SPARSE RECOVERY OVER NONLINEAR DICTIONARIES

Luiz F. O. Chamon, Yonina C. Eldar, and Alejandro Ribeiro

ICASSP 2019
May 14^{th}, 2019

Functional nonlinear dictionaries

- Dictionary has a continuum of atoms

$$\mathcal{D} = \{\boldsymbol{F}(\cdot, \beta) : \mathbb{R} \times \mathbb{R}^p \mid \beta \in \Omega \subset \mathbb{R}\}$$

Functional nonlinear dictionaries

- Dictionary has a continuum of atoms

$$\mathcal{D} = \{\boldsymbol{F}(\cdot, \beta) : \mathbb{R} \times \mathbb{R}^p \mid \beta \in \Omega \subset \mathbb{R}\}$$

- Atoms are (nonlinear) functions

$$\hat{\boldsymbol{y}} = \int \boldsymbol{F}(X(\beta), \beta) d\beta$$

Functional nonlinear dictionaries

- Dictionary has a continuum of atoms

$$\mathcal{D} = \{\boldsymbol{F}(\cdot, \beta) : \mathbb{R} \times \mathbb{R}^p \mid \beta \in \Omega \subset \mathbb{R}\}$$

- Atoms are (nonlinear) functions

- Sparse representations

$$\hat{\boldsymbol{y}} = \sum_{i=1}^{k} \boldsymbol{F}(x_i, \beta_i)$$

3 questions

- WHY functional nonlinear sparse models?
- WHY NOT functional nonlinear sparse models?
- HOW (if even possible) do we solve problems with them?

3 questions

- WHY functional nonlinear sparse models?
- WHY NOT functional nonlinear sparse models?
- HOW (if even possible) do we solve problems with them?

3 questions

- WHY functional nonlinear sparse models?
- WHY NOT functional nonlinear sparse models?
- HOW (if even possible) do we solve problems with them?

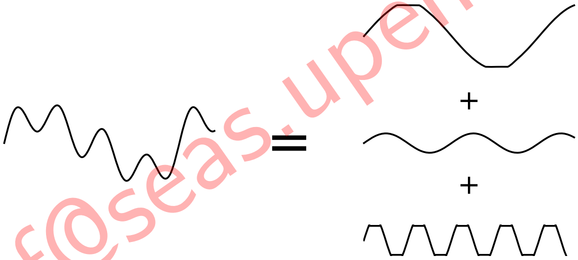
Functional nonlinear models: why?

- The physical world is continuous and nonlinear
- Linearity doesn't get you there
  - Super-resolution
  - Robustness

Discrete Continuous

Problem

Given samples from a mixture of few saturated sinusoids, determine their frequencies, amplitudes, and phases.



- Prototypical super-resolution problem: beamforming, radar, MRI...

Functional nonlinear models: why not?

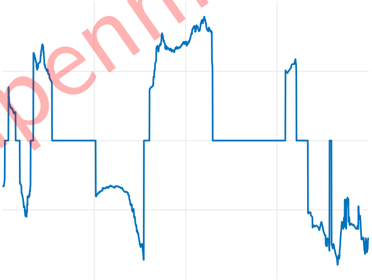
- Nonlinear models are hard (non-convex problems)
  - linearization
  - linear-in-the-parameters models (e.g., RKHS)
- Functional models are infinite dimensional
  - discretization
  - structure

3 questions

- WHY functional nonlinear sparse models?
- WHY NOT functional nonlinear sparse models?
- HOW (if even possible) do we solve problems with them?

Sparsity: why?

- Epistemological reasons
- Measurement and computational costs
- Interpretability



Sparsity: why not?

- Sparse models are hard (non-convex, often NP-hard)
  - convex relaxation (discrete case)
    - Minimize the ℓ_1-norm (*atomic norm*)
    - If the measurements are "incoherent" (NSP, RIP/REP), the relaxation yields the sparse solution
  - convex relaxation (continuous case)
    - Minimize L_1-norm or total variation
    - RIP-like "incoherence" conditions guarantee the relaxation yields the sparse solution

3 questions

- WHY functional nonlinear sparse models?
- WHY NOT functional nonlinear sparse models?
- HOW (if even possible) do we solve problems with them?

3 questions

- WHY functional nonlinear sparse models?
- WHY NOT functional nonlinear sparse models?
- HOW (if even possible) do we solve problems with them?
  - **Solution**: duality

Roadmap

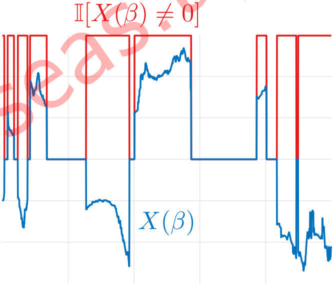
In words...

Variational problems that seek sparsest functions, i.e., functions with minimum support measure.

In words...

Variational problems that seek sparsest functions, i.e., functions with minimum support measure.

- "L_0-norm": $\|X\|_{L_0} = \mathfrak{m}\left[\text{supp}(X)\right] = \int_\Omega \mathbb{I}\left[X(\boldsymbol{\beta}) \neq 0\right] d\boldsymbol{\beta}$
  - Continuous counterpart of the "ℓ_0-norm" (not a norm!)



SFPs

$$\begin{aligned} \underset{X \in \mathcal{X}}{\text{minimize}} \quad & \int_\Omega F_0\left[X(\boldsymbol{\beta}), \boldsymbol{\beta}\right] d\boldsymbol{\beta} + \lambda \left\|X\right\|_{L_0} \\ \text{subject to} \quad & g_i(\boldsymbol{z}) \leq 0 \\ & \boldsymbol{z} = \int_\Omega \boldsymbol{F}\left[X(\boldsymbol{\beta}), \boldsymbol{\beta}\right] d\boldsymbol{\beta} \end{aligned} \tag{P-SFP}$$

- F_0 is an optional regularization, e.g., shrinkage: $\left\|X\right\|_{L_2}^2$
- g_i are convex losses, e.g., MSE, -LL, hinge loss
- $\boldsymbol{F}$ is a vector-valued nonlinear model, e.g., $\boldsymbol{F} \in \mathcal{D}$

Application: nonlinear spectral estimation

$$
\begin{aligned}
\underset{X \in L_2}{\text{minimize}} \quad & \|X\|_{L_2}^2 + \lambda \|X\|_{L_0} \\
\text{subject to} \quad & \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \leq \epsilon \\
& \hat{y}_i = B \int_0^{\frac{1}{2}} \rho\left[X(\varphi)\cos(2\pi\varphi t_i)\right] d\varphi
\end{aligned}
\tag{PI}
$$

- (y_i, t_i) are the measurements values and instants
- ρ models the saturation
- λ and B control sparsity and approximation
- ϵ is the fit slack ($\approx m\sigma_n^2$)

$$\begin{aligned} \underset{X \in \mathcal{X}}{\text{minimize}} \quad & \int_\Omega F_0\left[X(\boldsymbol{\beta}), \boldsymbol{\beta}\right] d\boldsymbol{\beta} + \lambda \left\| X \right\|_{L_0} \\ \text{subject to} \quad & g_i(\boldsymbol{z}) \leq 0 \\ & \boldsymbol{z} = \int_\Omega \boldsymbol{F}\left[X(\boldsymbol{\beta}), \boldsymbol{\beta}\right] d\boldsymbol{\beta} \end{aligned} \tag{P-SFP}$$

- Roadblocks:
  - Non-convexity $\Rightarrow$ ~~convex relaxation~~
  - Infinite dimensionality $\Rightarrow$ ~~discretization~~

Roadmap

SFP: The dual problem

- The primal problem (P-SFP)

$$\begin{aligned} \underset{X \in \mathcal{X}}{\text{minimize}} \quad & \int_\Omega F_0\left[X(\boldsymbol{\beta}), \boldsymbol{\beta}\right] d\boldsymbol{\beta} + \lambda \left\|X\right\|_{L_0} \\ \text{subject to} \quad & g_i(\boldsymbol{z}) \leq 0 \\ & \boldsymbol{z} = \int_\Omega \boldsymbol{F}\left[X(\boldsymbol{\beta}), \boldsymbol{\beta}\right] d\boldsymbol{\beta} \end{aligned} \tag{P-SFP}$$

- The dual problem of (P-SFP)

$$\underset{\boldsymbol{\mu},\ \nu_i \geq 0}{\text{maximize}} \quad d(\boldsymbol{\mu}, \nu_i) \triangleq \min_{\substack{\boldsymbol{z}, X \in \mathcal{X}, \\ X(\boldsymbol{\beta}) \in \mathcal{P}}} \mathcal{L}(X, \boldsymbol{z}, \boldsymbol{\mu}, \nu_i) \tag{D-SFP}$$

$$\begin{aligned}\mathcal{L}(X, \boldsymbol{z}, \boldsymbol{\mu}, \nu_i) = &\int_\Omega F_0 \left[X(\boldsymbol{\beta}), \boldsymbol{\beta}\right] d\boldsymbol{\beta} + \lambda \left\| X \right\|_{L_0} \\ &+ \boldsymbol{\mu}^T \left(\int_\Omega \boldsymbol{F} \left[X(\boldsymbol{\beta}), \boldsymbol{\beta}\right] d\boldsymbol{\beta} - \boldsymbol{z} \right) \\ &+ \sum_i \nu_i g_i(\boldsymbol{z})\end{aligned}$$

- Why?
  - (D-SFP) is convex and finite dimensional

Duality

- Why?
  - (D-SFP) is convex and finite dimensional
- Challenges
  - Non-convexity $\Rightarrow$ solving (D-SFP) $\neq$ solving (P-SFP)

Strong duality

Theorem

If F_0, $\boldsymbol{F}$, and $\mathcal{X}$ do not contain atoms and Slater's condition holds, then strong duality holds for (P-SFP), i.e., if P is the optimal value of (P-SFP) and D is the optimal value of (D-SFP), then $P = D$.

Corollary

We can obtain a solution of (P-SFP) from a solution of (D-SFP).

Duality

- Why?
  - (D-SFP) is convex and finite dimensional
- Challenges
  - Non-convexity $\not\Rightarrow$ strong duality

Duality

- Why?
  - (D-SFP) is convex and finite dimensional
- Challenges
  - Non-convexity $\Rightarrow$ strong duality
  - Can we even evaluate d?

SFP: The dual problem

- The dual problem of (P-SFP)

$$\underset{\boldsymbol{\mu},\ \nu_i \geq 0}{\text{maximize}} \quad d(\boldsymbol{\mu}, \nu_i) \triangleq \min_{\substack{\boldsymbol{z}, X \in \mathcal{X}, \\ X(\boldsymbol{\beta}) \in \mathcal{P}}} \mathcal{L}(X, \boldsymbol{z}, \boldsymbol{\mu}, \nu_i) \tag{D-SFP}$$

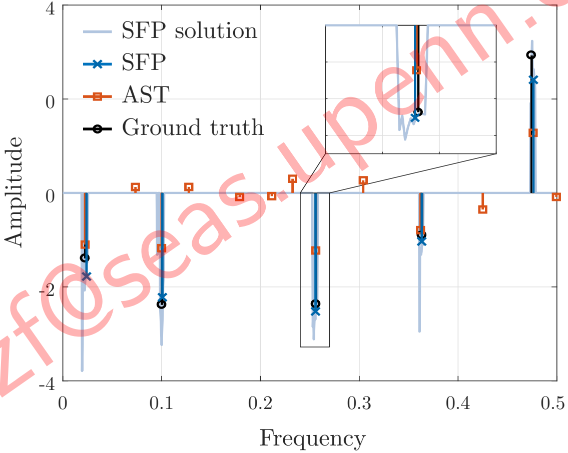
$$\begin{aligned} \mathcal{L}(X, \boldsymbol{z}, \boldsymbol{\mu}, \nu_i) = &\int_\Omega F_0\left[X(\boldsymbol{\beta}), \boldsymbol{\beta}\right] d\boldsymbol{\beta} + \lambda \left\|X\right\|_{L_0} \\ &+ \boldsymbol{\mu}^T \left(\int_\Omega \boldsymbol{F}\left[X(\boldsymbol{\beta}), \boldsymbol{\beta}\right] d\boldsymbol{\beta} - \boldsymbol{z}\right) \\ &+ \sum_i \nu_i g_i(\boldsymbol{z}) \end{aligned}$$

Duality

- Why?
  - (D-SFP) is convex and finite dimensional
- Challenges
  - Non-convexity $\Rightarrow$ solving (D-SFP) $\neq$ solving (P-SFP)
  - Can we even evaluate d? Yes (separability)

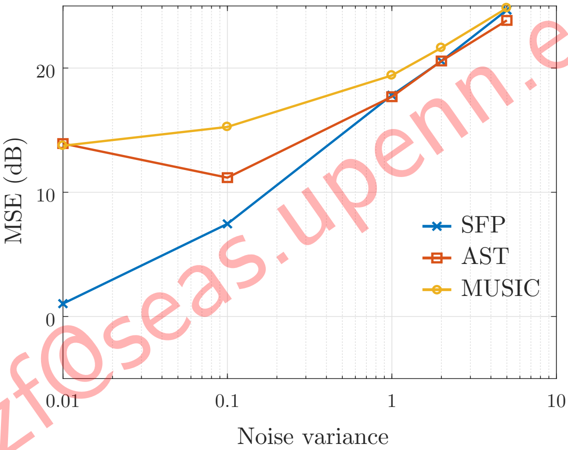
Roadmap

- SNR = 10 dB



SFP solution
SFP
AST
Ground truth
Amplitude
Frequency
4
0
2
0
-2
-4
0
0.1
0.2
0.3
0.4
0.5

Nonlinear spectral estimation: MSE vs. SNR



What else?

- Other applications:
  - robust classification [Chamon et al., ArXiV]
  - RKHS methods [Peifer et al., Wednesday, MLSP-P7.2]
- Other non-convexities
  - neural networks [Eisen et al., Friday, SPCOM-P4.1]

Conclusion

- WHY functional nonlinear sparse models?

- WHY NOT functional nonlinear sparse models?

- HOW (if even possible) do we solve problems with them?

Conclusion

- WHY functional nonlinear sparse models?
  - Functional nonlinear sparse models are versatile tools with a myriad of applications
- WHY NOT functional nonlinear sparse models?
- HOW (if even possible) do we solve problems with them?

Conclusion

- WHY functional nonlinear sparse models?
  - Functional nonlinear sparse models are versatile tools with a myriad of applications
- WHY NOT functional nonlinear sparse models?
  - Lead to non-convex and infinite dimensional optimization problems: SFPs
- HOW (if even possible) do we solve problems with them?

Conclusion

- WHY functional nonlinear sparse models?
  - Functional nonlinear sparse models are versatile tools with a myriad of applications
- WHY NOT functional nonlinear sparse models?
  - Lead to non-convex and infinite dimensional optimization problems: SFPs
- HOW (if even possible) do we solve problems with them?
  - SFPs can be solved exactly and efficiently using duality

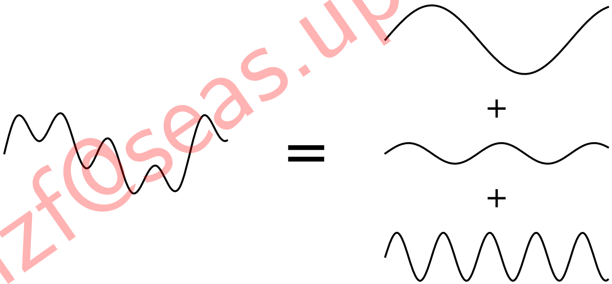
SPARSE RECOVERY OVER NONLINEAR DICTIONARIES

Luiz F. O. Chamon, Yonina C. Eldar, and Alejandro Ribeiro

"Functional nonlinear sparse models"
`https://arxiv.org/abs/1811.00577`

Problem

Given samples from a mixture of few ~~saturated~~ sinusoids, determine their frequencies, amplitudes, and phases.



Application: ~~non~~linear spectral estimation

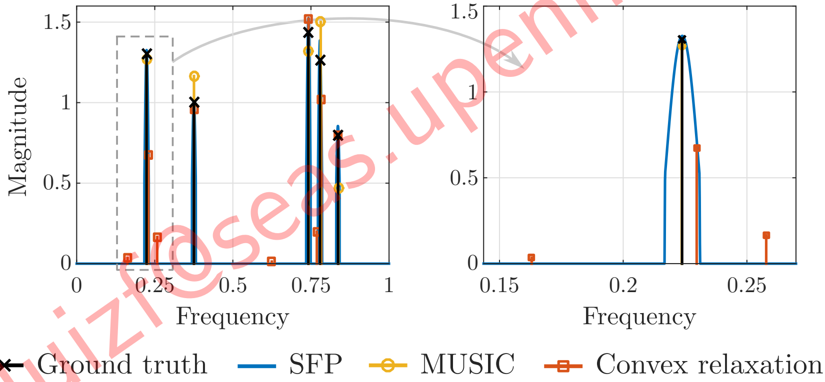
Problem

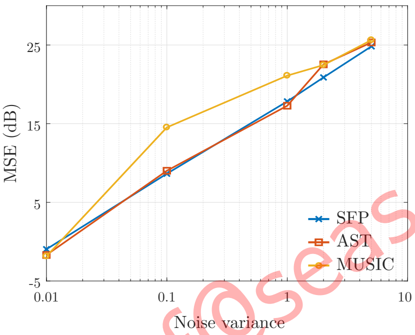
Given samples from a mixture of few ~~saturated~~ sinusoids, determine their frequencies, amplitudes, and phases.

- Classical solutions:
  - MUSIC (eigen-method)
  - AST (atomic norm relaxation)

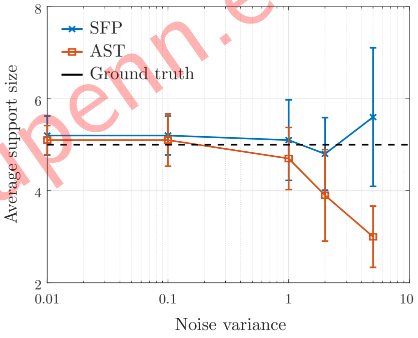
~~Non~~linear spectral estimation: typical solution

- SNR = 10 dB





MSE (dB)
25
15
5
-5
0.01
0.1
1
10
Noise variance
SFP
AST
MUSIC



Average support size
8
6
4
2
0.01
0.1
1
10
Noise variance
SFP
AST
Ground truth

Proposition

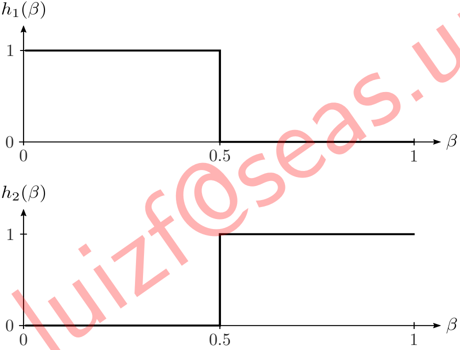
Consider the problem

$$
\begin{aligned}
P_q = \quad & \min_{X \in L_\infty} & & \|X\|_{L_q} \\
& \text{subject to} & & g_i(\boldsymbol{z}) \leq 0 \\
& & & \boldsymbol{z} = \int_\Omega \boldsymbol{F}[X(\boldsymbol{\beta}), \boldsymbol{\beta}] d\boldsymbol{\beta} \\
& & & |X| \leq \Gamma \text{ a.e.}
\end{aligned}
$$

Under mild conditions on $\boldsymbol{F}$, $P_0 = \dfrac{P_1}{\Gamma}$.

SFPs $\neq L_1$-norm minimization

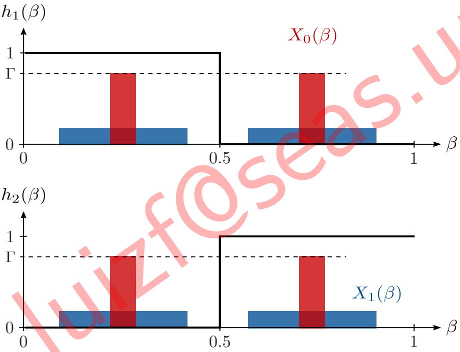
$$\begin{aligned} \min_{|X| \leq \Gamma} \quad & \|X\|_{L_q} \\ \text{subject to} \quad & \left\| \begin{bmatrix} \Gamma/8 \\ \Gamma/8 \end{bmatrix} - \boldsymbol{z} \right\|_2^2 \leq 0 \text{ and } \boldsymbol{z} = \int_0^1 \begin{bmatrix} h_1(\beta) \\ h_2(\beta) \end{bmatrix} X(\beta) d\beta \end{aligned}$$



SFPs $\neq L_1$-norm minimization

$$\min_{|X| \leq \Gamma} \quad \|X\|_{L_q}$$

$$\text{subject to} \quad \left\| \begin{bmatrix} \Gamma/8 \\ \Gamma/8 \end{bmatrix} - \boldsymbol{z} \right\|_2^2 \leq 0 \text{ and } \boldsymbol{z} = \int_0^1 \begin{bmatrix} h_1(\beta) \\ h_2(\beta) \end{bmatrix} X(\beta) d\beta$$



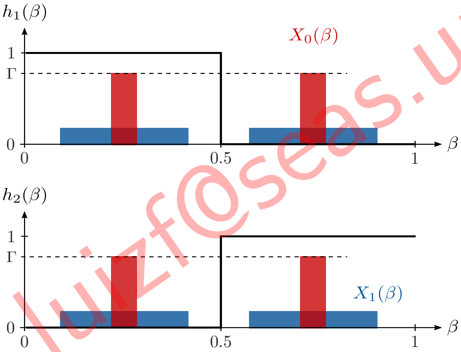
$$\|X_0\|_{L_1} = \frac{\Gamma}{4}$$

$$\|X_1\|_{L_1} = \frac{\Gamma}{4}$$

SFPs $\neq L_1$-norm minimization

$$\min_{|X| \leq \Gamma} \quad \|X\|_{L_q}$$

$$\text{subject to} \quad \left\| \begin{bmatrix} \Gamma/8 \\ \Gamma/8 \end{bmatrix} - \boldsymbol{z} \right\|_2^2 \leq 0 \text{ and } \boldsymbol{z} = \int_0^1 \begin{bmatrix} h_1(\beta) \\ h_2(\beta) \end{bmatrix} X(\beta) d\beta$$



$$\|X_0\|_{L_1} = \frac{\Gamma}{4} \qquad \|X_0\|_{L_0} = \frac{1}{4}$$

$$\|X_1\|_{L_1} = \frac{\Gamma}{4} \qquad \|X_1\|_{L_0} = \frac{5}{4}$$

Theorem

Under Slater's condition holds, strong duality holds for (P-SFP), i.e., if P *is the optimal value of (P-SFP) and* D *is the optimal value of (D-SFP), then* $P = D$.

Proof recipe

(i) Show cost-constraint set $\mathcal{C}$ is convex

$$\mathcal{C} = \left\{ (c, \boldsymbol{z}) \;\middle|\; \exists X \in \mathcal{X} \text{ with } X(\boldsymbol{\beta}) \in \mathcal{P} \right.$$
$$\left. \text{s.t. } c = f_0(X) \text{ and } \boldsymbol{z} = \int_\Omega \boldsymbol{F}\left[X(\boldsymbol{\beta}), \boldsymbol{\beta}\right] d\boldsymbol{\beta} \right\}$$

(ii) $\mathcal{C}$ convex $\Rightarrow$ perturbation function is convex

(iii) Convex perturbation function $\Rightarrow$ strong duality

Proposition

For fixed a, t, and f,

$$B \int_0^{\frac{1}{2}} \rho \left[X'(\varphi) \cos(2\pi\varphi t) \right] d\varphi \to \rho \left[a \cos(2\pi f t) \right] \text{ as } B \to \infty.$$