Geometry in the sparse methods

Zhiyue Huang and Paul Marriott

Department of Statistics and Actuarial Science
University of Waterloo

April 6, 2011

Outline

Signal representation

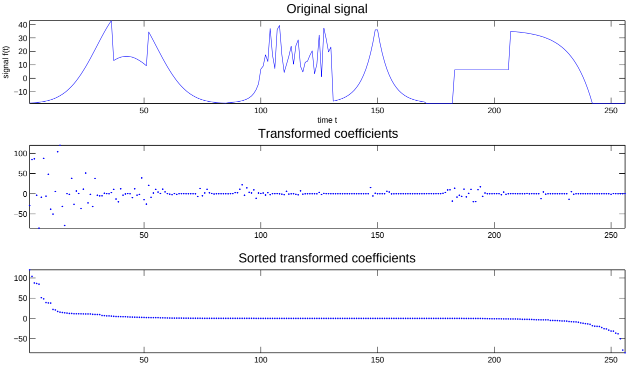
- Represent $f(t)$ as linear combination of basis elements

$$f(t) = \sum_i \alpha_i \psi_i(t) \quad \text{or} \quad f = \Psi\alpha \qquad (1)$$

- $\{\psi_i(t)\}$ could be sinusoids, wavelets, curvelets...
- The coefficient sequence α

$$\alpha_i = \langle f, \psi_i \rangle \quad \text{or} \quad \alpha = \Psi^T f. \qquad (2)$$

Sparsity


Original signal
signal f(t)
time t
Transformed coefficients
Sorted transformed coefficients

Basis pursuit [Chen et al., 1999]

- Consider underdetermined linear equation

$$\Psi_{n\times d}\beta_{d\times 1} = y_{n\times 1}, \tag{3}$$

where Ψ_j, the jth column of Ψ, is basis spanning the space of y and $d \gg n$.

- Basis pursuit

$$\min \|\beta\|_{\ell_1} \quad \text{subject to } \Psi\beta = y \tag{4}$$

can provide sparse solution β.

Compressed sensing [Donoho, 2006]

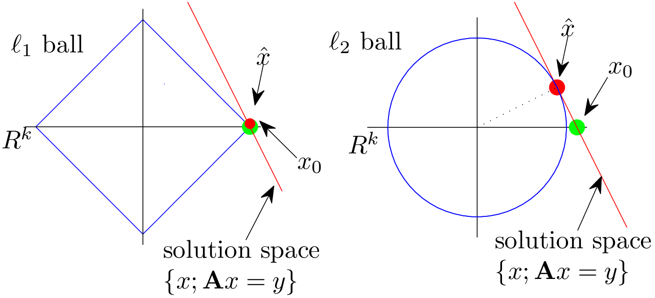
- x_0 is k-sparse in $\mathbb{R}^n$, a random matrix **A** obeys RIP of order $2k$.
- A vector is said to be k-sparse if it has at most k nonzero entries.
- x_0 can be compressed to $\mathbf{A}x_0$ in $\mathbb{R}^d$, $d = O(k \log n)$.
- x_0 can be recovered exactly by

$$\min ||x||_{\ell_1} \quad \text{subject to } \mathbf{A}x = y.$$

Outline

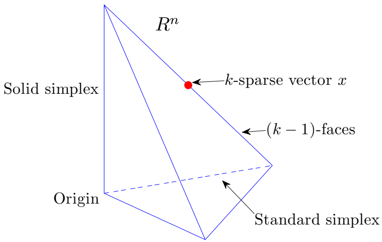
ℓ_1 minimization in basis pursuit

- $\min ||x||_{\ell_1}$ subject to $y = \mathbf{A}x$



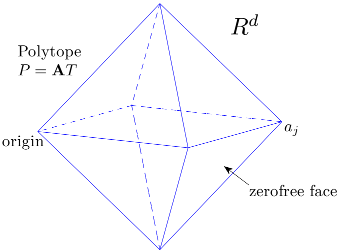
Outline

Solid simplex



- **Standard simplex** T^{n-1}: the convex hull of the unit basis vectors.
- **Solid simplex** T_0^n: the convex hull of T^{n-1} and the origin.
- T^{n-1} is the **outward** part of T_0^n.

Polytopes



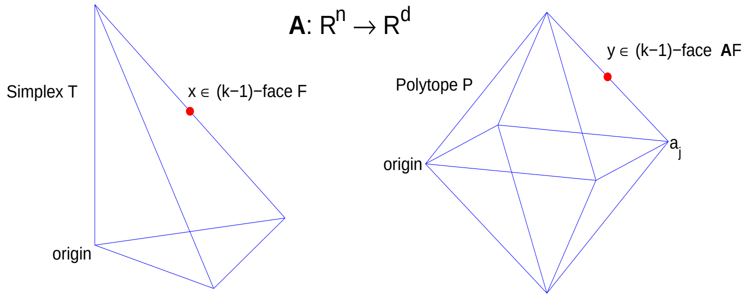
- $P = \mathbf{A}T_0^n$ is a convex polytope in $\mathbb{R}^d$, where $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^d$.
- $P = conv(\{0\} \cup \{a_j\}_{j=1}^n)$, where a_j is the jth column of $\mathbf{A}$.

k-neighborly polytope

- ***k*-neighborly ploytope**:
  every set of k vertices spans a $(k-1)$-face of P.
- **outwardly *k*-neighborly polytope**:
  every set of k vertices not including the origin spans a $(k-1)$-face of P.

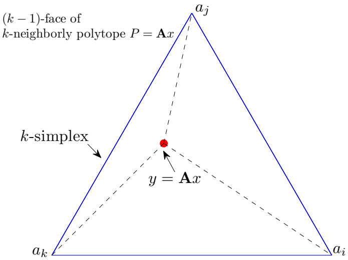
Neighborliness [Donoho and Tanner, 2005]

Suppose the polytope $P = \mathbf{A}T$ has N vertices and outwardly k-neighborly, if and only if, for all $\ell = 0, \cdots, k-1$, and for all ℓ-dimensional faces F of T^{n-1}, $\mathbf{A}F$ is a ℓ-dimensional face of P.



Unique representation [Donoho and Tanner, 2005]

- For k-neighborly polytopes, every low-dimensional face is a simplex.
- Each point y on the face has a unique convex combination by the vertices of the simplex.



Equivalent properties [Donoho and Tanner, 2005]

Let $\mathbf{A} \in \mathbb{R}^{d \times n}$, $d < n$. The two properties of $\mathbf{A}$ are equivalent

1. The polytope P is outwardly k-neighborly.
2. Whenever $y = \mathbf{A}x$ has a nonnegative solution x_0 having at most k nonzeros, x_0 is unique nonnegative solution to ℓ_1 minimization problem.

Outline

Restricted isometry property

- $\mathbf{A} \in \mathbb{R}^{d \times n}$ obeys a RIP for sets of size k if

$$(1-\delta_k)\frac{d}{n}\|\beta\|_{\ell_2}^2 \leq \|\mathbf{A}\beta\|_{\ell_2}^2 \leq (1+\delta_k)\frac{d}{n}\|\beta\|_{\ell_2}^2$$

  for every k-sparse vector β.
- RIP preserve all k and lower dimensional planes from $\mathbb{R}^n$ to $\mathbb{R}^d$.

Randomness and RIP [Baraniuk et al., 2008]

- **A** obeys RIP for

$$k \leq Cd/\log(n/d+1), \tag{5}$$

with high probability, when
  - **A** is random Gaussian matrix.
  - **A** is random binary matrix.

Outline

Local mixture model

- Given observed data $\mathcal{X} : x_1, \cdots, x_N$ with $t_1, \cdots, t_n$ distinct values.
- According to [Lindsay, 1995], embedding the likelihood of mixture model in $\mathbb{R}^n$ with -1-affine geometry.
- The likelihood of mixture model is

$$f(t;\theta,Q) = f(t;\theta) + \sum_{j=1}^{n} \beta_j s_j(t), \tag{6}$$

where $s_j(t)$ are the basis generated from data.

Robust local mixture model

- $s_j(t)$ can be chosen to be sparse for robustness of the model.
- For each t_i in observed data,

$$f(t_i; \theta, Q) = f(t_i; \theta) + \sum_{j \in \Omega_i} \beta_j s_j(t_i) \tag{7}$$

- For example, one $x = t_i$ is added to $\mathcal{X}$, it will not affect the inference on β_j, $j \notin \Omega_i$.
- Sparse basis $s_j(t)$ can be obtained by sparse methods.

Outline

Statistical model space

- Consider a family of statistical models, $\mathcal{M}_n$, whose likelihood can be embedded in $\mathbb{R}^n$ space.
- The true model locates in $\mathbb{R}^k$, $k \ll n$.
- By random orthogonal matrix **A** with RIP, $\mathcal{M}_n$ can be projected into $\mathbb{R}^m$, where model sufficiency is kept.
- What is the statistical inference property on such $\mathbb{R}^m$?

Outline

Summary

- ℓ_1 minimization.
- Sparse solution in ℓ_1 minimization.
- Outwardly k-neighborly polytope.
- RIP and randomness.
- Application to robust mixture models.
- Future work on statistical model space.

Reference I

R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008. ISSN 0176-4276.

S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1999. ISSN 1064-8275.

D.L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006. ISSN 0018-9448.

D.L. Donoho and J. Tanner. Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9446, 2005.

Reference II

B.G. Lindsay. Mixture models: theory, geometry and applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics*, pages 67–163. JSTOR, 1995.