Extreme Compressive Sampling for Covariance Estimation
[M. Azizyan, A. Krishnamurthy, and A. Singh. Arxiv Oct. 2015]

Lekshmi Ramesh

भारतीय विज्ञान संस्थान
INDIAN INSTITUTE OF SCIENCE

Signal Processing for Communications Lab
IISc, Bangalore

June 25, 2016

Problem setup

- Data: $X_1, \ldots, X_n \in \mathbb{R}^d$
- Measurements: $y_i = A_i^T X_i, \quad i \in [n]$,
  where $A_i \in \mathbb{R}^{d \times m}$ is an orthonormal basis for an m-dimensional subspace drawn uniformly at random.
- Goal: Estimate the population covariance matrix (distributional setting)/sample covariance matrix (distribution-free setting) from compressed measurements

- Main results
  - An estimator for the covariance matrix
  - Analysis of the estimator, upper bound on the estimation error
  - Lower bound on the risk of the covariance matrix estimation problem
  - Effective sample size shifts form n to $\frac{nm^2}{d^2}$
- No structural assumptions made on the target covariance matrix
- This presentation: lower bounds in the distributional setting

Lower bound

- Study the minimax risk: worst case error of the best estimator:

$$R_n(\Theta) = \inf_{\widehat{\Sigma}} \sup_{\Sigma \in \Theta} \mathbb{E}_{\substack{X_1^n \sim P_\Sigma \\ \Phi_1^n \sim \mathcal{U}}} [\|\widehat{\Sigma} - \Sigma\|],$$

  where $\|.\|$: ℓ_∞ or spectral norm.

- Distributional setting: set $P_\Sigma = \mathcal{N}(0, \Sigma)$
- We look at results that provide a lower bound on $R_n(\Theta)$ in ℓ_∞/spectral norm (Theorem 7, Theorem 17, Lemma 18)

Quick overview

- A class $\mathcal{P}$ of distributions, indexed by parameter $\theta \in \Theta$
- Samples $X_1, \ldots, X_n \in \mathcal{X}$ drawn i.i.d. from some $P \in \mathcal{P}$
- Estimate some parameter $\theta \in \Theta$ that depends on the true distribution P
- An estimator $\hat{\theta} : \mathcal{X}^n \to \Theta$ of θ based on the observations

$$\hat{\theta} \equiv \hat{\theta}(X_1, \ldots, X_n)$$

- A loss function $\ell : \Theta \times \Theta \to \mathbb{R}_+$
- The risk function associated with an estimator $\hat{\theta}$ is the expected loss,

$$R(\theta, \hat{\theta}) = \mathbb{E}_P \; \ell(\theta, \hat{\theta})$$

- Minimax: minimize the maximum risk $\sup_{\theta \in \Theta} R(\theta, \hat{\theta})$

  Bayes: minimize the average risk $\mathbb{E}_{\theta \sim \pi} R(\theta, \hat{\theta})$

Theorem

Let $\Theta(\ell_\infty, \eta, d)$ denote the set of d-dimensional positive semidefinite matrices with ℓ_∞ norm upper bounded by η. If $\frac{1}{15}\frac{d^2 \log d}{nm^2} \leq 1$ and $d \geq 2$, then we have,

$$R_n(\Theta(\ell_\infty, \eta, d)) \geq \frac{\eta}{7}\sqrt{\frac{d^2 \log d}{15nm^2}}.$$

Fano's method

- Key idea: Reduction to a testing problem

$$\inf_{\hat\theta} \sup_{\theta \in \Theta} P_\theta(d(\hat\theta, \theta) \geq s) \geq \inf_{\hat\theta} \max_{\theta \in \{\theta_1, \ldots, \theta_M\}} P_\theta(d(\hat\theta, \theta) \geq s)$$

- Design a test using the estimator.
  Consider the m-ary hypothesis testing problem with

$$\mathcal{H}_i : X \sim P_{\theta_i}, \quad 1 \leq i \leq m,$$

  where $\theta_1, \ldots, \theta_m$ are chosen such that

$$\min_{i,j} \|\theta_i - \theta_j\|_2 = \alpha.$$

Given an estimator $\hat{\theta} : \mathcal{X} \to \Theta$, consider the following test:

$$T(x) = \arg\min_{1 \le i \le m} \|\hat{\theta}(x) - \theta_i\|_2$$

Now bound the P_e for this test:

$$\begin{aligned} P_{\theta_i}(T(X) \neq i) &= P_{\theta_i}(i \neq \arg\min_j \|\hat{\theta}(X) - \theta_j\|_2) \\ &\le P_{\theta_i}(\|\hat{\theta}(X) - \theta_i\|_2 \ge \frac{\alpha}{2}) \\ &\le \frac{4}{\alpha^2} \mathbb{E}_{P_{\theta_i}} \|\hat{\theta}(X) - \theta_i\|_2^2, \qquad 1 \le i \le m. \end{aligned}$$

$$
\begin{aligned}
P_e^* \leq P_e &= \sum_{i=1}^{m} \frac{1}{m} P_{\theta_i}(T(X) \neq i) \\
&\leq \frac{1}{m} \sum_{i=1}^{m} 4 \frac{\mathbb{E}_{P_{\theta_i}} \|\hat{\theta}(X) - \theta_i\|_2^2}{\alpha^2} \\
&\leq \frac{4}{\alpha^2} \max_{\theta \in \Theta} \mathbb{E}_{P_\theta} \|\hat{\theta}(X) - \theta\|_2^2
\end{aligned}
$$

Lower bound on P_e^*:

$$
\begin{aligned}
P_e^* &= \min_T \frac{1}{m} \sum_{i=1}^{m} P(T(X) \neq i) \\
&\geq 1 - \frac{(I(M;X) + 1)}{\log m} \\
&\geq 1 - \frac{(\max_{i,j} D(P_{\theta_i} || P_{\theta_j}) + 1)}{\log m}.
\end{aligned}
$$

Thus,

$$\max_{\theta \in \Theta} \mathbb{E}_{P_\theta} \|\hat{\theta}(X) - \theta\|_2^2 \geq \frac{\alpha^2}{4} \left(1 - \frac{(\max_{i,j} D(P_{\theta_i} || P_{\theta_j}) + 1)}{\log m} \right)$$

Theorem

Assume that $M \geq 2$ *and suppose that the parameter space* Θ *contains elements* $\theta_0, \ldots, \theta_M$ *associated with probability measures* $P_0, \ldots, P_M$ *such that*
(i) $d(\theta_i, \theta_j) \geq 2s > 0$ *for all* $0 \leq j < k \leq M$*;*
(ii) P_j *is absolutely continuous with respect to* P_0 *for all* $j \in [M]$ *and,*

$$\frac{1}{M} \sum_{j=1}^{M} D(P_j || P_0) \leq \alpha \log M,$$

with $0 < \alpha < \frac{1}{8}$*.*
Then,

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} P_\theta(d(\hat{\theta}, \theta) \geq s) \geq \frac{\sqrt{M}}{1 + \sqrt{M}} \left(1 - 2\alpha - \sqrt{\frac{2\alpha}{\log M}}\right)$$

Lemma

Let P_0 be a distribution on (z, U) where U is an orthonormal basis for a uniform-at-random m-dimensional subspace, $x \sim \mathcal{N}(0, \eta I)$ and $z = U^T x$. Let P_1 be the same distribution but where $x \sim \mathcal{N}(0, \eta I + \gamma v v^T)$ for any unit vector v and any $\gamma \in \mathbb{R}$ such that $\gamma \geq -\eta$. Then,

$$D(P_1^n || P_0^n) \leq \frac{3}{2} \frac{\gamma^2}{\eta^2} \frac{nm^2}{d^2}.$$

Proof

- $Q_0 = \mathcal{N}(0, \eta I), \quad Q_1 = \mathcal{N}(0, \eta I + \gamma v v^T)$

$$
\begin{aligned}
D(Q_1 || Q_0) &= \frac{1}{2}\left(\log \frac{|\eta I|}{|\eta I + \gamma v v^T|} + \mathrm{Tr}((\eta I)^{-1}(\eta I + \gamma v v^T)) - d\right) \\
&= \frac{1}{2}\left(\log \frac{\eta}{\eta + \gamma} + \frac{\eta + \gamma}{\eta} - 1\right)
\end{aligned}
$$

Using

$$
\log x \geq \frac{x^2 - 1}{x^2 + 1} \quad (\text{for } x \geq 1)
$$

we have

$$D(Q_1||Q_0) \leq \frac{1}{2}\left(\frac{1-t^2}{1+t^2}+t-1\right)$$
$$=\frac{1}{2}\frac{t}{(1+t^2)}(t-1)^2,$$

where $t=\frac{\eta+\gamma}{\eta}$.
Thus,

$$D(Q_1^n||Q_0^n) \leq \frac{n\gamma^2}{2\eta^2}$$

- $\Sigma_0 = \eta I, \quad \Sigma_1 = \eta I + \gamma e_1 e_1^T$
- $P_0 = \mathcal{N}(0, U^T \Sigma_0 U), \quad P_1 = \mathcal{N}(0, U^T \Sigma_1 U)$

$$D(P_1 || P_0) = \int \mathcal{N}(0, U^T \Sigma_1 U)\text{Unif}(U) \log \left(\frac{\mathcal{N}(0, U^T \Sigma_1 U)\text{Unif}(U)}{\mathcal{N}(0, U^T \Sigma_0 U)\text{Unif}(U)} \right)$$
$$= \mathbb{E}_{U \sim \text{Unif}} D(\mathcal{N}(0, U^T \Sigma_1 U) || \mathcal{N}(0, U^T \Sigma_0 U))$$

$$= \mathbb{E}_{U \sim \text{Unif}} \; \frac{1}{2} \left(\frac{1}{\eta} \text{Tr}(\eta I_m + \gamma U^T e_1 e_1^T U) - m - \log \frac{det(\eta I_m + \gamma U^T e_1 e_1^T U)}{\eta^m} \right)$$

- Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of $U^T \Sigma_1 U$ $(= \eta I_m + \gamma U^T e_1 e_1^T U)$

Then,

$$
\begin{aligned}
D(P_1||P_0) &\leq \mathbb{E}_{U \sim \text{Unif}} \;\; \frac{1}{2} \sum_{i=1}^{m} (\frac{\lambda_i}{\eta} - 1)^2 \\
&= \mathbb{E}_{U \sim \text{Unif}} \;\; \frac{1}{2\eta^2} \| U^T (\Sigma_0 - \Sigma_1) U \|_F^2 \\
&= \mathbb{E}_{U \sim \text{Unif}} \;\; \frac{\gamma^2}{2\eta^2} \| U^T e_1 e_1^T U \|_F^2 \\
&= \frac{\gamma^2}{2\eta^2} \sum_{i,j=1}^{m} \mathbb{E}_{U \sim \text{Unif}} \;\; U_{1i}^2 U_{1j}^2 \\
&\leq \frac{\gamma^2}{2\eta^2} \frac{3m^2}{d^2}
\end{aligned}
$$

- Apply Lemma to a set of $d+1$ distributions
- Consider distribution P_0 with data drawn from $\mathcal{N}(0, \eta I)$ and distributions P_j with data drawn from $\mathcal{N}(0, \eta I - \gamma e_j e_j^T)$
- Note: $2s = \gamma$, $\gamma \leq \eta$ (for positive semidefiniteness)
- Using Lemma,

$$\frac{1}{d}\sum_{j=1}^{m} D(P_j || P_0) \leq \frac{3}{2}\frac{\gamma^2}{\eta^2}\frac{nm^2}{d^2}.$$

- Setting $\gamma = \eta\sqrt{\frac{2\alpha d^2 \log d}{3nm^2}}$, we have

$$\inf_{\widehat{\Sigma}} \sup_{\Sigma} P_{\Sigma}\left(\|\widehat{\Sigma} - \Sigma\|_{\infty} \geq \eta\sqrt{\frac{2\alpha d^2 \log d}{3nm^2}}\right) \geq \frac{\sqrt{d}}{1+\sqrt{d}}\left(1 - 2\alpha - \sqrt{\frac{2\alpha}{\log d}}\right)$$

- For $\alpha = \frac{1}{10}$ and $d \geq 2$,

$$\inf_{\widehat{\Sigma}} \sup_{\Sigma} P_{\Sigma}\left(\|\widehat{\Sigma} - \Sigma\|_{\infty} \geq \eta\sqrt{\frac{d^2 \log d}{3nm^2}}\right) \geq \frac{1}{7}$$

- Using Markov's inequality,

$$R_n(\Theta(\ell_\infty, \eta, d)) \geq \frac{\eta}{7}\sqrt{\frac{d^2 \log d}{15nm^2}}.$$