A Fast Numerical Method for the Optimal Data Fusion in the Presence of Unknown Correlations

FUSION 2018, Cambridge, UK

Xiaohai Zhang
xiaohai@yahoo.com

July 10, 2018

The Problem to Solve

Context

Given N unbiased measurement vectors of dimension M, $\hat{\boldsymbol{x}}_i$s, with measurement covariance matrices denoted by $\boldsymbol{V}_{ii}$ and cross correlation matrices denoted by $\boldsymbol{V}_{ij}$, $i \neq j$, under the context that some or all of the $\boldsymbol{V}_{ij}$ are unknown and linear matrix-weighted coefficients are used for data fusion, i.e., $\hat{\boldsymbol{x}} = \sum_{i=1}^{N} \boldsymbol{A}_i \boldsymbol{x}_i$ where $\boldsymbol{A}_i$ are $M \times M$ matrices with the constraint $\sum_{i=1}^{N} \boldsymbol{A}_i = \boldsymbol{I}_M$.

Goals

- To determine the optimal values of $\boldsymbol{A}_i$ in MMSE given $\boldsymbol{V}_{ii}$ and some available $\boldsymbol{V}_{ij}$ if any
- To present a fast numerical algorithm for the computation of $\boldsymbol{A}_i$

Notes

We implicitly assume $i, j = 1, 2, \cdots, N$ and $k, l = 1, 2, \cdots, M$.

Prior Arts: Optimal Fusion Algorithms

Independent measurements

$$\boldsymbol{A}_i = \left(\sum_{j=1}^{N} \boldsymbol{V}_{jj}^{-1}\right)^{-1} \boldsymbol{V}_{ii}^{-1}$$

Fully known cross correlations

$$[\boldsymbol{A}_1, \cdots, \boldsymbol{A}_N]^T = \boldsymbol{V}^{-1}\boldsymbol{e}\left(\boldsymbol{e}^T\boldsymbol{V}^{-1}\boldsymbol{e}\right)^{-1} \text{ where } \boldsymbol{V} = \{\boldsymbol{V}_{ij}\}, \boldsymbol{e} = [\boldsymbol{I}_M, \cdots, \boldsymbol{I}_M]^T$$

Missing cross correlations (Covariance Intersection)

$$\boldsymbol{A}_i = \left(\sum_{j=1}^{N} \omega_j \boldsymbol{V}_{jj}^{-1}\right)^{-1} \omega_i \boldsymbol{V}_{ii}^{-1} \text{ where } \omega_j \geq 0, \sum_{j=1}^{N} \omega_j = 1.$$

References

- S.-L. Sun and Z.-L. Deng, "Multi-sensor optimal information fusion Kalman filter," Automatica.
- S. J. Julier and J. K. Uhlmann, "Non-divergent estimation algorithm in the presence of unknown correlations", June 1997.

Prior Arts: the Optimal Fusion w/ Unknown Correlations

Theorem

The optimal values of $\boldsymbol{A}_i$ in the sense of MMSE is given by

$$\boldsymbol{A}_i = \boldsymbol{A}_i' \boldsymbol{\Lambda}_i^{-1/2} \boldsymbol{U}_i^T$$

where $\boldsymbol{V}_{ii} = \boldsymbol{U}_i \boldsymbol{\Lambda}_i \boldsymbol{U}_i^T$ is the diagonalization for $\boldsymbol{V}_{ii}$, and $\boldsymbol{A}_i'$s are the values that minimizes the term

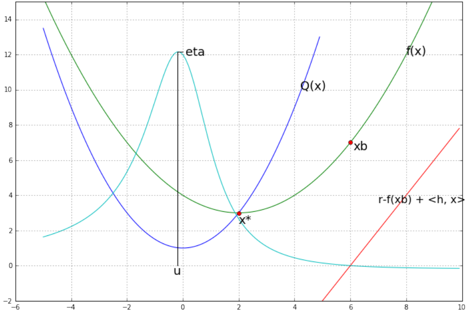
$$\sum_{i=1}^{N} Tr\left(\boldsymbol{A}_i' \boldsymbol{A}_i'^T\right) + \sum_{(i,j)\in\mathcal{D}^c} Tr\left(\boldsymbol{A}_i' \boldsymbol{V}_{ij}' \boldsymbol{A}_j'^T\right) + \sum_{(i,j)\in\mathcal{D}} \sum_{k=1}^{M} \sigma_k \left(\boldsymbol{A}_j'^T \boldsymbol{A}_i'\right)$$

subject to the constraint $\sum_{i=1}^{N} \boldsymbol{A}_i' \boldsymbol{\Lambda}_i^{-1/2} \boldsymbol{U}_i^T = \boldsymbol{I}_M$, where $\boldsymbol{V}_{ij}' \equiv \boldsymbol{\Lambda}_i^{-1/2} \boldsymbol{U}_i^T \boldsymbol{V}_{ij} \boldsymbol{U}_j \boldsymbol{\Lambda}_j^{-1/2}$, $\mathcal{D}$ and $\mathcal{D}^C$ are the index set with unknown and known cross correlations.

References

- X. Zhang, "An optimal data fusion algorithm in the presence of unknown correlations," Manuscript submitted for publication., 2018. [Online]. Available: https://xiaohaionline.com/wp-content/uploads/2018/03/xiaohai_zhang_psof_review.pdf

Prior Arts: OSGA/OSGA-V



- Uses a quadratic positive-valued proxy function (we HAVE)
- Is optimal in terms of the number of required oracle calls for target accuracy(smooth & nonsmooth)
- With no need for global parameters such as Lipschitz constants
- $\eta := E(\gamma_b, \boldsymbol{h}) := \sup_{\boldsymbol{x} \in \mathcal{C}} \left\{ -\frac{\gamma_b + \langle \boldsymbol{h}, \boldsymbol{x} \rangle}{Q(\boldsymbol{x})} \right\}, \gamma_b = \gamma - f(\boldsymbol{x}_b),$

  $\boldsymbol{u} := U(\gamma_b, \boldsymbol{h}) := \arg\max_{\boldsymbol{x} \in \mathcal{C}} \left\{ -\frac{\gamma_b + \langle \boldsymbol{h}, \boldsymbol{x} \rangle}{Q(\boldsymbol{x})} \right\}, f(\boldsymbol{x}_b) - f(\boldsymbol{x}^*) \leq \eta Q(\boldsymbol{x}^*)$

Adapt the OSGA-V algorithm

- To properly select the positive-valued strongly convex proxy function $Q(\cdot)$.
- Given the convex set $\mathcal{C}$ of the optimal fusion problem, to provide the formulation for the OSGA subproblem
- To adapt the OSGA-V algorithm to solve the optimal fusion problem

Verification

Simple experiment to verify faster convergence

Optimal Fusion Problem:

$$\sum_{i=1}^{N} Tr\left(\boldsymbol{A}'_i \boldsymbol{A}'^T_i\right) + \sum_{(i,j)\in\mathcal{D}^c} Tr\left(\boldsymbol{A}'_i \boldsymbol{V}'_{ij} \boldsymbol{A}'^T_j\right) + \sum_{(i,j)\in\mathcal{D}} \sum_{k=1}^{M} \sigma_k \left(\boldsymbol{A}'^T_j \boldsymbol{A}'_i\right)$$

the Proxy Function

$$Q(\boldsymbol{A}') = 1 + \frac{1}{2}\left(\sum_{i=1}^{N} Tr\left(\boldsymbol{A}'_i \boldsymbol{A}'^T_i\right) + \sum_{(i,j)\in\mathcal{D}^c} Tr\left(\boldsymbol{A}'_i \boldsymbol{V}'_{ij} \boldsymbol{A}'^T_j\right)\right)$$

that is equivalent to

$$Q(\boldsymbol{a}') := \frac{1}{2}\left\|\boldsymbol{a}'\right\|^2 + 1$$

with a positive semidefinite matrix

$$\mathcal{P} := \boldsymbol{I}_M \otimes \boldsymbol{V}'_{joint}.$$

Contributions

to turn the OSGA subproblem into a projection problem

- $\eta := \sup_{\boldsymbol{x} \in \mathcal{C}} \left\{ -\frac{\gamma_b + \langle \boldsymbol{h}, \boldsymbol{x} \rangle}{Q(\boldsymbol{x})} \right\}$, $\boldsymbol{u} := \arg\max_{\boldsymbol{x} \in \mathcal{C}} \left\{ -\frac{\gamma_b + \langle \boldsymbol{h}, \boldsymbol{x} \rangle}{Q(\boldsymbol{x})} \right\}$
- $Q(\boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{x}\|_{\mathcal{P}}^2 + 1$
- $g(\boldsymbol{x}) := \eta Q(\boldsymbol{x}) + \gamma_b + \langle \boldsymbol{h}, \boldsymbol{x} \rangle \geq 0$ over $\mathcal{C}$, and reaches 0 at $\boldsymbol{u}$
- For all $\boldsymbol{z} \in \mathcal{C}$, $\langle \triangledown g(\boldsymbol{u}), \boldsymbol{z} - \boldsymbol{u} \rangle \geq 0$, or $\langle \eta \mathcal{P} \boldsymbol{u} + \boldsymbol{h}, \boldsymbol{z} - \boldsymbol{u} \rangle \geq 0$
- For all $\boldsymbol{z} \in \mathcal{C}$, $\langle \triangledown f(\boldsymbol{u}), \boldsymbol{z} - \boldsymbol{u} \rangle \geq 0$ with $f(\boldsymbol{x}) := \frac{1}{2\eta} \left\| \eta \mathcal{P}^{1/2} \boldsymbol{x} + \mathcal{P}^{-1/2} \boldsymbol{h} \right\|_2^2$.
- $\boldsymbol{u} = \arg\min_{\boldsymbol{x} \in \mathcal{C}} f(\boldsymbol{x}) = \arg\min_{\boldsymbol{x} \in \mathcal{C}} \frac{1}{2} \left\| \mathcal{P}^{1/2} \boldsymbol{x} - \boldsymbol{y} \right\|_2^2 = \mathcal{P}^{-1/2} P_{\mathcal{C}_{\mathcal{P}^{1/2}}}(\boldsymbol{y})$ with $\boldsymbol{y} := -\eta^{-1} \mathcal{P}^{-1/2} \boldsymbol{h}$
- Additionally, $\eta Q(\boldsymbol{u}) = -\gamma_b - \langle \boldsymbol{h}, \boldsymbol{u} \rangle$

Contributions

to solve the projection problem

- Subspace $\mathcal{C}$: $\sum_{i=1}^{N} \boldsymbol{A}_i' \boldsymbol{B}_i = \boldsymbol{I}_M$ with $\boldsymbol{B}_i := \boldsymbol{\Lambda}_i^{-1/2} \boldsymbol{U}_i^T$
- $P_C^i(\overrightarrow{\boldsymbol{A}'}) = \boldsymbol{A}_i' + \left(\boldsymbol{I}_M - \sum_{j=1}^{N} \boldsymbol{A}_j' \boldsymbol{B}_j \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{B}_i^T$ with $\boldsymbol{\Sigma} := \sum_{j=1}^{N} \boldsymbol{B}_j^T \boldsymbol{B}_j = \sum_{j=1}^{N} \boldsymbol{V}_{jj}^{-1}$
- $P_{\mathcal{C}_{\mathcal{P}^{1/2}}}^i(\boldsymbol{a}') := \boldsymbol{A}_i' + \left(\boldsymbol{I}_M - \sum_{j=1}^{N} \boldsymbol{A}_j' \hat{\boldsymbol{B}}_j^T \right) \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{B}}_i$ where variables with hats have transformed values of corresponding variables without hats.

Contributions
to finally solve the OSGA subproblem

- $\boldsymbol{u} = \mathcal{P}^{-1/2} P_{\mathcal{C}_{\mathcal{P}^{1/2}}}(\boldsymbol{y})$ with $\boldsymbol{y} := -\eta^{-1}\mathcal{P}^{-1/2}\boldsymbol{h}$
- $\eta Q(\boldsymbol{u}) = -\gamma_b - \langle \boldsymbol{h}, \boldsymbol{u} \rangle$
- Combining the above two provides a quadratic scalar equation for η, and using the projection formulations with the η value leads to the solution for $\boldsymbol{u}$.

Contribution

to adapt OSGA-V algorithm

- Parameter updating schema - Algorithm 1
- Solution to the OSGA-V subproblem (SUB) - Algorithm 2
- OSGA-V for the optimal data fusion - Algorithm 3
- Details can be found in the paper
- Full implementation: https://github.com/xiaohai2016/osgav

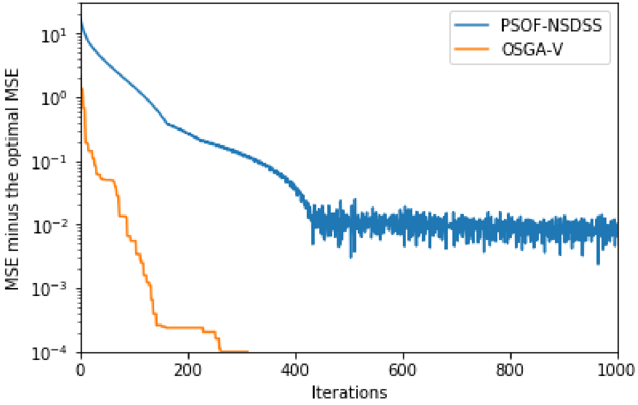


Figure: Convergence rate comparison between PSOF and OSGA-V

Discussions
Further work

- What about other fast subgradient algorithms?
- Even though the application of the OSGA-V for the optimal data fusion shows very promising convergence performance, the theoretical convergence speed (directly from the OSGA paper) bound is NOT better. Can we prove the convergence speed bound is indeed much better?