Multi-Pitch Estimation via Semidefinite Programming

August 24, 2016

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

Dept. of Electronic Systems

Aalborg University

AALBORG UNIVERSITY
DENMARK

Agenda

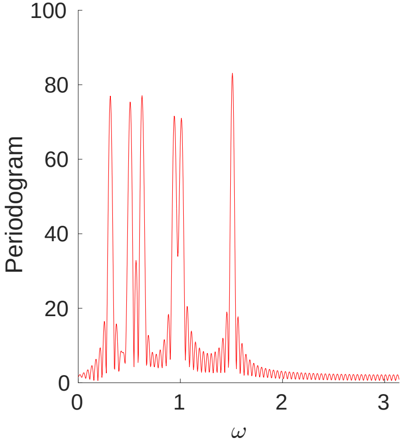
- Multi-pitch estimation.
- Superresolution/gridless/atomic norm using semidefinite programming.
- Bringing it together.
- Complex- and real-valued data.
- Simulations

Multi-pitch estimation I

- Harmonic signals: Fundamental 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, amplitudes (and K)[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

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

- Variables: Atoms $a_k \in \mathbb{C}^n$, coefficients $c_k \in \mathbb{C}^m, k = 1, \dots, r$ 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

$$\mathbb{A}_n = \left\{ \frac{s}{\sqrt{n}} \left[1, \exp(j\omega), \dots, \exp(j(n-1)\omega)\right]^T \right.$$

$$\left. \mid |\omega - \alpha| \leq \beta, |s| = 1, s \in \mathbb{C} \right\} \tag{2}$$

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

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

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

Frequency constraint

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

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

- 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 $F = \begin{bmatrix} 0 & I_{NL-1} \end{bmatrix}$, $G = \begin{bmatrix} I_{NL-1} & 0 \end{bmatrix}$.
- With the selection $\alpha = 0$, $\beta = \pi/L$, (20) is a real semidefinite cone constraint and Toeplitz.

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

Simulations I

- Monte Carlo, $R = 500$ repetitions, known model-order, $K = 2$, $L = 3$, real-valued data otherwise same setup as[4].
- The proposed estimators are implemented with a CVXOPT custom solver[5] based on a non-canonical semidefinite cone representation[6] and an alternating direction method of multipliers with fixed $k = 350$ iterations.
- 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.

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

[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.

[6] 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.

Simulations II

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

$$\mathbf{var}(\hat{\omega}_k) \geq \frac{24\sigma^2}{(N(N^2-1))\sum_{l=1}^{L} A_{k,l}^2 l^2} \tag{21}$$

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]

$$\mathrm{PSNR}_k = 10\log_{10} \frac{\sum_{l=1}^{L} A_{k,l}^2 l^2}{\sigma^2}\,. \tag{22}$$

[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

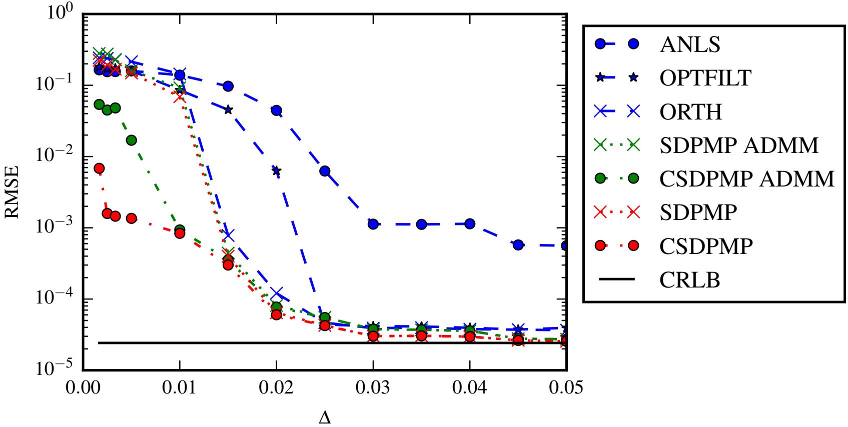


Figure : RMSE as a function of the fundamental frequency difference $\omega_2 - \omega_1 = \Delta$, $K = 2$, $N = 160$, $L = 3$, $\mathrm{PSNR}_1 = \mathrm{PSNR}_2 = 40\,[\mathrm{dB}]$.

Simulations IV: versus PSNR

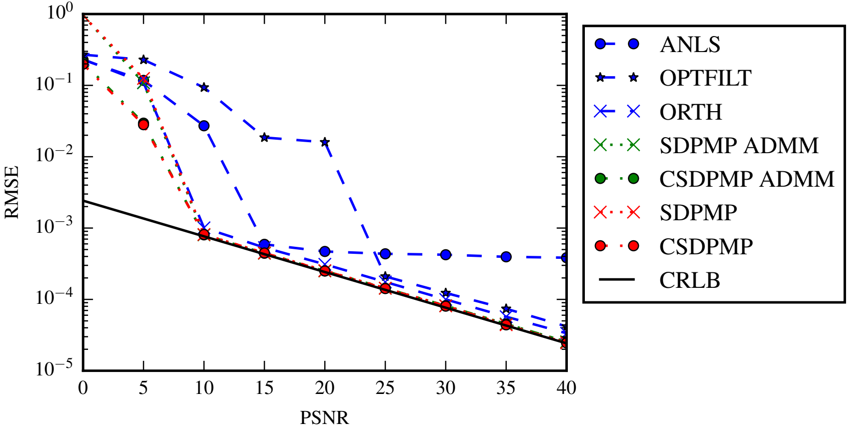


Figure : RMSE as a function of the $\text{PSNR} = \text{PSNR}_1 = \text{PSNR}_2$, $K = 2$, $N = 160$, $L = 3$, and $\omega_1 = 0.1580$, $\omega_2 = 0.6364$.

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 pitches with almost the same frequency).