Multi-Pitch Estimation via Semidefinite Programming

August 25, 2016

T. L. Jensen
Joint work with L. Vandenberghe, UCLA

Dept. of Electronic Systems

Aalborg University

AALBORG UNIVERSITY
DENMARK

Agenda

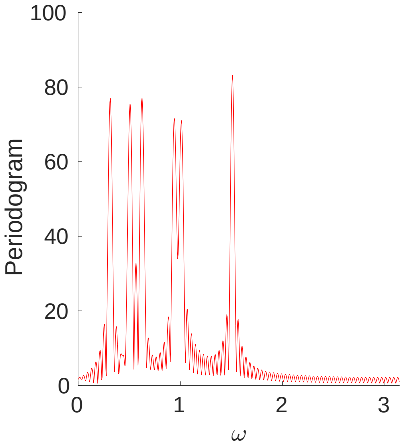
- Multi-pitch estimation.
- Superresolution/gridless/atomic norm using semidefinite programming.
- Bringing it together.
- Complex- and real-valued data.
- Solvers:
  - An interior point method based on a non-standard LMI representation.
  - An alternating direction method of multipliers.
- Simulations.

Multi-pitch estimation I

- Harmonic signals: Fundamental/pitch ω_k, first harmonic $2 \cdot \omega_k$, second harmonic $3 \cdot \omega_k$.
- Multi-pitch: superposition of $k = 1, \ldots, K$ harmonic signals.
- Application in music, speech, vibration analysis etc.

Multi-pitch estimation II

- $K = 2$ pitches
- $L = 3$ harmonics
- $N = 160$ samples
- SNR = 31 [dB]



- Multi-pitch estimation: Estimate ω_k and amplitudes[1].
- Problem may be ill-posed or ill-conditioned.

[1] M. G. Christensen and A. Jakobsson. *Multi-Pitch Estimation*. San Rafael, CA, USA: Morgan & Claypool, 2009.

Atomic decomposition

- Atomic decomposition over a continuous dictionary $\mathbb{A}_n \subseteq \mathbb{C}^n$ using a regularization term

$$\begin{array}{ll} \text{minimize} & f(\sum_{k=1}^r a_k c_k^H) + \sum_{k=1}^r \|c_k\|_2 \\ \text{subject to} & a_k \in \mathbb{A}_n,\ k = 1, \ldots, r \end{array} . \tag{1}$$

- Variables: Atoms $a_k \in \mathbb{C}^n$, coefficients $c_k \in \mathbb{C}^m, k = 1, \ldots, r$ and the number of selected atoms r.
- $m = 1$ single measurement, $m > 1$ multiple measurement case. Notice a kind of (group)-sparsity promoting term.
- In current literature: Often

$$\mathbb{A}_n = \left\{ \frac{s}{\sqrt{n}} \left[1, \exp(j\omega), \ldots, \exp(j(n-1)\omega)\right]^T \;\middle|\; |\omega - \alpha| \leq \beta, |s| = 1, s \in \mathbb{C} \right\} \tag{2}$$

with $\alpha = 0$ and $\beta = \pi$.

Atomic decomposition as a SDP

- With $\alpha = 0$ and $\beta = \pi$, f convex, the atomic decomposition is equivalent to the SDP

$$\begin{array}{ll} \text{minimize} & f(X_{12}) + \frac{1}{2}(\mathbf{tr}\, X_{11} + \mathbf{tr}\, X_{22}) \\ \text{subject to} & \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^H & X_{22} \end{bmatrix} \succeq 0 \\ & X_{11} \in \mathbb{T}^n \\ & X_{12} \in \mathbb{C}^{n \times m},\ X_{22} \in \mathbb{H}^m \end{array} \tag{3}$$

with $r = \mathbf{rank}(X_{11}^\star)$.[2]

[2] E. J. Candès and C. Fernandez-Granda. "Super-resolution from noisy data". In: *J. Fourier Anal. Appl.* 19.6 (2013), pp. 1229–1254; G. Tang et al. "Compressed Sensing Off the Grid". In: *IEEE Trans. Information Theory* 59.11 (2013), pp. 7465–7490; B. N. Bhaskar, G. Tang, and B. Recht. "Atomic Norm Denoising With Applications to Line Spectral Estimation". In: *IEEE Trans. Signal Processing* 61.23 (2013), pp. 5987–5999; Y. Li and Y. Chi. "Off-the-Grid Line Spectrum Denoising and Estimation With Multiple Measurement Vectors". In: *IEEE Trans. Signal Processing* 64.5 (2016), pp. 1257–1269.

Complex-valued multi-pitch model

The complex-valued multi-pitch model can be formulated as

$$x = \sum_{l=1}^{L} Z_K(l\omega)\bar{c}_l, \quad y = x + w \tag{4}$$

with

$$y = \left[y_0, \ldots, y_{N-1}\right]^T \tag{5}$$

$$\bar{c}_l = \left[\bar{c}_{l,1}, \ldots, \bar{c}_{l,K}\right]^T \tag{6}$$

$$\omega = \left[\omega_1, \ldots, \omega_K\right]^T \tag{7}$$

$$Z_K(\omega) = \left[z(\omega_1), \ldots, z(\omega_K)\right] \tag{8}$$

$$z(\omega_k) = \left[1, \exp(j\omega_k), \ldots, \exp(j(N-1)\omega_k)\right]^T \tag{9}$$

$$w = \left[w_0, \ldots, w_{N-1}\right]^T \sim \mathcal{CN}(0, \sigma^2 I)\,. \tag{10}$$

Bringing it together I

- Relating the formulations at $n = NL$

$$X_{12} = \sum_{k=1}^{r} a_k c_k^H, \quad a_k \in \mathbb{A}_{NL}\,. \tag{11}$$

- Define the selection matrix P_l that selects N elements $P_l v$ from every lth element of v, $P_l v = \left[v_1, v_{1+l}, \ldots, v_{1+(N-1)l}\right]$. Then

$$z(l\omega_k) = P_l a_k, \quad \text{for some } a_k \in \mathbb{A}_{NL} \tag{12}$$

and we may form the selection and add matrix

$$P = \begin{bmatrix} P_1 & P_2 & \cdots & P_L \end{bmatrix} \in \mathbb{R}^{N \times NL^2}, P_l \in \mathbb{R}^{N \times NL}\,. \tag{13}$$

Bringing it together II

- Let $c_k = \begin{bmatrix} [\bar{c}_1]_k & \cdots & [\bar{c}_L]_k \end{bmatrix}^H$.
- Then

$$\begin{aligned}
\sum_{l=1}^{L} Z_K(l\omega)\bar{c}_l &= \sum_{l=1}^{L}\sum_{k=1}^{K} z(l\omega_k)[\bar{c}_l]_k \\
&= \sum_{k=1}^{K}\sum_{l=1}^{L} P_l a_k [\bar{c}_l]_k \\
&= \sum_{k=1}^{K} P\,\mathbf{vec}(a_k c_k^H) \\
&= P\,\mathbf{vec}\left(\sum_{k=1}^{K} a_k c_k^H\right) \\
&= P\,\mathbf{vec}\left(X_{12}\right)
\end{aligned}$$

for some $a_k \in \mathbb{A}_{NL}, k = 1, \ldots, K$ and $K = r$.

A complex-valued SDP formulation

- A complex-valued multi-pitch estimator can then be formulated via the SDP

$$
\begin{array}{ll}
\text{minimize} & \frac{1}{2}(\mathbf{tr}(X_{11}) + \mathbf{tr}(X_{22})) \\
\text{subject to} & \|y - x\|_2 \leq \delta \\
& x = P\,\mathbf{vec}(X_{12}) \\
& \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^H & X_{22} \end{bmatrix} \succeq 0 \\
& X_{11} \in \mathbb{T}^{NL} \\
& X_{22} \in \mathbb{H}^{L}, X_{12} \in \mathbb{C}^{NL \times L}\,.
\end{array} \tag{14}
$$

A real-valued SDP formulation I

- The real-valued model is

$$x = \Re\left(\sum_{l=1}^{L} Z_K(l\omega)\bar{c}_l\right), \quad y = x + w \tag{15}$$

with $w \sim \mathcal{N}(0, \sigma^2 I)$.

- A real-valued $y \in \mathbb{R}^N$ atomic norm multi-pitch SDP estimator is

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}(\mathbf{tr}(X_{11}) + \mathbf{tr}(X_{22})) \\ \text{subject to} & \|y - P\,\mathbf{vec}(\Re(X_{12}))\|_2 \leq \delta \\ & \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^H & X_{22} \end{bmatrix} \succeq 0 \\ & X_{11} \in \mathbb{T}^{NL} \\ & X_{22} \in \mathbb{H}^L,\ X_{12} \in \mathbb{C}^{NL \times L} \end{array} \tag{16}$$

with a solution $(X_{11}^\star, X_{22}^\star, X_{12}^\star)$.

A real-valued SDP formulation II

- The optimal objective is $\frac{1}{2}(\mathbf{tr}(X_{11}^\star) + \mathbf{tr}(X_{22}^\star)) = \frac{1}{2}(\mathbf{tr}(\Re(X_{11}^\star)) + \mathbf{tr}(\Re(X_{22}^\star))$ and

$$\begin{bmatrix} X_{11}^\star & X_{12}^\star \\ (X_{12}^\star)^H & X_{22}^\star \end{bmatrix} \succeq 0 \Rightarrow \Re\left(\begin{bmatrix} X_{11}^\star & X_{12}^\star \\ (X_{12}^\star)^H & X_{22}^\star \end{bmatrix}\right) \succeq 0\,. \tag{17}$$

- If $X_{11}^\star$ is Toeplitz, then $\Re(X_{11}^\star)$ is also Toeplitz.
- So, $(\Re(X_{11}^\star), \Re(X_{22}^\star), \Re(X_{12}^\star))$ also solves the previous SDP.
- We can instead solve the equivalent real SDP

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}(\mathbf{tr}(X_{11}) + \mathbf{tr}(X_{22})) \\ \text{subject to} & \|y - P\,\mathbf{vec}(X_{12})\|_2 \le \delta \\ & \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \succeq 0 \\ & X_{11} \in \mathbb{S}^{NL} \cap \mathbb{T}^{NL} \\ & X_{22} \in \mathbb{S}^{L}, X_{12} \in \mathbb{R}^{NL \times L} \end{array} \tag{18}$$

with a solution that also solves the complex SDP (16).

Frequency constraint I

- If the signal y is Nyquist sampled: $-\pi \leq L\omega_k \leq \pi$.
- Recall the dictionary $\mathbb{A}_n$:

$$\mathbb{A}_n = \left\{ \frac{s}{\sqrt{n}} \left[1, \exp(j\omega), \ldots, \exp(j(n-1)\omega)\right]^T \;\middle|\; |\omega - \alpha| \leq \beta, |s| = 1, s \in \mathbb{C} \right\}. \quad (19)$$

- The constrained controlled by the parameters α, β can be imposed by adding a semidefinite cone constraint[3]

$$-e^{j\alpha} F X_{11} G^T - e^{-j\alpha} G X_{11} F^T + 2\cos(\beta) G X_{11} G^T \preceq 0 \quad (20)$$

where $F = \begin{bmatrix} 0 & I_{NL-1} \end{bmatrix}$, $G = \begin{bmatrix} I_{NL-1} & 0 \end{bmatrix}$.

[3] H.-H. Chao and L. Vandenberghe. "Extension of semidefinite programming methods for atomic decomposition". In: *ICASSP*. 2016, pp. 4757–4761.

Frequency constraint II

- With the selection $\alpha = 0$, $\beta = \pi/L$ it is a real semidefinte cone constraint.
- Let a Hermitian Toeplitz be given by

$$T(z) = \begin{bmatrix} z_0 & z_1^* & \cdots & z_n^* \\ z_1 & z_0 & \cdots & z_{n-1}^* \\ \vdots & \vdots & \ddots & \vdots \\ z_n & z_{n-1} & \cdots & z_0 \end{bmatrix}. \tag{21}$$

- Let x_{11} be such that $T(x_{11}) = X_{11}$, then the semidefinte cone constraint is

$$T(\begin{bmatrix} 2t_1 \\ t_0 + t_2 \\ \vdots \\ t_{NL-3} + t_{NL-1} \end{bmatrix}) - 2\cos(\beta) T(\begin{bmatrix} t_0 \\ t_1 \\ \vdots \\ t_{NL-2} \end{bmatrix}) = T(C_\beta x_{11}) \succeq 0 \tag{22}$$

Frequency constrained SDP based multi-pitch estimator

$$
\begin{array}{ll}
\text{minimize} & \frac{1}{2}(\mathbf{tr}(T(x_{11})) + \mathbf{tr}(X_{22})) \\
\text{subject to} & \|y - P\,\mathbf{vec}(X_{12})\|_2 \leq \delta \\
& \begin{bmatrix} T(x_{11}) & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \succeq 0 \\
& T(C_{\pi/L} x_{11}) \succeq 0 \\
& x_{11} \in \mathbb{R}^{NL} \\
& X_{22} \in \mathbb{S}^L, X_{12} \in \mathbb{R}^{NL \times L}\,.
\end{array} \tag{23}
$$

- The matrix $X_{11} = T(x_{11}) \in \mathbb{S}^{NL} \cap \mathbb{T}^{NL}$ will contain pairs of symmetric frequencies $\omega_k = -\omega_{k+1}$ but with $|\omega_k| = |\omega_{k+1}| \leq \beta = \pi/L$ and possible a single $\omega_k = 0$.

Solver based on an interior-point methods

- Need to exploit the Toeplitz part: NL versus $(NL)^2$ variables.
- Canonical cone program:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & s = h - Gx, s \in \mathbb{K} \end{array} \tag{24}$$

  where $\mathbb{K}$ represents a Cartesian product of linear, second order and semidefinite cones.
- Often it takes more operations to form the "normal equations form" than solving it

$$Kz = G^T W^{-T} W^{-1} Gz = b \tag{25}$$

  where W is an iteration dependent scaling matrix.

IP: another representation

- Assume that the vth LMI can be represented using:

$$-\,\mathbf{mat}(G_v x) = B_v\,\mathbf{diag}(A_v x) B_v^T \tag{26}$$

then[4]

$$K_v = G_v^T W_v^{-T} W_v^{-1} G_v \tag{27}$$
$$= A_v^T \left((B_v^T R_v^{-T} R_v^{-1} B_v) \odot (B_v^T R_v^{-T} R_v^{-1} B_v) \right) A_v \tag{28}$$

where $W_v = R_v^T \otimes R_v^T$.

[4] T. Roh and L. Vandenberghe. "Discrete transforms, semidefinite programming and sum-of-squares representations of nonnegative polynomials". In: *SIAM J. Optimiz.* 16 (2006), pp. 939–964.

IP: why another representation?

- Let the first n column of a $N \geq n$ discrete Fourier transform (DFT) matrix be F, Then

$$T(z) = \frac{1}{N} F^H \,\mathbf{diag}(\Re(Fz \odot \begin{bmatrix} 1 \\ 2 \\ \vdots \\ 2 \end{bmatrix})F \tag{29}$$

- If z is real, then $y = Fz \odot \begin{bmatrix} 1 & 2 & \cdots & 2 \end{bmatrix}^T$ has Hermitian symmetry such that

$$T(z) = \frac{1}{N}(\Re(F) + \Im(F))^T \,\mathbf{diag}(\Re(y))(\Re(F) + \Im(F)) \tag{30}$$

IP: example of the representation

- Then the representation of the second LMI $T(C_\beta x_{11}) \succeq 0$ is:

$$B_2 = \Re(F)^T + \Im(F)^T, \quad A_2 = \frac{1}{N}\Re(F)\,\mathbf{diag}(\begin{bmatrix}1 & 2 & \cdots & 2\end{bmatrix})C_\beta \tag{31}$$

- Notice that multiplications $B_v^T R_v^{-T}$ and $A_v^T X$ is possible via FFTs and IFFTs.
- Similar approach is possible with the first LMI

$$\begin{bmatrix} T(x_{11}) & X_{21} \\ X_{21}^T & X_{22} \end{bmatrix} \succeq 0 \tag{32}$$

but is a bit more complicated due to X_{21} and X_{22}.

An alternating direction method of multipliers approach.

- Main elements:
  - Projection on the set of symmetric Toeplitz matrices: averaging along the diagonal $\mathcal{O}(NL^2)$.
  - Projection on $\{x \mid \|y - Px\|_2 \leq \delta\}$, P has orthogonal rows: $\mathcal{O}(NL)$
  - Projection on the set of symmetric positive semidefinite matrices: eigenvalue decomposition $\mathcal{O}((NL)^3)$.

Quick and dirty comparison

- $N = 160$, $L = 3$: $NL + NL^2 + L^2 = 1929$ primal variables:
  - ADMM at 350 iterations: 30.4 s
  - CVXOPT[5] + custom solver: 29.7 s (21 iterations)
- Comments
  - Not done a huge effort in code optimization.
  - ADMM is easier to implement.

[5] M. S. Andersen et al. "Interior-point methods for large-scale cone programming". In: *Optimization for Machine Learning*. Ed. by S. Sra, S. Nowozin, and S. J. Wright. MIT Press, 2011.

Simulations I

- Monte Carlo, $R = 500$ repetitions, known model-order, $K = 2$, $L = 3$, real-valued data otherwise same setup as[6].
- δ: *1)* solve the SDP with δ selected by averaging the smallest $\frac{1}{3}$ of the coefficients of the periodogram *2)* extract the frequencies $\omega^\star$, re-select the regularization parameter as minimum of linear least-squares, re-solve the SDP.

[6] M. G. Christensen et al. "Multi-pitch estimation". In: *Signal Processing* 88.4 (Apr. 2008), pp. 972–983.

- The accuracy should at-least for unbiased estimators be governed by the asymptotic Cramér-Rao lower bound (CRLB) for estimating a single fundamental $\hat{\omega}_k$:

$$\mathbf{var}(\hat{\omega}_k) \geq \frac{24\sigma^2}{(N(N^2-1))\sum_{l=1}^{L} A_{k,l}^2 l^2} \tag{33}$$

where $A_{k,l} = |[\bar{c}_l]_k|$. These simulations $A_{k,l} = 1$.

- The bound depends on the "enhanced SNR"[7] (for a single pitch) or pseudo SNR (PSNR) for the kth pitch[8]

$$\text{PSNR}_k = 10\log_{10}\frac{\sum_{l=1}^{L} A_{k,l}^2 l^2}{\sigma^2}\,. \tag{34}$$

[7] A. Nehorai and B. Porat. "Adaptive comb filtering for harmonic signal enhancement". In: *IEEE Trans. Acoust., Speech, Signal Process."* 34.5 (Oct. 1986), pp. 1124–1138.

[8] M. G. Christensen et al. "Multi-pitch estimation". In: *Signal Processing* 88.4 (Apr. 2008), pp. 972–983.

Simulations III: closely spaced fundamentals

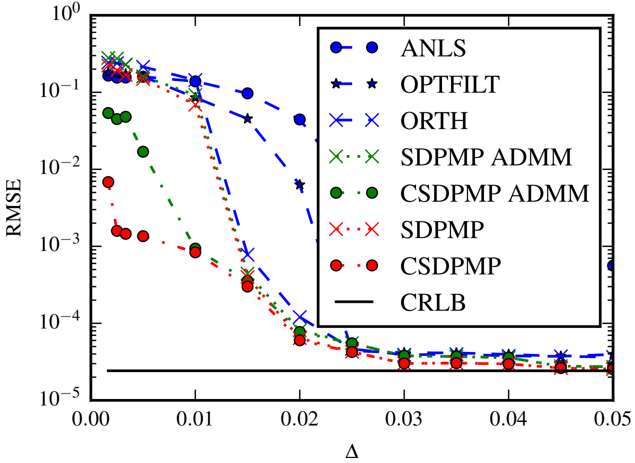


Figure : RMSE as a function of the fundamental frequency difference $\omega_2 - \omega_1 = \Delta$, $K = 2$, $N = 160$, $L = 3$, $\mathrm{PSNR}_1 = \mathrm{PSNR}_2 = 40\,[\mathrm{dB}]$.

Simulations IV: versus PSNR

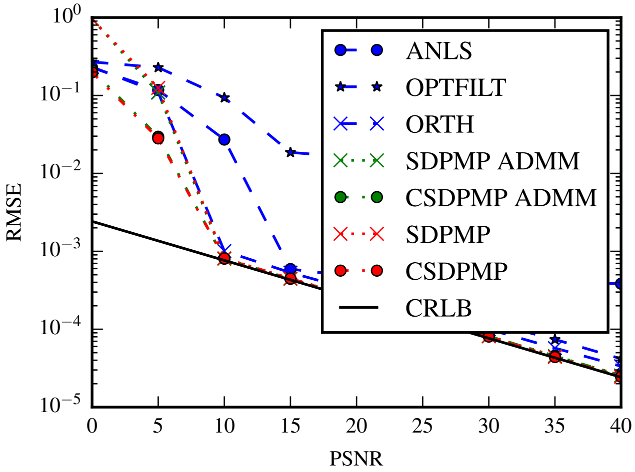


Figure : RMSE as a function of the $\text{PSNR} = \text{PSNR}_1 = \text{PSNR}_2$, $K = 2$, $N = 160$, $L = 3$, and $\omega_1 = 0.1580$, $\omega_2 = 0.6364$.

Summary

Multi-pitch estimation using semidefinite-programming:

- Convex optimization (semidefinite programming (SDP)).
- Gridless (atomic norm/superresolution, numerically: accuracy determined by the underlying method).
- The real-valued model is "easier"/"computational more efficient" compared to the complex-valued model.
- Approximately achieves the CRLB.
- High resolution (separating two fundamentals with almost the same frequency).
- Non-standard representation of LMIs is useful solving the problem using primal-dual interior point methods.