Resolution Limit for Atomic Decompositions

Gongguo Tang
Colorado School of Mines

SPARS

July 8, 2015

From ℓ_1 Norm to Atomic Norms

- The ℓ_1 norm enforces sparsity w.r.t. the canonical basis
- The nuclear norm enforces sparsity w.r.t. rank-one matrices
- Atomic norm generalizes these two norms and enforces sparsity w.r.t. a general dictionary/atomic set $\mathcal{A} = \{\mathbf{a}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ [Chandrasekaran et. al. 2010]:

$$\|\mathbf{x}\|_{\mathcal{A}} = \inf\{\sum_i |\lambda_i| : \mathbf{x} = \sum_i \lambda_i \mathbf{a}(\boldsymbol{\theta}_i), \boldsymbol{\theta}_i \in \Theta\}$$

- Connection to TV norm minimization: the atomic norm $\|\mathbf{x}\|_{\mathcal{A}}$ is equal to the optimal value of

$$\underset{\mu \in \mathcal{M}(\Theta)}{\text{minimize}} \ \|\mu\|_{\mathrm{TV}}$$
$$\text{subject to } \mathbf{x} = \int_{\Theta} \mathbf{a}(\boldsymbol{\theta}) d\mu(\boldsymbol{\theta})$$

Example Atoms

Example

- Canonical basis vectors $\mathbf{a}(i) = \mathbf{e}_i, i \in [n]$
- Finite collection of vectors $A = [\mathbf{a}_1, \cdots, \mathbf{a}_n]$
- Rank-1 matrices: $\mathbf{a}(\mathbf{u}, \mathbf{v}) = \mathbf{u} \otimes \mathbf{v}$
- Line spectral signals: $\mathbf{a}(f) = [1 \ e^{i2\pi f} \ \cdots \ e^{i2\pi n f}]^T, f \in [0, 1)$.
- High-dimensional line spectral signals
- Spherical harmonics
- Rank-1 tensors: $\mathbf{a}(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}$
- Translation-invariant signals: $\mathbf{a}(\tau) = [h(t_j - \tau)]_{j=1}^n$
- Radar signals: $\mathbf{a}(\tau, \nu) = [\psi(t_j - \tau)e^{i2\pi\nu t_j}]_{j=1}^n$
- Single-pole linear systems: $\mathbf{a}(w) = [\frac{1-|w|^2}{z_j - w}]_{j=1}^n$

Sparse Regularizer

- Given noisy linear measurements $\mathbf{y} = \Phi\mathbf{x}^\star + \mathbf{w}$ of a signal $\mathbf{x}^\star$, which has a sparse representation w.r.t. $\mathcal{A}$, recover $\mathbf{x}^\star$ via [Chandrasekaran et. al.]

$$\underset{\mathbf{x}}{\text{minimize}} \frac{1}{2}\|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \tau\|\mathbf{x}\|_{\mathcal{A}}$$

- To study the performance of the atomic norm regularizers, we'd like to understand $\|\cdot\|_{\mathcal{A}}$ as we do for the ℓ_1 and nuclear norms:
  - $\|\mathbf{x}\|_{\ell_1} = \sum_{i=1}^n |x_i|$ if $\mathbf{x} = \sum_{i=1}^n x_i\mathbf{e}_i$
  - $\|X\|_* = \sum_{i=1}^r \sigma_i$ if $X = \sum_{i=1}^r \sigma_i \mathbf{u}_i \otimes \mathbf{v}_i$ is the SVD
  - $\partial\|\mathbf{x}\|_{\ell_1} = \{\mathbf{z} : z_i = \text{sign}(x_i)$ for $x_i \neq 0; z_i \in [-1, 1]$ otherwise$\}$
  - $\partial\|X\|_* = \{UV + W : \|W\| \leq 1, U^TW = 0, WV^T = 0\}$.

Atomic Decompositions

Definition

We call a finite decomposition $\mathbf{x} = \sum_{i=1}^{r} \lambda_i \mathbf{a}(\boldsymbol{\theta}_i)$ an atomic decomposition if it achieves the atomic norm, i.e., $\|\mathbf{x}\|_{\mathcal{A}} = \sum_{i=1}^{r} |\lambda_i|$.

- The representing measure $\mu^\star = \sum_{i=1}^{r} \lambda_i \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_i) \in \mathcal{M}(\Theta)$ of an atomic decomposition is an optimal solution to the TV norm minimization problem.

Sufficient Conditions

Many sufficient conditions for atomic decompositions have been developed:

- Finite dictionary: restricted isometry property [Candès, Romberg, Tao, 2004]
- Line spectral signals: separation of frequencies [Candès, Fernandez-Granda, 2012]
- Rank-1 tensors: incoherence of the factors [Tang, Shah 2015]
- Translation invariant signals: separation of translations [Tang, Recht 2013; Bendory, Dekel, Feuer 2014]
- Spherical harmonics: separation of parameters [Bendory, Dekel, Feuer 2014]
- Radar signals: separation of time-frequency shifts [Heckel, Morgenshtern, and Soltanolkotabi, 2015]

Sufficient Conditions - 2

- Completion/Recovery: [Tang, Bhaskar, Shah, Recht, 2012], [Chi, Chen, 2015]
- Denoising: [Candès, Fernandez-Granda, 2012], [Bhaskar, Tang, Recht, 2012], [Tang, Bhaskar, Recht, 2015]
- Support recovery/parameter estimation: [Fernandez-Granda, 2013], [Duval, Peyré, 2014], [Denoyelle, Duval, Peyré, 2015],
- Effect of griding: [Tang, Bhaskar, Recht, 2013], [Duval, Peyré, 2015]

Questions

- Is certain separation in parameters also necessary for a decomposition to be an atomic decomposition?
- Does TV norm minimization have a resolution limit?

Outline

- Line Spectral Estimation
- Symmetric Tensor Decomposition
- Why is there a resolution limit?

Line Spectral Signals

- The atomic norm of $\mathbf{x}$ w.r.t. the atomic set

$$\mathcal{A} = \{\mathbf{a}(f) = \begin{bmatrix} 1 \\ e^{\mathrm{i}2\pi f} \\ e^{\mathrm{i}2\pi 2f} \\ \vdots \\ e^{\mathrm{i}2\pi nf} \end{bmatrix} : f \in [0,1]\}$$

- Computation of $\|\mathbf{x}\|_{\mathcal{A}}$ can be reformulated as an SDP [Bhaskar, Tang, Recht 2012].

Line Spectral Signals - 2

Theorem (Candès & Fernandez-Granda 2012)

A decomposition $\sum_i c_i \mathbf{a}(f_i)$ *is an atomic decomposition if*

$$\Delta = \min_{i \neq j} |f_i - f_j| > \frac{4}{n}$$

regardless the sign pattern of $\{c_i\}$*.*

Theorem (Tang 2015)

If a decomposition $\sum_i c_i \mathbf{a}(f_i)$ *is an atomic decomposition regardless the sign pattern of* $\{c_i\}$*, we must have*

$$\Delta = \min_{i \neq j} |f_i - f_j| \geq \frac{1}{n\pi}.$$

Symmetric Tensor Decomposition

- Symmetric tensor atoms

$$\mathcal{A} = \{\mathbf{a}(\mathbf{x}) = \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x} : \mathbf{x} \in \mathbb{S}^{n-1}\}$$

- The tensor nuclear norm

$$\|T\|_* = \inf\{\sum_j c_j : T = \sum_j c_j \mathbf{x}_j \otimes \mathbf{x}_j \otimes \mathbf{x}_j, c_j > 0, \mathbf{x}_j \in \mathbb{S}^{n-1}\}$$

- An equivalent definition:

$$\|T\|_* = \inf\{\|\mu\|_{\mathrm{TV}} : T = \int_{\mathbb{S}^{n-1}} \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x} d\mu, \mu \in \mathcal{M}(\mathbb{S}^{n-1})\}$$

- Best regularizer in low-rank tensor completion, recovery, and denoising.

Symmetric Tensor Decomposition - 2

- NP hard to compute in the worst case.
- Approximate using the Lasserre/SOS hierarchy ($d \geq 2$)

$$
\begin{aligned}
\underset{\mathbf{m}_{2d}}{\text{minimize}} \; & \mathbf{m}_{2d}(0) \\
\text{subject to} \; & P_3(\mathbf{m}_{2d}) = \operatorname{svec}(T) \\
& M(\mathbf{m}_{2d}) \succcurlyeq 0 \\
& L(\mathbf{m}_{2d}) = 0.
\end{aligned}
$$

Symmetric Tensor Decomposition - 3

Theorem (Tang 2015)

For $\sum_{i=1}^{r} \lambda_i \mathbf{x}_i \otimes \mathbf{x}_i \otimes \mathbf{x}_i$ *to be atomic tensor decompositions regardless the sign pattern of* $\{\lambda_i\}$*, we must have*

$$\Delta = \min_{i \neq j} \arccos(|\langle \mathbf{x}_i, \mathbf{x}_j \rangle|) \geq \frac{2}{3}.$$

Theorem (Tang, Shah 2015)

Denote $X = [\mathbf{x}_1, \ldots, \mathbf{x}_r]$*. If*

$$\|X'X - I_r\| \leq 0.0016.$$

then $\sum_{i=1}^{r} \lambda_i \mathbf{x}_i \otimes \mathbf{x}_i \otimes \mathbf{x}_i$ *is an atomic decomposition regardless the sign pattern of* $\{\lambda_i\}$*.*

Symmetric Tensor Decomposition - 4

Theorem (Tang, Shah 2015)

If

$$\|X'X - I_r\| \leq 0.0016.$$

then the smallest $(d = 2)$ SDP in the Lasserre hierarchy is exact.

- The resolution limit condition can be (easily) extended to higher-order and/or non-symmetric tensors.
- The sufficient results are also likely to be extended to these cases (but much harder).
- Use the relaxation norm for tensor completion, denoising, and robust principal component analysis.

Why is there a resolution limit?

- Using similar atoms to represent a signal is not economical in the ℓ_1 norm sense.
- The dual problem is

$$\underset{\mathbf{q}}{\text{maximize}} \ \langle \mathbf{q}, \mathbf{y} \rangle \ \text{ subject to } \ \underbrace{\sup_{\boldsymbol{\theta}} |\langle \mathbf{q}, \mathbf{a}(\boldsymbol{\theta}) \rangle|}_{\|\mathbf{q}\|_{\mathcal{A}}^{*}} \leq 1.$$

Why is there a resolution limit? - 2

Dual certificate

Suppose strong duality holds, then $\sum_{j=1}^{r} c_j \mathbf{a}(\boldsymbol{\theta}_j)$ is an atomic decomposition iff there exists a dual "polynomial"

$$Q(\boldsymbol{\theta}) := \langle \mathbf{q}, \mathbf{a}(\boldsymbol{\theta}) \rangle = \sum_i q_i a_i^*(\boldsymbol{\theta})$$

such that

$$Q(\boldsymbol{\theta}_j) = \text{sign}(c_j), \forall j$$
$$|Q(\boldsymbol{\theta})| \leq 1, \forall \boldsymbol{\theta} \in \Theta.$$

- Ex: For line spectral signals, the dual polynomial is a trigonometric polynomial $Q(f) = \sum_k q_k e^{-\mathrm{i}2\pi k f}$.
- Ex: For symmetric tensors, the dual polynomial is a third order polynomial $Q(\mathbf{x}) = \sum_{i,j,k} q_{ijk} x_i x_j x_k$.

Why is there a resolution limit? - 3

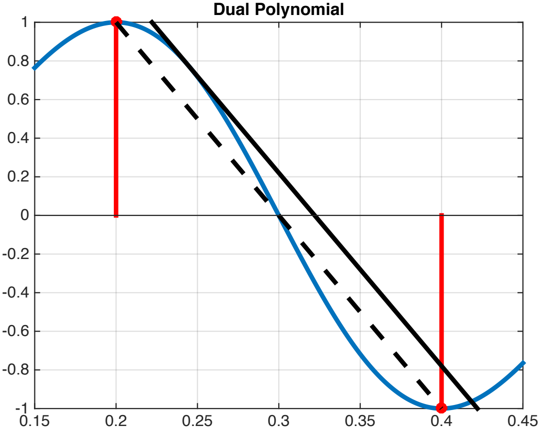
- To simultaneously interpolate $\mathrm{sign}(c_i) = +1$ and $\mathrm{sign}(c_j) = -1$ at $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ respectively while remain bounded imposes constraints on the derivative of $Q(\boldsymbol{\theta})$:

$$\|\nabla Q(\hat{\boldsymbol{\theta}})\|_2 \geq \frac{|Q(\boldsymbol{\theta}_i) - Q(\boldsymbol{\theta}_j)|}{\Delta_{i,j}} = \frac{2}{\Delta_{i,j}}$$

Why is there a resolution limit? - 4

- For $\Theta \subset \mathbb{R}$, there exists $\hat{f} \in (f_i, f_j)$ such that

$$Q'(\hat{f}) = 2/(f_j - f_i)$$



Why is there a resolution limit? - 5

- For certain classes of functions $\mathcal{F}$, if the function values are uniformly bounded by 1, this limits the maximal achievable derivative, i.e.,

$$\sup_{g \in \mathcal{F}} \frac{\|g'\|_\infty}{\|g\|_\infty} < \infty.$$

- For $\mathcal{F} = \{$trigonometric polynomials of degree at most $n\}$,

$$\|g'(f)\|_\infty \leq 2\pi n \|g(f)\|_\infty.$$

- This is the classical Markov-Bernstein's inequality.

Sign Pattern

- Sign pattern of $\{c_j\}$ plays a big role. The argument breaks down if, e.g., all c_j are positive.

Theorem

[a] *Suppose the atom components $\{a_i(t)\}_{i=0}^n$ form a Chebyshev system on $[a, b]$. Define $\omega(t) = 2$ if $t \notin \{a, b\}$ and 1 otherwise. A decomposition $\mathbf{y} = \sum_j c_j \mathbf{a}(t_j)$ with $c_j > 0$ and $\sum_j \omega(t_j) \leq n$ is unique.*

[a] [de Castro, Gamboa 2011; Denoyelle, Duval, Peyre 2015; Bendory, 2015; Morgenshtern, Candès 2015; Tang 2015; Schiebinger, Robeva, Recht 2015]

- Chebyshev system: no non-trivial "polynomial" $\sum_{i=0}^n c_i a_i(t)$ has more than n distinct zeros.

Sign Pattern - 2

Example (Chebyshev systems)

- algebraic polynomials: $\{t^i\}_{i=0}^n$ on any interval.
- trigonometric polynomials: $\{1, \sin(k\theta), \cos(k\theta)\}_{k=1}^n$ on $[0, 2\pi)$.
- rational functions: $\{\frac{1}{s_i+t}\}_{i=0}^n$ with $s_i > 0$ on $(0, \infty)$.
- exponentials: $\{e^{\alpha_i t}\}_{i=0}^n$ with $\alpha_i > 0$ on any interval.
- Gaussian functions: $\{e^{-(s_j-t)^2}\}$ for $s_j > 0$ on $(-\infty, \infty)$.
- Totally positive kernels: A continuous function $G(t, s)$ defined on $[a, b] \times [c, d]$ is called a totally positive kernel if for any n and any points $(a \leq)t_0 < t_1 < \cdots < t_n(\leq b), (c \leq)s_0 < s_1 < \cdots < s_n(\leq d)$, the determinant $\det([G(t_j, s_k)]_{j,k=0}^n) > 0$. The system $\{\phi_k(t) = G(t, s_k)\}_{k=0}^n$ is a Chebyshev system if $G(t, s)$ is totally positive.

Conclusions

- Atomic norm can only be achieved by decompositions involving incoherent or well-separated atoms.
- TV norm minimization has a limit in resolving parameters.
- Positive combination of atoms typically requires no separation condition (when there is zero noise).
- Connections to stability. [Moitra 2014]