Journal Watch
IEEE Transactions on Information Theory (February, 2017)

Lekshmi Ramesh

Signal Processing for Communications Lab
IISc, Bangalore

April 22, 2017

- Goal: Given p samples from $\mathbb{R}^n$, $Y = [y_1, \ldots, y_p]$, find a concise representation for these samples. That is, find:
  a complete (square and invertible) matrix $A \in \mathbb{R}^{n \times n}$,
  and a sparse coefficient matrix $X \in \mathbb{R}^{n \times p}$,
  such that $Y \approx AX$ when $n < p$.

- Usual approach

$$\min_{A,X} \|Y - AX\|_F^2 + \lambda \|X\|_1$$
$$\text{s.t. } A \in \mathcal{A}$$

  - Objective non convex in A, X; $\mathcal{A}$ typically non convex too
  - For a permutation matrix Π and a diagonal matrix Σ with diagonal entries in $\{+1, -1\}$:
    (A, X) and $(A\Pi\Sigma, \Sigma^{-1}\Pi^\top X)$ result in the same objective value: combinatorially many global minima

- A different formulation:
  - Rowspace(Y) = Rowspace(X), rows of X are sparse vectors in the known subspace Rowspace(Y)
  - First recover rows of X, then recover A

$$\min_q \ \|q^\top Y\|_0 \ \text{ s.t. } \ q^\top Y \neq 0$$

  - Replace above formulation with a convex objective and a spherical constraint

$$\min_q \ \frac{1}{p}\sum_{i=1}^{p} h_\mu(q^\top y_i) \ \text{ s.t. } \ \|q\|_2 = 1,$$

    where h_μ: a convex, smooth appproximation to $|.|$

- Contributions
  - Geometric characterization of the objective, explanation for the effectiveness of non convex heuristics
  - First efficient algorithm that provably recovers A, X where X can have $O(n)$ non zeros per column
  - Under the assumption that $X_{ij} = \Omega_{ij} V_{ij}$, with $\Omega_{ij} \sim Ber(\theta)$ and $V_{ij} \sim \mathcal{N}(0,1)$ (denoted $X_0 \overset{iid}{\sim} BG(\theta)$):

    For $\theta \in (0, \frac{1}{3})$, given $Y = A_0 X_0$ with A_0 a complete dictionary and $X_0 \overset{iid}{\sim} BG(\theta)$, there exists a polynomial-time algorithm that recovers A_0 and X_0 (upto sign, scale and permutation) with high probability.

- Setup
  - $\theta \in \mathbb{R}^k$ containing locations of k objects, $k \geq 1$ is known
  - Choose subsets A_i of $\mathbb{R}$, query the number of objects in each subset and obtain a sequence $\{X_i\}$ of noiseless answers
  - Formally, for the n^{th} question A_n, the answer X_n is

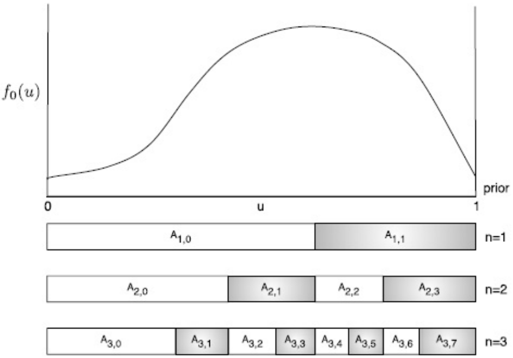
$$X_n = \mathbb{1}_{A_n}(\theta_1) + \ldots + \mathbb{1}_{A_n}(\theta_k)$$

  - Bayesian setting: $\theta_i \stackrel{iid}{\sim} f_0$ with joint density $p_0 = \Pi_{i=1}^k f_0(\theta_i)$

- Goal: Devise a method for choosing questions so that θ can be found as accurately as possible form a finite budget of questions (accuracy measured in terms of entropy of posterior distribution of θ)

- Prior work
  - $k = 1$ case: noiseless, noisy, Bayesian setting considered
  - $k > 1$ case: Group testing: "Is $A \cap S \neq \phi$?"
    Subset guessing: "Is $S \subset A$?"
    Binary answers, non-Bayesian setting
- Contributions
  - A non-adaptive dyadic policy and an adaptive greedy policy for noiseless group testing under sum observations.
    Both algorithms based on minimizing the expected entropy of the posterior on θ.
  - Dyadic policy: shown to be optimal among non adaptive policies
    Greedy policy: at least as good as dyadic policy, strictly better in some cases.

The dyadic policy



Prior density f_0 with support $[0, 1]$. The question set A_n is the union of the shaded subsets.

- Setup
  - Set of N variables/features $X_1, \ldots, X_N$, outcome Y (both known)
  - Only k variables, indexed by $S \subset [N]$ (unknown), relevant for predicting outcome Y
  - Latent random quantity β_S affecting observations

$$P(Y|X, \beta_S, S) = P(Y|X_S, \beta_S, S)$$

- Goal: Given T sample pairs $\{X_i, Y_i\}_{i=1}^T$, observation model $P(Y|X_S, \beta_S, S)$ and prior $p(\beta_S)$, find necessary and sufficient conditions on T in order to recover S with arbitrarily small error probability

- Contributions
  - Necessary and sufficient conditions on T for various sparsity models (sparse linear regression, binary regression, group testing, models with missing data)
  - Results for both linear and non-linear models in a unifying manner

Other interesting papers

- Sensing Tensors With Gaussian Filters. *S. Chrétien and T. Wei*
- Blind Recovery of Sparse Signals From Subsampled Convolution. *K. Lee, Y. Li, M. Junge, and Y. Bresler*
- Compressive Sampling Using Annihilating Filter-Based Low-Rank Interpolation. *J. C. Ye, J. M. Kim, K. H. Jin, and K. Lee*