| < 1$ for $j \notin T$:

- Naively, $|v_j| \leqslant \sqrt{s} \left\| u_j \right\|_2$, so we need for $j \notin T$, $\left\| \Phi_T^* \Phi_{\{j\}} \right\|_2 \lesssim \frac{1}{\sqrt{s}}$ which forces $m \gtrsim s^2$ measurements.

Comment on our $s^{1.5}$ bound

To explain the random signs requirement, consider the Fuchs certificate in the finite dimensional case,

$$v = \Phi^* \Phi_T (\Phi_T^* \Phi_T)^{-1} \operatorname{sign}(a_T) = (\langle \operatorname{sign}(a_T), u_j \rangle)_{j=1}^N$$

where $u_j = (\Phi_T^* \Phi_T)^{-1} \Phi_T^* \Phi_{\{j\}}$, and we need to show $|v_j| < 1$ for $j \notin T$:

- Naively, $|v_j| \leqslant \sqrt{s} \, \|u_j\|_2$, so we need for $j \notin T$, $\left\| \Phi_T^* \Phi_{\{j\}} \right\|_2 \lesssim \frac{1}{\sqrt{s}}$ which forces $m \gtrsim s^2$ measurements.
- If $\operatorname{sign}(a_T)$ is made of random signs, then by Hoeffding's inequality, with high probability, $|v_j| = |\langle u_j, \operatorname{sign}(a_T) \rangle| \lesssim \|u_j\|_2$ which yields $m \gtrsim s$ (up to log).

Comment on our $s^{1.5}$ bound

To explain the random signs requirement, consider the Fuchs certificate in the finite dimensional case,

$$v = \Phi^* \Phi_T (\Phi_T^* \Phi_T)^{-1} \operatorname{sign}(a_T) = (\langle \operatorname{sign}(a_T), u_j \rangle)_{j=1}^N$$

where $u_j = (\Phi_T^* \Phi_T)^{-1} \Phi_T^* \Phi_{\{j\}}$, and we need to show $|v_j| < 1$ for $j \notin T$:

- Naively, $|v_j| \leqslant \sqrt{s} \, \|u_j\|_2$, so we need for $j \notin T$, $\left\| \Phi_T^* \Phi_{\{j\}} \right\|_2 \lesssim \frac{1}{\sqrt{s}}$ which forces $m \gtrsim s^2$ measurements.
- If $\operatorname{sign}(a_T)$ is made of random signs, then by Hoeffding's inequality, with high probability, $|v_j| = |\langle u_j, \operatorname{sign}(a_T) \rangle| \lesssim \|u_j\|_2$ which yields $m \gtrsim s$ (up to log).
- But we can also write

$$v_j = \langle ((\Phi_T^* \Phi_T)^{-1} - \mathrm{Id}) \Phi_T^* \Phi_{\{j\}}, \operatorname{sign}(a_T) \rangle + \langle \Phi_T^* \Phi_{\{j\}}, \operatorname{sign}(a_T) \rangle$$

  So, we simply need to ensure that $\left\| \Phi_T^* \Phi_T - \mathrm{Id}_T \right\|_{2 \to 2} \lesssim s^{-1/4}$ and $\left\| \Phi_T^* \Phi_{\{j\}} \right\|_2 \lesssim s^{-1/4}$ which is true w.h.p. when $m \gtrsim s^{1.5}$ (up to log factor).

Outline

Ideas from (finite dimensional) compressed sensing

Instead of requiring that $v_j = \mathrm{sign}(a_j)$, it is enough that this holds approximately.

Ideas from (finite dimensional) compressed sensing

Instead of requiring that $v_j = \mathrm{sign}(a_j)$, it is enough that this holds approximately.

Theorem (Gross (2011); Candès and Plan (2011))

Let T index the largest s entries of $|a|$. Suppose that there exists $v = \Phi^* p$ such that

$$\|v_T - \mathrm{sign}(a_T)\|_2 \leqslant \frac{1}{4} \quad \text{and} \quad \|v_{T^c}\|_\infty \leqslant \frac{1}{4}$$

and

$$\left\|(\Phi_T^* \Phi_T)^{-1}\right\|_{2\to 2} \leqslant 2 \quad \text{and} \quad \max_{i\in T^c} \left\|\Phi_T^* \Phi_{\{i\}}\right\|_2 \leqslant 1,$$

then one can guarantee that $\|\hat{a} - a\|_2 \lesssim \|p\|_2 \, \delta + \sigma_1(a)_s$ provided that $\lambda \sim \delta$.

Alternative proof: $\exists$ inexact certificate $\implies$ $\exists$ dual certificate

Theorem (Gross (2011); Candès and Plan (2011))

Let T index the largest s entries of $|a|$. Suppose that there exists $v = \Phi^* p$ such that

$$\|v_T - \mathrm{sign}(a_T)\|_2 \leqslant \frac{1}{4} \quad \textit{and} \quad \|v_{T^c}\|_\infty \leqslant \frac{1}{4}$$

and

$$\left\|(\Phi_T^* \Phi_T)^{-1}\right\|_{2\to 2} \leqslant 2 \quad \textit{and} \quad \max_{i\in T^c} \left\|\Phi_T^* \Phi_{\{i\}}\right\|_2 \leqslant 1,$$

then one can guarantee that $\|\hat{a} - a\|_2 \lesssim (1 + \|p\|_2)\delta + \sigma_1(x^0)_s$ provided that $\lambda \sim \delta / \|p\|$.

Proof:

1. Define $u \stackrel{\text{def.}}{=} v + \tilde{v}$ where $\tilde{v} \stackrel{\text{def.}}{=} \Phi^* \Phi_T (\Phi_T^* \Phi_T)^{-1} e$ and $e = \mathrm{sign}(a_T) - v_T$.
2. By definition, $u_T = v_T + e_T = \mathrm{sign}(a_T)$.
3. Note that

$$\|\tilde{v}_{T^c}\|_\infty \leqslant \|\Phi_{T^c}^* \Phi_T\|_{2\to\infty} \left\|(\Phi_T^* \Phi_T)^{-1}\right\|_{2\to 2} \|e\|_2 \leqslant \frac{1}{2},$$

so $\|u_{T^c}\|_\infty \leqslant \|v_{T^c}\|_\infty + \|\tilde{v}_{T^c}\|_\infty \leqslant \frac{3}{4}$.

Key steps of our proof

- Apply the golfing scheme [Gross '09, Candès & Plan '11] to construct $\tilde{\eta} \in \mathrm{Im}(\Phi^*)$ which is approximately nondegenerate on a finite grid:
  - The vector $V = (\tilde{\eta}(x_j), \mathrm{D}_1[\tilde{\eta}](x_j))_{j=1}^s$ satisfies $\left\| V - \binom{\mathrm{sign}(a)}{0_{sd}} \right\| \leqslant \delta$,
  - For all $x \in \mathcal{X}_{\mathrm{grid},j}^{\mathrm{near}}$, $\mathrm{sign}(a_j) \cdot \mathrm{D}_2[\tilde{\eta}](x) \prec -\varepsilon_2$.
  - For all $x \in \mathcal{X}_{\mathrm{grid}}^{\mathrm{far}}$, $|\tilde{\eta}(x)| < 1 - \varepsilon_0$.

Key steps of our proof

- Apply the golfing scheme [Gross '09, Candès & Plan '11] to construct $\tilde{\eta} \in \mathrm{Im}(\Phi^*)$ which is approximately nondegenerate on a finite grid:
  - The vector $V = (\tilde{\eta}(x_j), \mathrm{D}_1[\tilde{\eta}](x_j))_{j=1}^s$ satisfies $\left\| V - \binom{\mathrm{sign}(a)}{0_{sd}} \right\| \leqslant \delta$,
  - For all $x \in \mathcal{X}^{\mathrm{near}}_{\mathrm{grid},j}$, $\mathrm{sign}(a_j) \cdot \mathrm{D}_2[\tilde{\eta}](x) \prec -\varepsilon_2$.
  - For all $x \in \mathcal{X}^{\mathrm{far}}_{\mathrm{grid}}$, $|\tilde{\eta}(x)| < 1 - \varepsilon_0$.
- Show that provided that grid is sufficiently dense, this holds on the entire domain $\mathcal{X}$ (depends on L_1 and L_3).

Key steps of our proof

- Apply the golfing scheme [Gross '09, Candès & Plan '11] to construct $\tilde{\eta} \in \mathrm{Im}(\Phi^*)$ which is approximately nondegenerate on a finite grid:
  - The vector $V = (\tilde{\eta}(x_j), \mathrm{D}_1[\tilde{\eta}](x_j))_{j=1}^s$ satisfies $\left\| V - \binom{\mathrm{sign}(a)}{0_{sd}} \right\| \leqslant \delta$,
  - For all $x \in \mathcal{X}^{\mathrm{near}}_{\mathrm{grid},j}$, $\mathrm{sign}(a_j) \cdot \mathrm{D}_2[\tilde{\eta}](x) \prec -\varepsilon_2$.
  - For all $x \in \mathcal{X}^{\mathrm{far}}_{\mathrm{grid}}$, $|\tilde{\eta}(x)| < 1 - \varepsilon_0$.
- Show that provided that grid is sufficiently dense, this holds on the entire domain $\mathcal{X}$ (depends on L_1 and L_3).
- Add a small perturbation to $\tilde{\eta}$ to obtain a true certificate.

Key steps of our proof

- Apply the golfing scheme [Gross '09, Candès & Plan '11] to construct $\tilde{\eta} \in \mathrm{Im}(\Phi^*)$ which is approximately nondegenerate on a finite grid:
  - The vector $V = (\tilde{\eta}(x_j), \mathrm{D}_1[\tilde{\eta}](x_j))_{j=1}^s$ satisfies $\left\| V - \binom{\mathrm{sign}(a)}{0_{sd}} \right\| \leqslant \delta$,
  - For all $x \in \mathcal{X}^{\mathrm{near}}_{\mathrm{grid},j}$, $\mathrm{sign}(a_j) \cdot \mathrm{D}_2[\tilde{\eta}](x) \prec -\varepsilon_2$.
  - For all $x \in \mathcal{X}^{\mathrm{far}}_{\mathrm{grid}}$, $|\tilde{\eta}(x)| < 1 - \varepsilon_0$.
- Show that provided that grid is sufficiently dense, this holds on the entire domain $\mathcal{X}$ (depends on L_1 and L_3).
- Add a small perturbation to $\tilde{\eta}$ to obtain a true certificate.

We still construct a dual certificate, but it is *not* of minimal norm.

The subsampled setting

Assumption 1

- K is admissible, $\mu_0 = \sum_{j=1}^s a_j \delta_{x_j}$ with $\min_{j \neq k} d_{\mathbf{H}}(x_j, x_k) \geqslant \Delta$ and $s \leqslant s_{\max}$.
- $\mathcal{X}$ is a compact domain wth $R_{\mathcal{X}} \stackrel{\text{def.}}{=} \sup_{x, x' \in \mathcal{X}} d_{\mathbf{H}}(x, x')$,

Let $L_r(\omega) \stackrel{\text{def.}}{=} \sup_{x \in \mathcal{X}} \|\mathrm{D}_r[\varphi_\omega](x)\|$ and let F_r be such that $\mathbb{P}_\omega(L_r(\omega) > t) \leqslant F_r(t)$.

Assumption 2

For $\rho > 0$ (probability of failure) choose $m \in \mathbb{N}$ (number of measurements), and $\{\bar{L}_i\}_{i=0}^3$ such that

$$\sum_{j=0}^{3} F_j(\bar{L}_j) \leqslant \frac{\rho}{m} \quad \text{and} \quad \bar{L}_j^2 \sum_{i=0}^{3} F_i(\bar{L}_i) + 2 \int_{\bar{L}_j}^{\infty} t F_j(t) \mathrm{d}t \leqslant \frac{\varepsilon}{m}.$$

and $m \gtrsim C \cdot s \cdot (\log^2(s) + \log(N^d))$ where

$$N \stackrel{\text{def.}}{=} \frac{1}{\varepsilon} \mathbb{L}_3 R_{\mathcal{X}} d \sqrt{s} \quad \text{and} \quad C \stackrel{\text{def.}}{=} \frac{1}{\varepsilon^2} \left(\frac{\log\left(\frac{\mathbb{L}_2}{\varepsilon \rho}\right)}{\log(s)} + 1 \right) \left(\mathbb{L}_1^2 B + \mathbb{L}_2^2\right),$$

$B = B_{00} + B_{02} + B_{10} + B_{12}$, $\varepsilon = \min\{\varepsilon_0, \varepsilon_2\}, \mathbb{L}_r = \max_{i \leqslant r} \bar{L}_i$

Stability without the random signs assumption

Theorem

Let

$$\mathcal{X}_j^{near} \overset{\text{def.}}{=} \{x \in \mathcal{X} \; ; \; d_{\mathbf{H}}(x, x_j) \leqslant r_{\text{near}}\} \quad \textit{and} \quad \mathcal{X}^{far} \overset{\text{def.}}{=} \mathcal{X} \setminus \bigcup_{j=1}^{s} \mathcal{X}_j^{near}. \tag{5.1}$$

Suppose that $\|w\| \leqslant \delta$ *and* $\lambda \sim \delta/\sqrt{s}$ *(ignoring* log *factors), then any solution* $\hat{\mu}$ *to* $\mathcal{P}_\lambda(y)$ *is approximately* s*-sparse: by defining the "projection" of* $|\hat{\mu}|$ *onto* $X \overset{\text{def.}}{=} \{x_j\}$ *by* $P_X(|\hat{\mu}|) \overset{\text{def.}}{=} \sum_{j=1}^{s} |\hat{\mu}| (\mathcal{X}_j^{\text{near}})\delta_{x_j}$ *we have*

$$\mathcal{T}_{\mathbf{H}}^2(|\hat{\mu}| \, , P_X(|\hat{\mu}|)) \lesssim \frac{\delta\sqrt{s}}{\varepsilon}.$$

where $\mathcal{T}_{\mathbf{H}}^2 \overset{\text{def.}}{=} \inf_{\mu,\nu} W_{\mathbf{H}}^2(\hat{\mu}, \hat{\nu}) + |\mu - \hat{\mu}| (\mathcal{X}) + |\nu - \hat{\nu}| (\mathcal{X})$.

Stability without the random signs assumption

Theorem

Let

$$\mathcal{X}_j^{near} \overset{\text{def.}}{=} \{x \in \mathcal{X} \; ; \; d_{\mathbf{H}}(x, x_j) \leqslant r_{\text{near}}\} \quad \textit{and} \quad \mathcal{X}^{far} \overset{\text{def.}}{=} \mathcal{X} \setminus \bigcup_{j=1}^{s} \mathcal{X}_j^{near}. \tag{5.1}$$

Suppose that $\|w\| \leqslant \delta$ *and* $\lambda \sim \delta/\sqrt{s}$ *(ignoring* log *factors), then any solution* $\hat{\mu}$ *to* $\mathcal{P}_\lambda(y)$ *is approximately* s*-sparse: by defining the "projection" of* $|\hat{\mu}|$ *onto* $X \overset{\text{def.}}{=} \{x_j\}$ *by* $P_X(|\hat{\mu}|) \overset{\text{def.}}{=} \sum_{j=1}^{s} |\hat{\mu}| (\mathcal{X}_j^{\text{near}})\delta_{x_j}$ *we have*

$$\mathcal{T}_{\mathbf{H}}^2(|\hat{\mu}|, P_X(|\hat{\mu}|)) \lesssim \frac{\delta\sqrt{s}}{\varepsilon}.$$

where $\mathcal{T}_{\mathbf{H}}^2 \overset{\text{def.}}{=} \inf_{\mu,\nu} W_{\mathbf{H}}^2(\hat{\mu}, \hat{\nu}) + |\mu - \hat{\mu}| (\mathcal{X}) + |\nu - \hat{\nu}| (\mathcal{X})$.
Moreover, we have

$$\sum_j \left|a_j - \hat{\mu}(\mathcal{X}_j^{\text{near}})\right|^2 \lesssim \frac{\mathbb{L}_1}{\varepsilon}(1 + |\mu_0| (\mathcal{X})) \left(\delta\sqrt{s}\right)$$

Stability without the random signs assumption

Theorem

Let

$$\mathcal{X}_j^{near} \overset{\text{def.}}{=} \{x \in \mathcal{X} \,;\, d_{\mathbf{H}}(x, x_j) \leqslant r_{\text{near}}\} \quad \textit{and} \quad \mathcal{X}^{far} \overset{\text{def.}}{=} \mathcal{X} \setminus \bigcup_{j=1}^{s} \mathcal{X}_j^{near}. \tag{5.1}$$

Suppose $\mu_0 = \sum_{j=1}^s a_j \delta_{x_j} + \nu_0$ *where* $\nu_0 \perp \sum_j a_j \delta_{x_j}$.
Suppose that $\|w\| \leqslant \delta$ *and* $\lambda \sim \delta/\sqrt{s}$ *(ignoring* log *factors), then any solution* $\hat{\mu}$ *to* $\mathcal{P}_\lambda(y)$ *is approximately s-sparse: by defining the "projection" of* $|\hat{\mu}|$ *onto* $X \overset{\text{def.}}{=} \{x_j\}$ *by* $P_X(|\hat{\mu}|) \overset{\text{def.}}{=} \sum_{j=1}^s |\hat{\mu}| (\mathcal{X}_j^{\text{near}})\delta_{x_j}$ *we have*

$$\mathcal{T}_{\mathbf{H}}^2(|\hat{\mu}|\,, P_X(|\hat{\mu}|)) \lesssim \frac{\delta\sqrt{s} + |\nu_0|\,(\mathcal{X})}{\varepsilon}.$$

where $\mathcal{T}_{\mathbf{H}}^2 \overset{\text{def.}}{=} \inf_{\mu,\nu} W_{\mathbf{H}}^2(\hat{\mu}, \hat{\nu}) + |\mu - \hat{\mu}|\,(\mathcal{X}) + |\nu - \hat{\nu}|\,(\mathcal{X})$.

Summary

Papers:

- Support Localization and the Fisher Metric for off-the-grid Sparse Regularization, arXiv 1810.03340, AISTATS 2019.
- A Dual Certificates Analysis of Compressive Off-the-Grid Recovery, arXiv 1802.08464

Summary: Extended existing results to general measurement operators and the multivariate setting.

- Introduction of the Fisher metric, which offers a natural way of imposing the separation condition and allows a unified way of approaching nontranslational invariant problems.
- Quantitative support stability under a random signs assumption.
- Removal of the random signs condition (with support concentration guarantees).

Outlook:

- Algorithmic implications of the Fisher metric/natural gradient?
- Our results are optimal wrt s, but what about d?
- How should we quantify noise stability in general?

Summary

Papers:

- Support Localization and the Fisher Metric for off-the-grid Sparse Regularization, arXiv 1810.03340, AISTATS 2019.
- A Dual Certificates Analysis of Compressive Off-the-Grid Recovery, arXiv 1802.08464

Summary: Extended existing results to general measurement operators and the multivariate setting.

- Introduction of the Fisher metric, which offers a natural way of imposing the separation condition and allows a unified way of approaching nontranslational invariant problems.
- Quantitative support stability under a random signs assumption.
- Removal of the random signs condition (with support concentration guarantees).

Outlook:

- Algorithmic implications of the Fisher metric/natural gradient?
- Our results are optimal wrt s, but what about d?
- How should we quantify noise stability in general?

Thanks for listening!