Journal Watch:

IEEE TIT, March, 2017 and IEEE TSP, June 15, 2017

Saurabh Khanna,
Signal Processing for Communication Lab, ECE, IISc

High Dimensional Estimation of Structured Signals From Non-Linear Observations With General Convex Loss Functions

Authors: Martin Genzel

- Recovery of structured signals from a convex set $K \subset \mathbb{R}^n$.
- Non-linear observations:

$$y_i := f(\langle \mathbf{a}_i, \mathbf{x}_0 \rangle) + \epsilon, \quad i = 1, \ldots, m,$$

  where $\mathbf{a}_i \sim \mathcal{N}(0, \mathbf{\Sigma})$ are independent mean-zero Gaussian vectors.
- f plays the role of non-linearity.
- **Key Idea:** Non linearity treated as noise which disturbs a linear measurement process.
- Generalized estimator:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \frac{1}{2m} \sum_{i=1}^{m} (\tilde{y}_i - \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle)^2 \quad \text{s.t. } \boldsymbol{x} \in K.$$

- **Question:** Can the above estimator still work? (stably recover $\mathbf{x}_0$?)
- Answer depends on quantities like:.
  - Global Gaussian width of a set K: $w(K) := \mathbb{E}[\sup_{\boldsymbol{x} \in K} \langle \boldsymbol{g}, \boldsymbol{x} \rangle], \quad g \sim \mathcal{N}(0, I)$.
  - Local Gaussian width of a set K: $w_t(L) := w(L \cap tB_2^n) = \mathbb{E}[\sup_{\boldsymbol{x} \in L \cap tB_2^n} \langle \boldsymbol{g}, \boldsymbol{x} \rangle], \quad t \geq 0$.

High Dimensional Estimation of Structured Signals From Non-Linear Observations With General Convex Loss Functions Authors: Martin Genzel

- and answer also depends on entities like...

$$\mu := \mathbb{E}[f(g)\cdot g],$$
$$\sigma^2 := \mathbb{E}[(f(g)-\mu g)^2],$$
$$\eta^2 := \mathbb{E}[(f(g)-\mu g)^2\cdot g^2]$$

  μ models the correlation between linear and non-linaer model. σ^2, η^2 model the variance between them.

- Also, the loss function $\mathcal{L}$ should be well beahaved (continuously diffrentiable, $\mathcal{L}'$ Lipschitz cont., convex in first argument).

  If $m > C.w(K-\mu\mathbf{x}_0)^2$, then

$$\|\hat{\mathbf{x}}-\mu\mathbf{x}_0\| \le C_{\sigma,\eta}\cdot\left(\left(\frac{w(K-\mu\mathbf{x}_0)^2}{m}\right)^{1/4}+\varepsilon\right).$$

- An improved result: If $m > w_t(K-\mu\mathbf{x}_0)^2$, then

$$\|\hat{\mathbf{x}}-\mu\mathbf{x}_0\| \le C\cdot\left(\frac{\sigma\cdot\sqrt{d_t(K-\mu\mathbf{x}_0)}+\eta}{\sqrt{m}}+\varepsilon\right).$$

A Nonconvex Splitting Method for Symmetric Nonnegative Matrix Factorization: Convergence Analysis and Optimality

Authors: Songtao Lu et al., Iowa St. Univ.

- Symmetrix Nonnegative Matrix Factorization (SymNMF):
  Decompose given PSD matrix $\mathbf{Z}$ as $\mathbf{Z} = \mathbf{X}\mathbf{X}^T$ where $\mathbf{X}$ is componentwise nonnegative.
- Proposes algorithm with is capable of convergence to the set of KKT points with a provable global convergence rate.
- Key idea is to relax the symmetry constraint and enforce it at a slower rate.
- Conventional approach:

$$\min_{\mathbf{X}=\mathbf{Y},\mathbf{Y}\geq 0} \quad \frac{1}{2}||\mathbf{X} - \mathbf{X}\mathbf{Y}^T||_F^2$$

- Reformulated problem:

$$\min_{\mathbf{X},\mathbf{Y}} \quad \frac{1}{2}||\mathbf{X} - \mathbf{X}\mathbf{Y}^T||_F^2 \quad \text{s.t. } \mathbf{Y} \geq 0, \mathbf{X} = \mathbf{Y}, ||\mathbf{Y}_i||_2^2 \leq \tau, \forall i.$$

- For τ sufficiently large (depending on $\mathbf{Z}$), the above two problems have identical KKT points !.
- Proposes a primal dual algorithm for the reformulated problem with an additional proximal penalty with variable stepsize. This results in vastly improved convergence guarantees.

The β model - Maximum likelihood, Cramér Rao Bounds, and Hypothesis Testing

Authors: Johan Walhstrom, Arye Nehorai and others

- β-models belong to the class of exponential random graph models (ERGMs or p^* models).
- Random graph models are useful in analysis work (realistic, tractable models are need of the hour).
- ERGMs are popular as their probability distribution can be specified in terms of graph attributes such as max degree, etc.
- In practical problems, random graphs must reflect/incorporate some exogenous information.

  Exohenous information available in two forms:
  - Nodal covariates - gender, status etc.
  - Dyadic covariates - age difference, spatial distance

  Random graphs with nodal covariates are modelled as stochastic block models.
- β model is capable of incorporating covariates on a graph-level. This allows us to perform regression with a random graph as the dependent variable.

The β model - Maximum likelihood, Cramér Rao Bounds, and Hypothesis Testing

Authors: Johan Walhstrom, Arye Nehorai and others

- The undirected β model

$$p_{ij} = \frac{e^{\beta_i + \beta_j}}{1 + e^{\beta_i + \beta_j}}$$

- The directed β model

$$p_{ij} = \frac{e^{\alpha_i + \beta_j}}{1 + e^{\alpha_i + \beta_j}}$$

- The covariate based β model

$$p_{ij}(\mathbf{x}) = \frac{e^{\boldsymbol{\alpha}_i^\mathsf{T}\mathbf{x} + \boldsymbol{\beta}_j^\mathsf{T}\mathbf{x}}}{1 + e^{\boldsymbol{\alpha}_i^\mathsf{T}\mathbf{x} + \boldsymbol{\beta}_j^\mathsf{T}\mathbf{x}}}$$

  The probability depends on both the $2K$ regression coefficients α_i and β_j, and the K covariates $\mathbf{x} = [x_1, \ldots, x_K]^T$, representing e.g., time, space, or other variables describing the state of the network. Here, $\alpha_{i,k}$ and $\beta_{i,k}$ describe the effect that the kth covariate has on the tendency of the ith node to form edges with other nodes.

- Cramér Rao bounds are derived and hypothesis testing framework is set up for all the above models.

Quantitative Recovery Conditions for Tree-Based Compressive Sensing

Authors: Coralia Cartis and Andrew Thompson,
Univ. of Oxford, U.K.

- Wavelet representations have a multiscale tree structure. (wavelet coefficients have tree like nested sparsity structure)
- Derives explicit sufficient conditions for exact and stable recovery structured signal from noisless and noisy compressive linear measurements.
- Decoder is Iterative Tree Projection (ITP) algorithm:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}||\mathbf{y}-\mathbf{A}\mathbf{x}||_2^2 \quad \text{subject to } supp(\mathbf{x}) \in T_k,$$

  where T_k is the set of all k-tree sparse vectors in $\mathbb{R}^n$.
- Surprisingly or not surprisingly, we need to only ensure that

$$\frac{k}{m} \to \rho \quad \text{as } (k, m, n) \to \infty.$$

  if the measurement map $\mathbf{A}$ is Gaussian distributed.
- Tree based RIP, RIP constants are defined and characterized.

Other Interesting Papers:

- Invariant Adaptive Detection of Range-Spread Targets Under Structured Noise Covariance
- Proximity Without Consensus in Online Multiagent Optimization
- On Spectrum Sensing of OFDM Signals at Low SNR: New Detectors and Asymptotic Performance
- Effective Low-Complexity Optimization Methods for Joint Phase Noise and Channel Estimation in OFDM
- Nonadaptive Group Testing Based on Sparse Pooling Graphs
- Adaptive Compressed Sensing for Support Recovery of Structured Sparse Sets
- On the Limitation of Spectral Methods: From the Gaussian Hidden Clique Problem to Rank One Perturbations of Gaussian Tensors
- Information-Theoretic Lower Bounds on Bayes Risk in Decentralized Estimation