STRONG DUALITY OF SPARSE FUNCTIONAL OPTIMIZATION

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

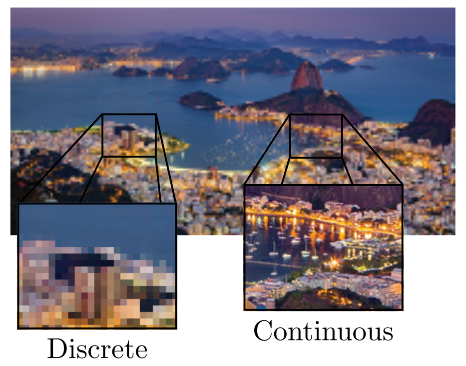
e-mail: luizf@seas.upenn.edu, yonina@ee.technion.ac.il, aribeiro@seas.upenn.edu

* Electrical and Systems Engineering, University of Pennsylvania, † Department of Electrical Engineering, Technion—Israel Institute of Technology

INTRODUCTION

Sparse functional optimization: why?

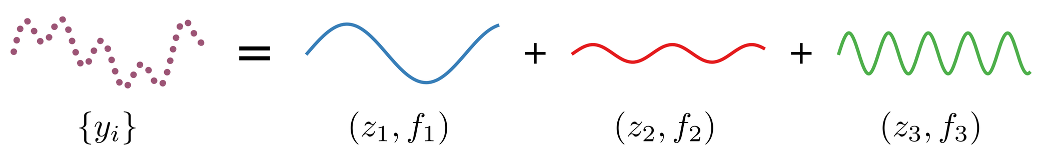
- Signal processing is rich in inherently continuous problems (e.g., imaging, radar, continuous dictionaries...)
- Finite # of samples ⇒ underdetermined problem
- Exploit structure: sparsity



Line spectral estimation

- Given noisy samples $\{y(i)\}$ taken at instants $\{t_i\}$ from a superposition of n complex exponentials, find their frequencies ($f_k \in [0,1]$), amplitudes, and phases ($z_k \in \mathbb{C}$).

$$y(i) = \sum_{k=1}^{n} z_k \exp(j2\pi f_k t_i) + v(i)$$



- Reparametrization using an overcomplete, continuous dictionary:

$$y(i) = \int_0^1 \exp(j2\pi\beta t_i) X^o(\beta) d\beta + v(i)$$

$X^o(\beta) = \sum_{k=1}^{n} z_k \delta(\beta - f_k)$ (ideal)

$X^o(\beta) = \sum_{k=1}^{n} z_k \kappa(\beta - f_k)$ (real)

- Few components ⇒ X^o is sparse (small support)

SPARSE FUNCTIONAL PROBLEMS

- $(\Omega, \mathcal{B})$ measurable space with $\Omega \subset \mathbb{R}$ compact
- Measurement vector $\boldsymbol{y} \in \mathbb{C}^m$
- Functional linear model: $\int_\Omega \boldsymbol{h}(\beta) X(\beta) d\beta$
  - $\boldsymbol{h} : \Omega \to \mathbb{C}^m$ with measurable components $h_i \in L_2(\Omega)$

Problem

Find the sparsest functional linear model that fits the measurements $\boldsymbol{y}$.

$$\begin{aligned} \underset{X \in L_2(\Omega)}{\text{minimize}} \quad & \int_\Omega \mathbb{I}\left[X(\beta) \neq 0\right] d\beta + \lambda \|X\|_{L_2}^2 \\ \text{subject to} \quad & \left\| \boldsymbol{y} - \int_\Omega \boldsymbol{h}(\beta) X(\beta) d\beta \right\|_2^2 \leq w \end{aligned} \tag{PI}$$

Roadblocks

- Infinite dimensionality
  - discretization [Tang'13, Duval'17]
  - duality [Shapiro'06, Tang et al.'13, Bhaskar et al.'13]
- Non-convexity
  - convex relaxation [Tang'13, Bhaskar'13, Adcock'17, Puy'17]

OUR APPROACH

Solve (PI) exactly using duality

The ingredients

- **Separability** ⇒ closed form for the dual problem of (PI)
- **Strong duality** ⇒ solving the dual problem yields a solution of (PI)

THE DUAL PROBLEM

- Reformulate (PI)

$$\begin{aligned} \underset{X \in L_2(\Omega)}{\text{minimize}} \quad & \int_\Omega \mathbb{I}\left[X(\beta) \neq 0\right] d\beta + \lambda \|X\|_{L_2}^2 \\ \text{subject to} \quad & \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2^2 \leq w \\ & \hat{\boldsymbol{y}} = \int_\Omega \boldsymbol{h}(\beta) X(\beta) d\beta \end{aligned} \tag{PI′}$$

- The dual function

$$\begin{aligned} d(\boldsymbol{\mu}, \nu) &= \min_{X \in L_2(\Omega), \hat{\boldsymbol{y}}} \mathcal{L}(X, \hat{\boldsymbol{y}}, \boldsymbol{\mu}, \nu) \\ &= \min_{X \in L_2(\Omega)} \int_\Omega \left[\mathbb{I}(X(\beta) \neq 0) + \lambda |X(\beta)|^2 + \operatorname{Re}\left(\boldsymbol{\mu}^H \boldsymbol{h} X(\beta)\right)\right] d\beta \\ &+ \min_{\hat{\boldsymbol{y}}} \nu \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2^2 - \operatorname{Re}[\boldsymbol{\mu}^H \hat{\boldsymbol{y}}] - \nu w \end{aligned}$$

THE DUAL MINIMIZERS

- Convex quadratic minimization: $\hat{\boldsymbol{y}}_d(\boldsymbol{\mu}, \nu) = \boldsymbol{y} + \dfrac{\boldsymbol{\mu}}{2\nu}$
- $L_2(\Omega)$ is separable and the integrand is normal, so we can solve for each β separately [Rockafellar'76]

$$\inf_{X \in L_2(\Omega)} \int_\Omega F(\beta, X(\beta)) d\beta = \int_\Omega \inf_{X(\beta) \in \mathbb{C}} F(\beta, X(\beta)) d\beta$$

$$X_d(\beta, \boldsymbol{\mu}) = -\frac{1}{2\lambda} \boldsymbol{h}(\beta)^H \boldsymbol{\mu} \times \mathbb{I}\left(\left|\boldsymbol{h}(\beta)^H \boldsymbol{\mu}\right|^2 > 4\lambda\right)$$

SOLVING THE DUAL PROBLEM

$$\begin{aligned} \underset{\boldsymbol{\mu} \in \mathbb{C}^m,\ \nu \geq 0}{\text{maximize}} \quad & \mathfrak{m}(\mathcal{S}) - \frac{1}{4\lambda} \boldsymbol{\mu}^H \boldsymbol{H} \boldsymbol{\mu} - \operatorname{Re}\left[\boldsymbol{\mu}^H \boldsymbol{y}\right] - \frac{\|\boldsymbol{\mu}\|_2^2}{4\nu} - \nu w \\ \text{subject to} \quad & \boldsymbol{H} = \int_{\mathcal{S}} \boldsymbol{h}(\beta) \boldsymbol{h}(\beta)^H d\beta \\ & \mathcal{S} = \left\{\beta \in \Omega \mid \left|\boldsymbol{h}(\beta)^H \boldsymbol{\mu}\right|^2 > 4\lambda\right\} \end{aligned} \tag{DI}$$

Stochastic Gradient Ascent

$\boldsymbol{\mu}_0 = \mathbf{0}$, $\nu_0 = 1$

for $t = 1, \ldots, T$

Draw β_j uniformly at random in Ω

$$\bar{\boldsymbol{H}} = \frac{1}{p} \sum_{j=1}^{p} \boldsymbol{h}(\beta_j) \boldsymbol{h}(\beta_j)^H \times \mathbb{I}\left(\left|\boldsymbol{\mu}^H \boldsymbol{h}(\beta_j)\right|^2 > 4\lambda\right)$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} - \eta_t \left[\frac{1}{2\lambda\nu_{t-1}} \left(\nu_{t-1} \bar{\boldsymbol{H}} + \lambda \boldsymbol{I}\right) \boldsymbol{\mu}_{t-1} + \boldsymbol{y}\right]$$

$$\nu_t = \left[\nu_{t-1} + \eta_t \left(\frac{\|\boldsymbol{\mu}_{t-1}\|_2^2}{4\nu_{t-1}^2} - w\right)\right]_+$$

end

STRONG DUALITY OF SPARSE FUNCTIONAL PROGRAMS

Theorem (Strong duality of sparse functional optimization)

Suppose that $\boldsymbol{h}$ in (PI) has no point masses (Dirac deltas). Then, strong duality holds for (PI), i.e., if P is the optimal value of (PI) and D is the optimal value of (DI), then $P = D$.

Corollary

Let $X^\star$ be the solution of (PI) and $\boldsymbol{\mu}^\star$ be the solution of (DI). Then, $X^\star(\beta) = X_d(\beta, \boldsymbol{\mu}^\star)$.

Proof sketch

- Reformulate (PI) as

$$\underset{c \in \mathcal{C}}{\text{minimize}} \ c \tag{PI″}$$

$$\mathcal{C} = \left\{ s \in \mathbb{R} \;\middle|\; \exists X \in L_2(\Omega) \text{ s.t. } s = \int_\Omega \mathbb{I}(X(\beta) \neq 0) d\beta + \lambda \|X\|_{L_2}^2 \text{ and } \left\| \boldsymbol{y} - \int_\Omega \boldsymbol{h}(\beta) X(\beta) d\beta \right\|_2^2 \leq w \right\}$$

Lemma

The cost-constraint set $\mathcal{C}$ is convex.

- Take $c, c' \in \mathcal{C}$ achieved for $X, X' \in L_2(\Omega)$. Construct a vector measure $\mathfrak{p}$ over $(\Omega, \mathcal{B})$ such that

$$\mathfrak{p}(\mathcal{Z}) = \begin{bmatrix} \int_{\mathcal{Z}} \left[\mathbb{I}(X(\beta) \neq 0) + \lambda |X(\beta)|^2\right] d\beta \\ \int_{\mathcal{Z}} \left[\mathbb{I}(X'(\beta) \neq 0) + \lambda |X'(\beta)|^2\right] d\beta \end{bmatrix}$$

Note that $\mathfrak{p}(\emptyset) = \mathbf{0}$ and

$$\mathfrak{p}(\Omega) = \begin{bmatrix} \int_\Omega \mathbb{I}(X(\beta) \neq 0) d\beta + \lambda \|X\|_{L_2}^2 \\ \int_\Omega \mathbb{I}(X'(\beta) \neq 0) d\beta + \lambda \|X'\|_{L_2}^2 \end{bmatrix} = \begin{bmatrix} c \\ c' \end{bmatrix}$$

Lyapunov's convexity theorem

If $\mathfrak{p}$ is a non-atomic measure, then its range is convex. Explicitly, for every $\theta \in [0,1]$ there exists $\mathcal{T}_\theta \in \mathcal{B}$ such that

$$\mathfrak{p}(\mathcal{T}_\theta) = \theta \mathfrak{p}(\Omega) + (1-\theta) \mathfrak{p}(\emptyset) = \theta \mathfrak{p}(\Omega)$$

- By additivity: $\mathfrak{p}(\Omega \setminus \mathcal{T}_\theta) = (1-\theta) \mathfrak{p}(\Omega)$.
- Let $X_\theta(\beta) = \begin{cases} X(\beta), & \text{for } \beta \in \mathcal{T}_\theta \\ X'(\beta), & \text{for } \beta \in \Omega \setminus \mathcal{T}_\theta \end{cases}$
- Then, $X_\theta \in L_2(\Omega)$ and (PI)-feasible

$$\begin{aligned} &\left\| \boldsymbol{y} - \int_\Omega \boldsymbol{h}(\beta) X_\theta(\beta) d\beta \right\|_2^2 = \\ &\left\| \boldsymbol{y} - \theta \int_\Omega \boldsymbol{h}(\beta) X(\beta) d\beta - (1-\theta) \int_\Omega \boldsymbol{h}(\beta) X'(\beta) d\beta \right\|_2^2 \leq \\ &\theta \left\| \boldsymbol{y} - \int_\Omega \boldsymbol{h}(\beta) X(\beta) d\beta \right\|_2^2 + (1-\theta) \left\| \boldsymbol{y} - \int_\Omega \boldsymbol{h}(\beta) X'(\beta) d\beta \right\|_2^2 \leq w \end{aligned}$$

- So the objective evaluated at X_θ is in $\mathcal{C}$ and

$$\int_\Omega \mathbb{I}(X_\theta(\beta) \neq 0) d\beta + \lambda \|X_\theta(\beta)\|_{L_2}^2 = \theta c + (1-\theta) c'$$

- Hence, $c, c' \in \mathcal{C} \Rightarrow \theta c + (1-\theta) c' \in \mathcal{C}$ for $\theta \in [0,1]$. ■

APPLICATION: LINE SPECTRAL ESTIMATION

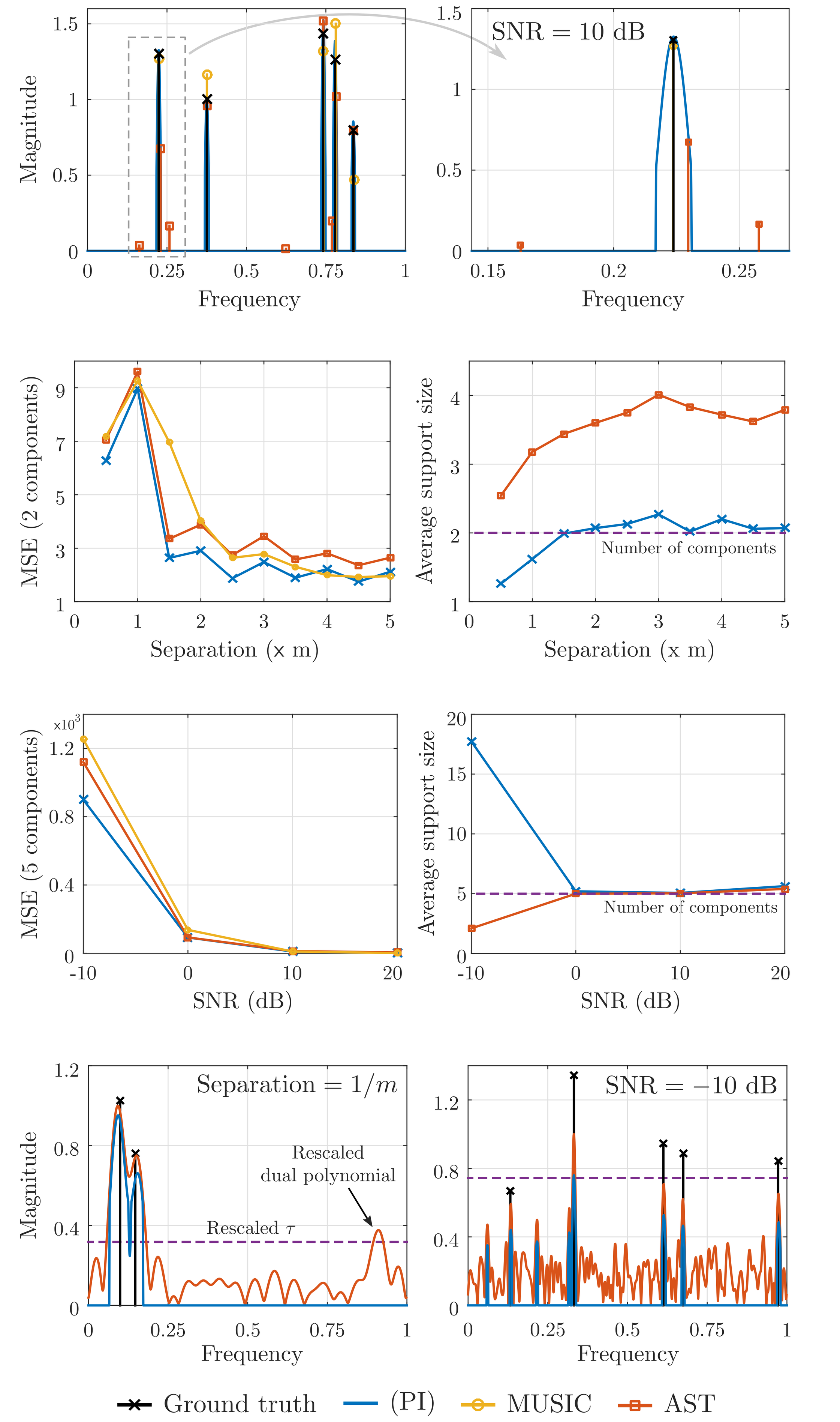
Other methods

- **MUSIC**: EVD-based method; assumes number of components n is known.
- **Atomic norm relaxation**: The support is estimated from the maxima of a polynomial obtained from $\boldsymbol{x}^\star$, the solution of the SDP

$$\begin{aligned} \underset{\boldsymbol{u}, \boldsymbol{x}, t}{\text{minimize}} \quad & \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \frac{\tau}{2}(t + u_1) \\ \text{subject to} \quad & \begin{bmatrix} T(\boldsymbol{u}) & \boldsymbol{x} \\ \boldsymbol{x}^H & t \end{bmatrix} \succeq 0 \end{aligned} \tag{AST}$$

where $T(\boldsymbol{a})$ is a Hermitian Toeplitz matrix with entries from $\boldsymbol{a}$ and $\tau > 0$ is a regularization parameter [Bhaskar'13].

Simulations



CONCLUSION

Although sparse functional optimization problem are infinite dimensional and non-convex, they have zero duality gap. They can therefore be solved efficiently by leveraging duality and stochastic gradient methods.