EE378B Inference, Estimation, and Information Processing

Practice Final: Group estimation

Andrea Montanari

This final will be assigned Friday, March 19 at 12:00 pm PST and be due on Saturday, March 20 at 12:00 PM PST. Solutions should be submitted by e-mail to Andrea (montanari@stanford.edu) and Qijia (qjiang2@stanford.edu).

You are allowed to use books, notes and papers, but you should provide arguments for anything that was not proved in class. You can also contact Andrea and Kabir for reasonable questions about the text. You are not allowed to consult/collaborate with your colleagues or anybody else.

We want to determine n unknown matrices $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n \in \mathsf{SO}(3)$ (the group of 3×3 orthogonal matrices with determinant equal to 1), which are independent and uniformly distributed in $\mathsf{SO}(3)$. We are given noisy observations of the relative group element between any pair $\boldsymbol{\theta}_i$, $\boldsymbol{\theta}_j$. Namely, for each $i \neq j$, $i, j \leq n$, we observe a matrix $\boldsymbol{Y}_{ij} \in \mathbb{R}^{3\times 3}$ given by

$$\boldsymbol{Y}_{ij} = \boldsymbol{\theta}_i^{\mathsf{T}} \boldsymbol{\theta}_j + \sigma_n \, \boldsymbol{Z}_{ij} \,, \tag{1}$$

where $(\boldsymbol{Z}_{ij})_{i\neq j, i,j \leq n}$ are independent 3×3 matrices, with independent entries $(\boldsymbol{Z}_{ij})_{a,b\leq 3} \sim \mathsf{N}(0,1)$. We are interested in estimators $\widehat{\boldsymbol{\theta}} : \boldsymbol{Y} \mapsto \widehat{\boldsymbol{\theta}}(\boldsymbol{Y})$ that take as input $\boldsymbol{Y} = (\boldsymbol{Y})_{i<j\leq n}$ and return n matrices $\widehat{\boldsymbol{\theta}}(\boldsymbol{Y}) = (\widehat{\boldsymbol{\theta}}_1(\boldsymbol{Y}), \dots, \widehat{\boldsymbol{\theta}}_n(\boldsymbol{Y}))$, $\widehat{\boldsymbol{\theta}}_i(\boldsymbol{Y}) \in \mathbb{R}^{3\times 3}$. Throughout, we will require $\widehat{\boldsymbol{\theta}}_i(\boldsymbol{Y})$ to be an orthogonal matrix (i.e. $\widehat{\boldsymbol{\theta}}_i(\boldsymbol{Y})^{\mathsf{T}} \widehat{\boldsymbol{\theta}}_i(\boldsymbol{Y}) = \boldsymbol{I}_3$).

(1) We want to define a good metric for estimation in this problem. Consider the, for instance, the following options (despite the notation, these are not real distances):

$$\mathsf{dist}_0(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \frac{1}{6n} \sum_{i=1}^{n} \|\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_i\|_F^2 \,, \tag{2}$$

$$\mathsf{dist}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \frac{1}{6n} \min_{\boldsymbol{R}:\, \boldsymbol{R}^{\mathsf{T}}\boldsymbol{R} = \boldsymbol{I}} \sum_{i=1}^{n} \|\boldsymbol{R}\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_i\|_F^2 \,. \tag{3}$$

Why is dist more appropriate than dist_0? Prove the following properties: (i) dist admits the representation (here $\|\cdot\|_*$ denotes the nuclear norm)

$$\mathsf{dist}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = 1 - \left\| \frac{1}{3n} \sum_{i=1}^{n} \boldsymbol{\theta}_i \widehat{\boldsymbol{\theta}}_i^{\mathsf{T}} \right\|_* \,. \tag{4}$$

(ii) $\mathsf{dist}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$; (iii) $\mathsf{dist}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) \in [0, 1]$.

(2) Consider the baseline 'random guessing algorithm', which draws at each i an independent uniformly random orthogonal matrix $\widehat{\boldsymbol{\theta}}_i^{\mathrm{RG}}$. What is the baseline value $D_0 = \lim_{n\to\infty} \mathbb{E}\mathsf{dist}(\widehat{\boldsymbol{\theta}}^{\mathrm{RG}}, \boldsymbol{\theta})$?

[We don't require a full proof for this point.]

(3) Derive the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}^{\mathrm{ML}}(\boldsymbol{Y})$ for this problem. Show that, for $\sigma_n = 0$, $\mathsf{dist}(\widehat{\boldsymbol{\theta}}^{\mathrm{ML}}(\boldsymbol{Y}), \boldsymbol{\theta}) = 0$.

(4) Construct a spectral algorithm $\widehat{\boldsymbol{\theta}}^{\text{SP}}(\boldsymbol{Y})$ for this problem. Recall that the algorithm must return orthogonal matrices $\widehat{\boldsymbol{\theta}}_i^{\text{SP}}(\boldsymbol{Y})$, $i \le n$. Show that, for $\sigma_n = 0$, $\mathsf{dist}(\widehat{\boldsymbol{\theta}}^{\text{SP}}(\boldsymbol{Y}), \boldsymbol{\theta}) = 0$.

[Hint: A possible approach is to proceed in two steps. First construct non-orthogonal estimates, and then 'round them.']

(5) Assume $\sigma_n = n^a$, for some $a \in \mathbb{R}$. What is the critical value a_* for the spectral algorithm? By this, we mean the value of the exponent such that $\lim_{n\to\infty} \mathbb{E}\mathsf{dist}(\widehat{\boldsymbol{\theta}}^{\text{SP}}(\boldsymbol{Y}), \boldsymbol{\theta}) = 0$ for $a < a_*$ while $\limsup_{n\to\infty} \mathbb{E}\mathsf{dist}(\widehat{\boldsymbol{\theta}}^{\text{SP}}(\boldsymbol{Y}), \boldsymbol{\theta}) > 0$ for $a > a_*$.

Justify your answer as rigorously as you can. Provide heuristics for steps that are not fully proved.

(6) Implement your spectral algorithm and check the conclusion at point (5) via numerical simulations (by running it for several values of a and of n). Present your results in figures, and submit your code as well.

[It might be useful to know how to generate $\boldsymbol{\theta}_i$ uniformly random on $\mathsf{SO}(3)$. A possible way to do it is as follows. Generate a random matrix $\boldsymbol{G}$ with i.i.g. entries $G_{ij} \sim \mathsf{N}(0,1)$, and take its singular value decomposition $\boldsymbol{G} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathsf{T}}$. Return $\boldsymbol{U}$.]

(7) Construct a semidefinite programming (SDP) relaxation of the maximum likelihood problem. Define the SDP estimator $\widehat{\boldsymbol{\theta}}^{\text{SDP}}(\boldsymbol{Y}) = (\widehat{\boldsymbol{\theta}}_i^{\text{SDP}}(\boldsymbol{Y}))_{i\le n}$. Do you think that there exists a critical noise for exact reconstruction? Namely, is there $\sigma_n^{**} > 0$ such that $\mathsf{dist}(\widehat{\boldsymbol{\theta}}^{\text{SDP}}(\boldsymbol{Y}), \boldsymbol{\theta}) = 0$ with high probability for $\sigma_n < \sigma_n^{**}$?

[For the last question (existence of $\sigma_n^{**} > 0$) we are only expecting an educated guess. Of course, if you have a proof of your claim, go ahead and write it!]