Journal Watch
IEEE Transactions on Signal Processing
August 15 2019

Chirag Ramesh
SPC Lab, Indian Institute of Science

September 7, 2019

One-Bit Compressive Sensing via Schur-Concave Function Minimization

Contributions

- Notion of Schur-concavity and its application to 1-bit CS
- Recovery model for minimizing ℓ_1 Shannon entropy function (SEF) with Schur-concave functions

System Model

- 1-bit CS problem: $\mathbf{y} = \text{sign}(\mathbf{\Phi x}) \in \{1, -1\}^M$

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) := \|\mathbf{x}\|_0 \\ \text{s.t.} \quad & \mathbf{y} = \text{sign}(\mathbf{\Phi x}) \quad \equiv \quad \text{diag}(\mathbf{y})\mathbf{\Phi x} \succeq \mathbf{0} \\ & \|\mathbf{x}\|_2 = 1 \end{aligned}$$

What is Schur-concavity?

- $f : \mathbb{R}^N \to \mathbb{R}$ is Schur-concave if $\mathbf{x} \prec \mathbf{y} \Rightarrow f(\mathbf{x}) \geq f(\mathbf{y})$
- Here, '$\prec$' means $\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}, \quad k = 1, \ldots, N-1$

Why use ℓ_1-SEF?

$$\psi_{\ell_1\text{-SEF}}(\mathbf{x}) := -\sum_{i=1}^{N} \frac{|x_i|}{\|\mathbf{x}\|_1} \log \frac{|x_i|}{\|\mathbf{x}\|_1}$$

- SEF is scale invariant, Schur-concave and concave wrt $|x|$

$$\langle \nabla \psi_{\ell_1\text{-SEF}}(|\mathbf{x}|), |\mathbf{x}| \rangle = 0$$

$$\nabla \psi_{\ell_1\text{-SEF}}(|x_i|)(|x_i| - \mu) \leq 0, \; \mu = \exp\left(\left(\sum_i |x_i| \log |x_i|\right) / \|\mathbf{x}\|_1\right)$$

- 1-bit CS problem reframed as:

$$\min_{\|\mathbf{x}\|_2=1} \quad \psi_{\ell_1\text{-SEF}}(|\mathbf{x}|) + \lambda \sum_{i=1}^{M} \max\{0, -[\text{diag}(\mathbf{y})\mathbf{\Phi x}]_i\}$$

- Algorithm proposed to solve the above by constructing surrogate functions and minimizing them
- Results compared with other 1-bit penalty functions such as Gaussian entropy, sorted ℓ_1 penalty

Stochastic Successive Convex Approximation for Non-Convex Constrained Stochastic Optimization

Contributions

- A general SSCA method and its convergence proof

$$\min_{\mathbf{x}\in\mathcal{X}} f_0(\mathbf{x}) \triangleq \mathbb{E}\left[g_0(\mathbf{x},\eta)\right]$$
$$s.t.\ f_i(\mathbf{x}) \triangleq \mathbb{E}\left[g_i(\mathbf{x},\eta)\right] \leq 0, i = 1,\ldots,m$$

- Parallel constrained SSCA

Novelty

- Current works use stochastic gradient/ stochastic MM/ SSCA with deterministic feasible region
- Parallel SSCA can help in large scale optimization

Applications

- MIMO transmit signal design with imperfect CSI

$$\min_{\{\boldsymbol{Q}_k \succeq \boldsymbol{0}\}} \sum_{k=1}^{K} Tr\left(\boldsymbol{Q}_k\right) \;\; s.t. \; \mathbb{E}\left[\log\left(1 + \frac{\boldsymbol{h}_k^H \boldsymbol{Q}_k \boldsymbol{h}_k}{\sum_{j \neq k} \boldsymbol{h}_k^H \boldsymbol{Q}_j \boldsymbol{h}_k + \sigma_k^2}\right)\right] \geq r_k$$

- MIMO robust beamforming design

$$\min_{\{\boldsymbol{w}_k\}} \sum_{k=1}^{K} \|\boldsymbol{w}_k\|^2 \;\; s.t. \;\; \Pr\left[SINR_k \triangleq \frac{\left|\boldsymbol{h}_k^H \boldsymbol{w}_k\right|^2}{\sum_{i \neq k} \left|\boldsymbol{h}_k^H \boldsymbol{w}_i\right|^2 + \sigma_k^2} \leq \eta_k\right] \leq \epsilon$$

- Massive MIMO hybrid beamforming design

$$\max_{\boldsymbol{\Theta}, \boldsymbol{p}} \sum_{k=1}^{K} \mathbb{E}\left[\log\left(1 + \frac{\left|\boldsymbol{h}_k^H \boldsymbol{F} \boldsymbol{g}_k\right|^2}{\sum_{i \neq k} \left|\boldsymbol{h}_k^H \boldsymbol{F} \boldsymbol{g}_i\right|^2 + 1}\right)\right]$$
$$s.t. \; \mathbb{E}\left[Tr\left(\boldsymbol{F}\boldsymbol{G}\boldsymbol{G}^H \boldsymbol{F}^H\right)\right] \leq P$$

Methodology

- Construct approximate problem using surrogate functions

$$\bar{\boldsymbol{x}}^t = \underset{\boldsymbol{x}\in\mathcal{X}}{\operatorname{argmin}}\ \bar{f}_0^t(\boldsymbol{x})$$
$$s.t.\ \bar{f}_i^t(\boldsymbol{x}) \leq 0, i = 1, \ldots, m$$

$$\boldsymbol{x}^{t+1} = \left(1 - \gamma^t\right)\boldsymbol{x}^t + \gamma^t \bar{\boldsymbol{x}}^t$$

Here, $\gamma^t \downarrow 0$ with $\sum_t \gamma^t = \infty$ and $\sum_t (\gamma^t)^2 < \infty$

- Construction of smooth surrogates with appropriate assumptions
- Convergence analysis performed
- Parallel implementation discussed
- Results shown for the three applications

On Global Linear Convergence in Stochastic Nonconvex Optimization for Semidefinite Programming

Problem

- Nonlinear stochastic semidefinite optimization problem is considered in the scenario of statistical learning
- More general and weaker conditions assumed
- Applications include matrix sensing, subspace tracking, community detection, PCA & recommendation systems

Contributions

- Establish global linear convergence of SGD for a more general non-linear objective function
- Also propose an initialization scheme to ensure faster descent

Problem

$$\min_{X\in\mathbb{R}^{p\times p}} \; f(X) = \frac{1}{n}\sum_{i=1}^{n} f_i(X) \;\; \text{s.t.} \;\; \mathrm{X} \succeq 0$$

$$\min_{X\in\mathbb{S}^{p}_{+}} \; f(X) = \frac{1}{n}\sum_{i=1}^{n} f_i(X) \;\; \text{s.t.} \;\; \mathrm{rank(X)} \leq \mathrm{r} \leq \mathrm{r}^*$$

- Relaxed into a matrix factorization problem

$$\min_{U\in\mathbb{R}^{p\times r}} \; f(UU^T), \quad X = UU^T$$

- Stochastic Gradient Descent:

$$U^{t+1} = U^t - \eta_t \nabla f_{i_t}(X^t) U^t$$

- Second order lemma proven for SGD
- Compared with other techniques such as projected gradient descent, factored gradient descent
- Matrix sensing:

$$\min_{X \succeq 0} f(X) = \frac{1}{2n} \sum_{i=1}^{n} (b_i - \langle A_i, X \rangle)^2$$

- Numerical results show upto 100x speed-up in convergence of SGD from FGD for the above problem
- Error decomposition consisting of approximation error, optimization error, and statistical error is analysed

Other Interesting Papers

1. Optimization Algorithms for Graph Laplacian Estimation via ADMM and MM
2. Provable Subspace Tracking From Missing Data and Matrix Completion
3. Randomized Two-Timescale Hybrid Precoding for Downlink Multicell Massive MIMO Systems
4. Channel Estimation for Orthogonal Time Frequency Space (OTFS) Massive MIMO
5. Constrained Sampling: Optimum Reconstruction in Subspace With Minimax Regret Constraint
6. High-Dimensional Filtering Using Nested Sequential Monte Carlo