PARAMETER ESTIMATION FROM HETEROGENEOUS/MULTIMODAL DATA SETS

Inbar Fijalkow[1], Elad Heiman[2] and Hagit Messer[2]
[1] ETIS / ENSEA - Université Cergy-Pontoise - CNRS, France
[2] School of Electrical Engineering, Tel Aviv University, Israel

Introduction

Goal: **fusion of non-linear estimators**
Example: measure precipitation from both wireless telecom [Messer2015] and rain gauges.
$\Rightarrow$ **We propose to study linear weighting of Maximum Likelihood (ML) estimators**.

Problem:

- find optimal weights
- for non-linear data model, the optimal weights depend on the parameters to be estimated

Our apporach:

- derive heuristic weights from the optimal ones
- improve them iteratively

Results:

- proposition of optimal, heuristic weights
- conditions for heuristic weights to improve performance with respect to individual estimators

Heterogeneous sensor networks

Assumptions and notations:

- $\boldsymbol{\theta} = (\theta_1, ..., \theta_P)^t$ to be estimated from D data sets,

$$\mathbf{x}_i = \mathbf{s}_i(\boldsymbol{\theta}) + \mathbf{n}_i, i = 1, ..., D.$$

of size N_i, $N_i \geq P$. $\mathbf{n}_i$ follows $\mathcal{N}(\mathbf{0}, \sigma_i^2 I_{N_i})$.

- $\mathbf{x}_i \Rightarrow$ ML estimator $\widehat{\boldsymbol{\theta}}_i$ with covariance $\mathcal{C}ov(\widehat{\boldsymbol{\theta}}_i) \rightarrow$ FIM: $F_i^{-1}(\boldsymbol{\theta}) = \frac{1}{\sigma_i^2} G_i(\boldsymbol{\theta}) G_i(\boldsymbol{\theta})^t = \frac{N_i}{\sigma_i^2} \tilde{F}_i^{-1}(\boldsymbol{\theta})$ asymptotically.

Problem setting:
We are looking for $\alpha_i \geq 0$, $\sum_{i=1}^{D} \alpha_i = 1$,

$$\widehat{\boldsymbol{\theta}}_\alpha = \sum_{i=1}^{D} \alpha_i \widehat{\boldsymbol{\theta}}_i$$

minimizing the asymptotical variance of $\widehat{\boldsymbol{\theta}}_\alpha$.

Optimal weights for linear fusion

Proposition 1 *The weights minimizing $var_\infty(\widehat{\boldsymbol{\theta}}_\alpha)$ are*

$$\alpha_i^*(\boldsymbol{\theta}) = \frac{1}{\sum_{j=1}^{D-1} \frac{\rho_i \lambda_i^*(\boldsymbol{\theta})}{\rho_j \lambda_j^*(\boldsymbol{\theta})} + \rho_i \lambda_i^*(\boldsymbol{\theta})}, i = 1, ..., D-1 \quad (1)$$

with $\lambda_i^*(\boldsymbol{\theta}) = \frac{tr(\tilde{F}_i^{-1}(\boldsymbol{\theta}))}{tr(\tilde{F}_D^{-1}(\boldsymbol{\theta}))}$ *and* $\rho_i = \frac{\sigma_i^2}{\sigma_D^2} \frac{N_D}{N_i}$.

Unfortunatly, $\boldsymbol{\alpha}^*(\boldsymbol{\theta})$ depend on the unknown $\boldsymbol{\theta}$! [a]
For $D = 2$, $\alpha_1^*(\boldsymbol{\theta}) = \frac{1}{1+\rho_1 \lambda_1^*(\boldsymbol{\theta})} \in [\frac{1}{1+\rho_1 \lambda_{max}(\boldsymbol{\theta})}, \frac{1}{1+\rho_1 \lambda_{min}(\boldsymbol{\theta})}]$ where $\lambda_{min/max}(\boldsymbol{\theta})$ are eigenvalues of $\mathcal{F}(\boldsymbol{\theta}) = \tilde{F}_2^{1/2}(\boldsymbol{\theta}) \tilde{F}_1^{-1}(\boldsymbol{\theta}) \tilde{F}_2^{1/2}(\boldsymbol{\theta})$.
Trends: $\rho_1 >> 1 \Rightarrow \widehat{\boldsymbol{\theta}}_{\alpha^*} \approx \widehat{\boldsymbol{\theta}}_2$; $\rho_1 << 1 \Rightarrow \widehat{\boldsymbol{\theta}}_{\alpha^*} \approx \widehat{\boldsymbol{\theta}}_1$.

[a] *Proof:* in "Parameter estimation from heterogeneous/multimodal data sets", I. Fijalkow, E. Heiman, H. Messer, to appear in IEEE Signal Processing Letters, 2016.

Proposed sub-optimal weights

We follow the "indication" of $\rho_i \Rightarrow$ we propose the heuristic:

$$\alpha_i^h = \frac{1}{\sum_{j=1}^{D-1} \frac{\rho_i}{\rho_j} + \rho_i} \text{ for } i = 1, ..., D-1$$

and $\alpha_D^h = 1 - \sum_{i=1}^{D-1} \alpha_i^h = \frac{1}{1+\sum_{i=1}^{D-1} \rho_i}$.

Conditions for improved performance

For $D = 2$, we know $var(\widehat{\boldsymbol{\theta}}_{\alpha^*(\boldsymbol{\theta})}) \leq var(\widehat{\boldsymbol{\theta}}_{\alpha^h})$ and

$$tr((F_1(\boldsymbol{\theta}) + F_2(\boldsymbol{\theta}))^{-1}) \leq var(\widehat{\boldsymbol{\theta}}_{\alpha^*(\boldsymbol{\theta})}) \leq var(\widehat{\boldsymbol{\theta}}_i)$$

Proposition 2 *Asymptotically, $var(\widehat{\boldsymbol{\theta}}_{\alpha^h}) \leq var(\widehat{\boldsymbol{\theta}}_i)$, iff* $\frac{1}{2+\rho} \leq \lambda^*(\boldsymbol{\theta}) \leq 2 + \frac{1}{\rho}$.

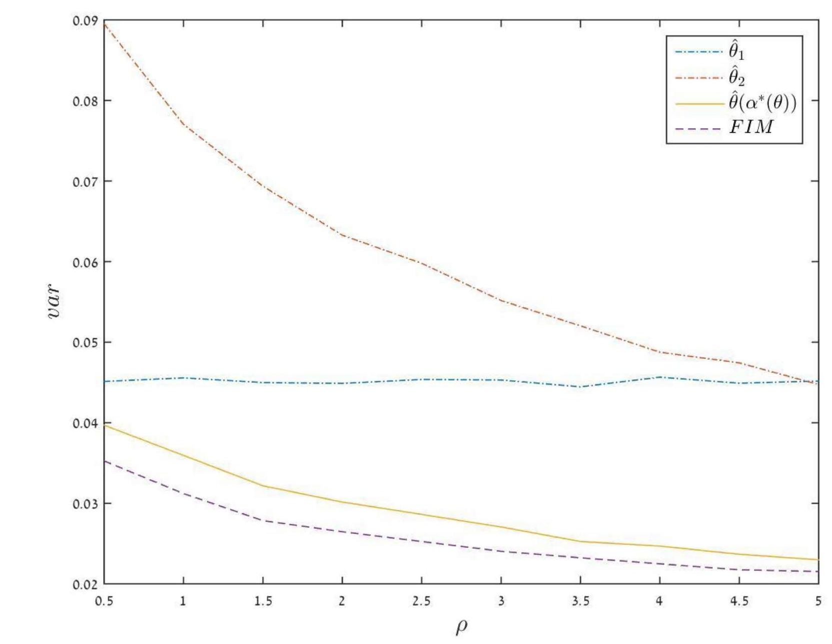
$\Rightarrow$ when $1/2 \leq \lambda^*(\boldsymbol{\theta}) \leq 2$, $\widehat{\boldsymbol{\theta}}_{\alpha^h}$ overcomes $\widehat{\boldsymbol{\theta}}_i$.

Numerical illustration

Simulation settings satisfying Proposition 2 conditions: $P = 2$

- data set 1: linear, $N_1 = 2$, $\sigma_1 = 0.05$ and $tr(F_1^{-1}) = 2$
- data set 2: $\mathbf{s}_2(\boldsymbol{\theta}) = H_2(\sin(\theta_1) + \sin(\theta_2), \cos(\theta_1) + \cos(\theta_2))^t$, $\sigma_2 = 0.05$, N_2 varies

Variance of $\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2, \widehat{\boldsymbol{\theta}}(\alpha^*(\boldsymbol{\theta}))$ and tr(FIM) v.s ρ, $\boldsymbol{\theta} = (\pi/4, \pi/8)^t$:



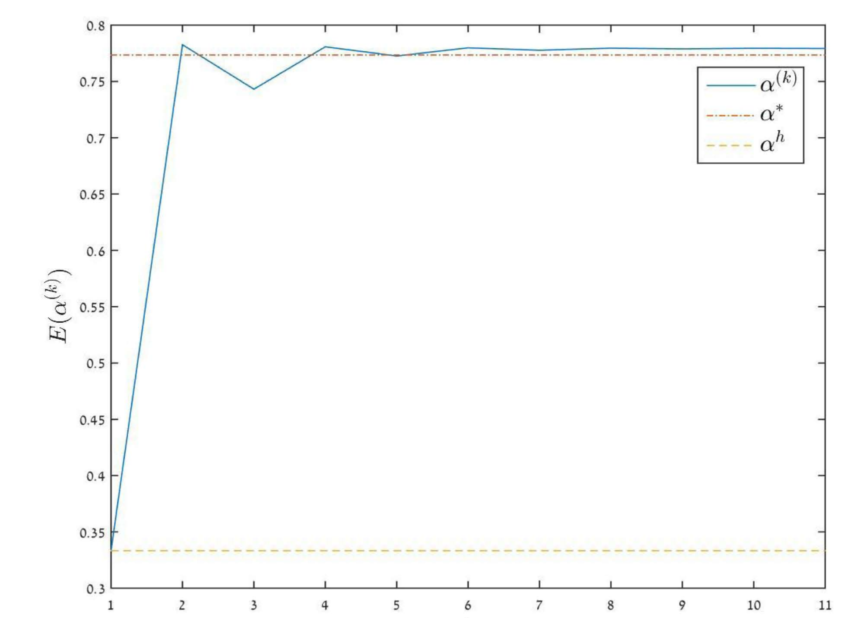
When Proposition 2 conditions are not satisfied:
$N_2 = 2 \Rightarrow$ for $\boldsymbol{\theta} \approx (0, \pi/2)^t$ $\alpha^*(\boldsymbol{\theta}) \approx 1$ whereas $\alpha^h \approx \frac{1}{3}$:
$\Rightarrow$ We want to approach α^h numerically.

Numerical computation of weights

Proposed iterative algorithm:

- **Initialization:** $\lambda_i^{(0)} = 1$ **(implying** $\boldsymbol{\alpha}^{(0)} = \boldsymbol{\alpha}^h$**)**
- **Iteration** k:
  - $\alpha_i^{(k)} = 1/(\sum_{j=1}^{D-1} \frac{\rho_i \lambda_i^{(k)}}{\rho_j \lambda_j^{(k)}} + \rho_i \lambda_i^{(k)}), i = 1, ..., D-1$
  - $\widehat{\boldsymbol{\theta}}^{(k)} = \sum_{i=1}^{D-1} \alpha_i^{(k)} \widehat{\boldsymbol{\theta}}_i + (1 - \sum_{i=1}^{D-1} \alpha_i^{(k)}) \widehat{\boldsymbol{\theta}}_D$
  - $\lambda_i^{(k+1)} = \frac{tr(\tilde{F}_i^{-1}(\widehat{\boldsymbol{\theta}}^{(k)}))}{tr(\tilde{F}_D^{-1}(\widehat{\boldsymbol{\theta}}^{(k)}))}$
- **Repeat until** $|\lambda_i^{(k+1)} - \lambda_i^{(k)}| \leq \epsilon$.

$\boldsymbol{\theta} = (0, \pi/8)^t$, $\rho = 2$, $\rightarrow \alpha^h = 1/3$, $\alpha^* = 0.7735$. $E[\alpha^{(k)}]$ vs k:



Convergence of proposed fixed point algorithm depends on non-linear functions $tr(\tilde{F}_i^{-1}())$.