18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

OptFuse: Low-rank Factor Estimation by Optimal Data-Driven Linear Fusion of Multiple Signal-Plus-Noise Matrices

Himanshu Nayar
EECS Department
University of Michigan, Ann Arbor
Ann Arbor, Michigan, 48104
Email: hnayar@umich.edu

Raj Rao Nadakuditi
EECS Department
University of Michigan, Ann Arbor
Ann Arbor, Michigan, 48104
Email:rajnrao@umich.edu

***Abstract*—We consider the setting where we are given multiple signal-plus-noise matrices. The signal matrices are modeled as low-rank with the same factors (or eigenvectors) but arbitrary (modulo a fixed ordering) eigen-SNRs. One motivating example is the determination of community structure from multiple, independent adjacency matrices. The objective is to combine them linearly so that the eigenvectors of the resulting matrix are as close as possible to the unknown, latent factors. We utilize recent results from random matrix theory to recast this as a constrained data-driven optimization problem and develop an efficient algorithm (OptFuse) for solving it. We demonstrate the improved performance of the algorithm relative to an equal weighting scheme.**

I. Introduction

Fusion of diverse information sources can potentially yield significant detection and classification performance gains and provide robustness relative to sensing using just a single modality. There are significant unmet challenges to multi-modal fusion in environments where the per-modality SNR is low and time-varying. This is particularly so for time-varying operating environments where the physically driven, time-varying characteristics of the battlespace and/or jamming by the adversary introduces distortion/interference such that at various time instances, different modalities will have higher SNRs and the identity of the modalities with better SNRs will keep changing. The time-varying nature of the problem makes computationally intense Bayesian schemes, e.g. POMDP or multi-armed bandit sequential decisions, intractable. Simple methods that average the statistics across the modalities will not realize the performance or diversity gains expected of multi-modal sensing systems.

The key idea that can be exploited is that signals of interest or targets will occupy different low dimensional subspaces for each modality the *expressiveness* of a modality depends on the target signature for that modality (e.g. hyperspectral versus EO/IR). A combination of these probing multiple modalities yields optimal detection or classification performance.

Spectral clustering is a powerful technique for classification of such low dimensional signals encoded via large similarity (or adjacency) matrices. Representative applications where spectral learning has proven to be effective are community detection [7], [11], [12], image segmentation [9], [10]. [5], [6], [8].

In this work, motivated by such fusion problems. we consider a setting where we are given q symmetric signal-pluse-noise type matrices modeled as

$$\widetilde{\mathbf{X_i}} \;=\; \mathbf{X_i} \;+\; \theta_i \mathbf{u}\mathbf{u}^H,$$

where $\mathbf{u}$ is a latent signal eigenvector containing information, and $\mathbf{X}_i$'s are noise matrices. The objective is to derive the best possible estimate of $\mathbf{u}$ using all the samples $\widetilde{\mathbf{X}}_i$'s. Since, eigenvectors are the principle objects in spectral learning, improving the quality of eigenvectors can be expected to lead to performance gains in detection, estimation and classification based applications. Thus the primary technical challenge is to automatically compute the weighting coefficient that is to be assigned to each modality modalities with greater informational content should receive a higher weight while modalities with lower information content should receive a lower weight or not be used at all.

The main contribution of our paper is the development of a data driven algorithm, *OptFuse* for doing that exploits the information content in the 'noise-portion' of the eigen-distribution of the individual matrices. The algorithm also gives us a proxy for the quality of estimation of the signal eigenvector which could later on be used for comparison with other techniques.

The paper is organized as follows: in Section II, we discuss the linear fusion problem setup and discuss the special case of Wigner matrices where we can derive the optimal coefficients in closed form. In Section III, we extend our analysis using *Additive Free Convolution* to propose our data driven algorithm *OptFuse* for finding the optimal coefficients. In Section IV, we show the results when our algorithm is applied to image segmentation. We give concluding remarks where we propose open questions and future directions of work in Section V.

II. Model for Wigner modeled noise matrices

For $i \in \{1, \ldots, q\}$, let

$$\widetilde{\mathbf{X}_\mathbf{i}} = \mathbf{X}_\mathbf{i} + \theta_i \mathbf{u}\mathbf{u}^H, \tag{1}$$

be $N \times N$ symmetric matrices, where $\mathbf{u}$ is an arbitrary unit vector. For $i \in \{1 \ldots, q\}$, $\mathbf{X}_i$ is a noise matrix with independent diagonal entries having zero mean, variance $2/N$ and independent off diagonal entries with zero mean, variance $1/N$. The matrices $\mathbf{X_i}$'s are themselves independent of each other. θ_i are SNR parameters which may be positive or negative.

For an arbitrary $\mathbf{w} \in \mathbb{R}^k$, define $\widetilde{\mathbf{X}}(\mathbf{w}) := \sum \mathbf{w}_i \widetilde{\mathbf{X}}_i$ and $\mathbf{X}(\mathbf{w}) := \sum \mathbf{w}_i \mathbf{X}_i$. Let $\tilde{\mathbf{u}}(\mathbf{w})$ be the estimate of $\mathbf{u}$ obtained from the eigenvalue decomposition of $\tilde{\mathbf{X}}$.

We wish to solve the optimization problem,

$$\mathbf{w}_{\text{opt}} = \underset{\mathbf{w}\in\mathbb{R}^q, ||\mathbf{w}||_1=1}{\arg\max} \; |\langle \tilde{\mathbf{u}}(\mathbf{w}), \mathbf{u}\rangle|^2 \tag{2}$$

The objective function value of this problem gives a measure of reliability of the estimate $\tilde{\mathbf{u}}(\mathbf{w})$.

Theorem 1. *For the model given in* (1), $\mathbf{w}_{\text{opt}}$ *defined in* (2), *we have that,*

$$\mathbf{w}_{\text{opt}_i} \xrightarrow{a.s.} \bar{\mathbf{w}}_{\text{opt}_i} := \frac{\theta_i}{\sum_j \theta_j} \tag{3}$$

Proof. For an arbitrary $\mathbf{w} \in \mathbb{R}^\mathbf{q}$ the matrix $\widetilde{\mathbf{X}}(\mathbf{w})$,

$$\begin{aligned}\widetilde{\mathbf{X}}(\mathbf{w}) &= \sum_{i=1}^{q} \mathbf{w}_i \widetilde{\mathbf{X}}_i \\ &= \underbrace{\left(\sum_{i=1}^{q} \mathbf{w}_i \theta_i\right)}_{\mathbf{w}^\mathbf{H}\theta} \mathbf{u}\mathbf{u}^\mathbf{H} + \underbrace{\sum_{i=1}^{q} \mathbf{w}_i \mathbf{X}_i}_{\mathbf{X}(\mathbf{w})}.\end{aligned}$$

The entries of the matrix $\mathbf{X}(\mathbf{w})$ are independent because each entry of $\mathbf{X}(\mathbf{w})$ is a linear combination of the entries of the matrices $\mathbf{X}_i$ which are themselves independent. Also the means and the variances of the entries of $\mathbf{X}(\mathbf{w})$ are given by,

$$\begin{aligned}\mathbb{E}(\mathbf{X}(\mathbf{w})_{k,l}) &= \sum \mathbf{w}_i \mathbb{E}(\mathbf{X}_{k,l}) = 0 \\ \mathbb{E}(\mathbf{X}(\mathbf{w})^2_{k,l}) &= \sum \mathbf{w}_i^2 \mathbb{E}(\mathbf{X}^2_{k,l}) = \begin{cases} \|\mathbf{w}\|_2^2 \frac{2}{N}, & \text{if } \; l = k, \\ \|\mathbf{w}\|_2^2 \frac{1}{N}, & \text{if } \; l \neq k \end{cases}\end{aligned}$$

Therefore, the matrix $\mathbf{X}(\mathbf{w})$ is just a scaled wigner matrix itself with a scaling factor of $\|\mathbf{w}\|$ and its spectral distribution is just a scaled semicircular law.

Noting that all wigner matrices have the same semicircular eigenvalue distribution, and using the result presented in *Example 3.1* of [1] on low rank perturbations of GOE matrices, the cosine similarity between the observed eigenvector $\tilde{\mathbf{u}}(\mathbf{w})$ and the vector $\mathbf{u}$ is given by,

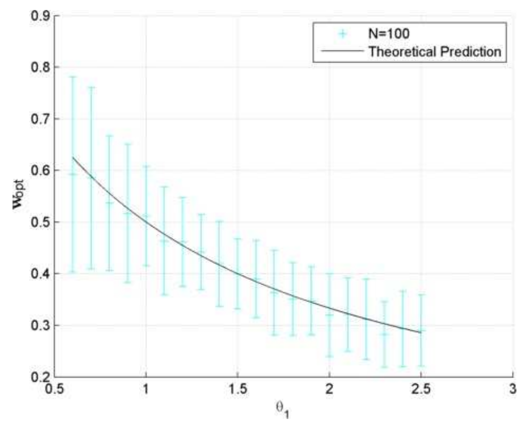
$$|\langle \tilde{\mathbf{u}}(\mathbf{w}), \mathbf{u}\rangle|^2 \xrightarrow{a.s.} \begin{cases} 1 - \dfrac{\|\mathbf{w}\|_2^2}{(\mathbf{w}^\mathbf{H}\theta)^2}, & \text{if } |\mathbf{w}^\mathbf{H}\theta| > \|\mathbf{w}\|_2 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

To maximize the function given above, it suffices to maximize $\dfrac{(\mathbf{w}^\mathbf{H}\theta)^2}{\|\mathbf{w}\|^2}$. By the Cauchy Schwarz Inequality, we have that

$$(\theta^\mathbf{H}\mathbf{w})^\mathbf{2} \leq \|\mathbf{w}\|^\mathbf{2}\|\theta\|^\mathbf{2}$$

$$\frac{\|\mathbf{w}\|_\mathbf{2}^\mathbf{2}}{(\mathbf{w}^\mathbf{H}\theta)^\mathbf{2}} \geq \frac{\mathbf{1}}{\|\theta\|^\mathbf{2}} \implies \mathbf{1} - \frac{\|\mathbf{w}\|_\mathbf{2}^\mathbf{2}}{(\mathbf{w}^\mathbf{H}\theta)^\mathbf{2}} \leq \mathbf{1} - \frac{\mathbf{1}}{\|\theta\|^\mathbf{2}}$$

with the equality holding whenever $\mathbf{w_i} \propto \theta_\mathbf{i}$. By applying the $L1$ normalization constraint , the cosine similarity between $\tilde{\mathbf{u}}(\mathbf{w})$ and $\mathbf{u}$ attains a maxima when $\mathbf{w}_i = \dfrac{\theta_i}{\sum_j \theta_j}$. □



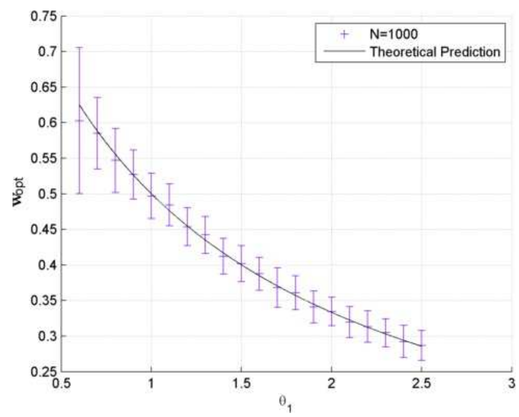


Fig. 1: Optimal Coefficients $\mathbf{w}_{\text{opt}}$ obtained from scanning vs theoretical prediction $\bar{\mathbf{w}}_{\text{opt}}$. In this experiment, two sample matrices of rank one perturbations of GOE noise were taken. The size of the matrices, N was varied. The second sample SNR θ_2 was fixed at 1. MATLAB's fmincon (with random initializations) was used to determine the global maximizer for $|\langle \tilde{\mathbf{u}}, \mathbf{u}\rangle|^2$. The solid line represents our theoretical prediction $\bar{\mathbf{w}}_{\text{opt}}$.

Let $\mathbf{u}_{\mathsf{opt}}$ denote $\mathbf{u}(\bar{\mathbf{w}}_{\mathsf{opt}})$ and let $\mathbf{u}_{\mathsf{avg}}$ denote the estimate of $\mathbf{u}$ obtained from equal weighted averaging of the matrices samples. The following corollary quantifies the performance gain obtained in using optimal coefficients as opposed to a naíve equal weighted averaging scheme.

Corollary 1. *For the model given in* (8),*and using the coefficients* $\bar{\mathbf{w}}_{\mathsf{opt}}$ *as suggested in* (3)*, the following holds,*

$$|\langle \tilde{\mathbf{u}}_{\mathsf{avg}}, \mathbf{u}\rangle|^2 \xrightarrow{a.s.} \begin{cases} 1 - \dfrac{q}{\left(\sum_{j=1}^{q} \theta_j\right)^2}, & \text{if } |\sum_{j=1}^{q} \theta_j| > \sqrt{q} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

and

$$|\langle \tilde{\mathbf{u}}_{\mathsf{opt}}, \mathbf{u}\rangle|^2 \xrightarrow{a.s.} \begin{cases} 1 - \dfrac{1}{\left(\sum_{j=1}^{q} \theta_j^2\right)}, & \text{if } \sum_{j=1}^{q} \theta_j^2 > 1 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

In particular,

$$\lim |\langle \tilde{\mathbf{u}}_{\mathsf{avg}}, \mathbf{u}\rangle|^2 \geq \lim |\langle \tilde{\mathbf{u}}_{\mathsf{opt}}, \mathbf{u}\rangle|^2 \tag{7}$$

Proof. Using (4), plugging in $\mathbf{w} = \bar{\mathbf{w}}_{\mathsf{opt}}$ and $\mathbf{w} = \frac{1}{q}[1, \ldots, 1]$, we get the desired limiting values. To prove the inequality we have to consider three cases:

- $|\sum_{j=1}^{q} \theta_j| > \sqrt{q}$ and $\sum_{j=1}^{q} \theta_j^2 > 1$,
- $|\sum_{j=1}^{q} \theta_j| < \sqrt{q}$ and $\sum_{j=1}^{q} \theta_j^2 > 1$
- $|\sum_{j=1}^{q} \theta_j| < \sqrt{q}$ and $\sum_{j=1}^{q} \theta_j^2 < 1$

In the first case, note that by the "Arithmetic Mean$\leq$ Quadratic Mean" inequality, we have that ,

$$\frac{\left(\sum_{j=1}^{q} \theta_j^2\right)}{q} \geq \left(\frac{\sum_{j=1}^{q} \theta_j}{q}\right)^2$$

$$\implies 1 - \frac{1}{\left(\sum_{j=1}^{q} \theta_j^2\right)} \geq 1 - \frac{q}{\left(\sum_{j=1}^{q} \theta_j\right)^2}.$$

In the second case, if $|\sum_{j=1}^{q} \theta_j| < \sqrt{q}$ and $\sum_{j=1}^{q} \theta_j^2 > 1$, then we have that,

$$1 - \frac{1}{\left(\sum_{j=1}^{q} \theta_j^2\right)} \geq 0.$$

In the third case, the inequality holds trivially. $\square$

Another consequence of this corollary is that the phase transition threshold associated with partial recovery for the case of Wigner Matrices is "lowered" i.e. the range of recoverable SNR's $(\theta_i)_{i=1}^{k}$ increases. In particular, we illustrate the difference between the recoverable regions for our proposed optimal set of coefficients versus equal weighted averaging.

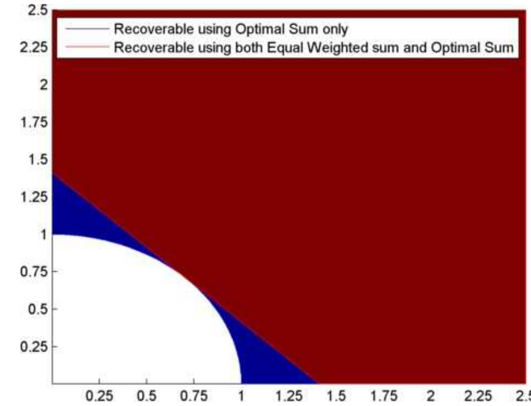


Fig. 2: Theoretical plots for the recoverable regions for Optimal Weighted Scheme vs Equal Weighted Scheme for a two sample case. Note the increase in the partial recovery region. The white region represents the set of parameters (θ_1, θ_2) where the estimated eigenvector $\tilde{\mathbf{u}}(\mathbf{w})$ is orthogonal to $\mathbf{u}$.

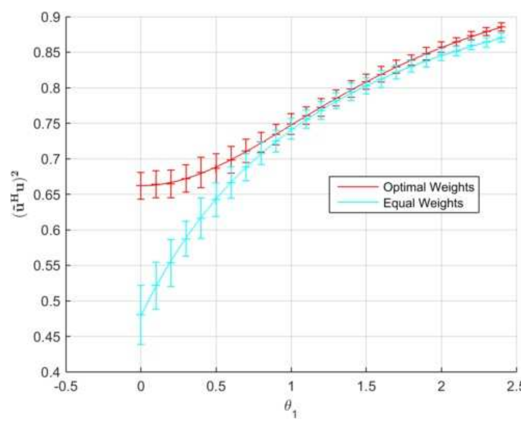


Fig. 3: Simulations illustrating the inner Product Squared of the recovered eigenvector estimate with the original eigenvector for Equal Weighted Averaging vs Optimal Weighted Averaging. In this experiment, we took three sample matrices os size 1000×1000 of rank one perturbations of GOE matrices. Two SNR parameters θ_2 and θ_3 were fixed at 1 and 1.4. The solid lines represent the theoretical predictions.

III. OptFuse: A data-driven fusion algorithm

The results in the previous section make a strong assumption that the noise-only random matrix is Wigner distributed. We now derive a data-driven fusion algorithm that relaxes the assumption on the noise-only matrices. To that end, for $i \in \{1, \ldots, q\}$, we model the symmetric signal-plus-noise matrices as

$$\widetilde{\mathbf{X_i}} = \mathbf{X_i} + \theta_i \mathbf{u}\mathbf{u}^H, \quad (8)$$

be $N \times N$ symmetric matrices, where $\mathbf{u}$ is a unit vector chosen uniformly at random. For $i \in \{1 \ldots, q\}$, $\mathbf{X}_i$ has Haar invariant eigenvectors and its eigenvalues are drawn from the compact real measures μ_i. The matrices $\mathbf{X}_i$'s are assumed to be independent of each other. The matrix $\mathbf{X}(\mathbf{w})$ and the vector $\tilde{\mathbf{u}}(\mathbf{w})$ are defined as before.
It is desired to solve the same optimization problem (2) in this general setting. We propose an efficient optimization routine to solve for the problem of finding limiting values of the $\bar{\mathbf{w}}_{\text{opt}}$.

Alternative Formulation:

Since each eigenvalue of $\mathbf{X}(\mathbf{w})$ is a function of the matrices $\{\mathbf{X}_i : i = 1, \ldots, q\}$, the law of eigenvalues $\mathbf{X}(\mathbf{w})$ itself is a function of $\{\mu_i : i = 1, \ldots, q$. Let the limiting spectral distribution of $\mathbf{X}(\mathbf{w})$ be $\mu_{\mathbf{w}}$. Also let $\theta(\mathbf{w}) = \theta^H \mathbf{w}$.

The following theorem asserts that the optimization problem 2 can be reformulated in terms of only the limiting spectral measure of the matrix $\mathbf{X}(\mathbf{w})$.

Theorem 2. *For $\mathbf{w} \in \mathbb{R}^q$, the following holds ,*

$$|\langle \tilde{\mathbf{u}}(\mathbf{w}), \mathbf{u} \rangle|^2 \xrightarrow{a.s.} -\frac{1}{\theta(\mathbf{w})^2} G_{\mu_{\mathbf{w}}}^{-1'} \left(\frac{1}{\theta(\mathbf{w})} \right) \quad (9)$$

where $G_\mu(.)$ is the Cauchy Transform of the distribution μ i.e.

$$G_\mu(z) = \int \frac{d\mu(x)}{z - x}$$

, if effective SNR satisfies ,

$$\theta(\mathbf{w}) > \lim_{h \downarrow 0} \frac{1}{G_{\mu_{\mathbf{W}}}(b + h)}$$

or

$$\theta(\mathbf{w}) < \lim_{h \downarrow 0} \frac{1}{G_{\mu_{\mathbf{W}}}(a - h)}$$

, where $b = \sup supp(\mu_{\mathbf{w}}) \quad a = \inf supp(\mu_{\mathbf{w}})$.

Proof. Using *Theorem 2.2* of [1], and using the fact that $\mathbf{u}$ is chosen uniformly at random from the unit sphere, we have that

$$|\langle \tilde{\mathbf{u}}(\mathbf{w}), \mathbf{u} \rangle|^2 \xrightarrow{a.s.} \frac{-1}{\theta(\mathbf{w})^2 G'_{\mu_{\mathbf{w}}} \left(G_{\mu_{\mathbf{w}}}^{-1} \left(\frac{1}{\theta(\mathbf{w})} \right) \right)} \quad (10)$$

$$= -\frac{1}{\theta(\mathbf{w})^2} G_{\mu_{\mathbf{w}}}^{-1'} \left(\frac{1}{\theta(\mathbf{w})} \right) \quad (11)$$

□

Using (9), we can assert that the limiting value of the objective function evaluated at $\mathbf{w}$ is a function of the limiting spectral distribution of the matrix $\mathbf{X}(\mathbf{w})$. Hence, the following optimization problem can be used as a proxy instead of (2),

$$\mathbf{w}_{\text{opt}} = \underset{\mathbf{w} \in \mathbb{R}^q, ||\mathbf{w}||_1 = 1}{\arg\max} -\frac{1}{\theta(\mathbf{w})^2} G_{\mu_{\mathbf{w}}}^{-1'} \left(\frac{1}{\theta(\mathbf{w})} \right) \quad (12)$$

It should be noted that by using the assumption that $\mathbf{u}$ is drawn from a uniform distribution, we have eliminated any explicit dependence on the vector $\mathbf{u}$. This means that even though we do not explicitly know the true underlying vector $\mathbf{u}$, we still get a measure of reliability for our estimate.
The next theorem gives us a method to compute the function $-\frac{1}{\theta(\mathbf{w})^2} G_{\mu_{\mathbf{w}}}^{-1'} \left(\frac{1}{\theta(\mathbf{w})} \right)$ without explicitly evaluating the limiting eigenvalue measure of the matrix $\mathbf{X}(\mathbf{w})$.

Theorem 3. *For $\mathbf{w} \in \mathbb{R}^q$ and for $z \in \mathbb{R}$, the following holds ,*

$$-z^2 G_{\mu_{\mathbf{w}}}^{-1'}(z) = \left(\sum_{i=1}^{q} - (\mathbf{w_i} z)^2 G_{\mu_i}^{-1'}(\mathbf{w_i} z) \right) - q + 1 \quad (13)$$

Proof. Using the *free additive convolution identity*, ([2] , [3], [4]), we have that the limiting law of the matrix $\mathbf{X}(\mathbf{w})$ is given by

$$\begin{aligned} \mu_{\mathbf{w}} &= \mathbf{w_1} \mu_1 \boxplus \cdots \boxplus \mathbf{w_q} \mu_q \\ \implies R_{\mathbf{w}}(z) &= \sum_{i=1}^{q} R_{\mathbf{w}_i \mu_i}(z) \\ &= \sum_{i=1}^{q} \mathbf{w_i} R_{\mu_i}(\mathbf{w_i} z). \end{aligned}$$

where $\mathbf{w}_i \mu_i$ is the limiting spectral distribution of $\mathbf{w}_i \mathbf{X}_i$ and $R_\mu(z) = G_\mu^{-1}(z) - \frac{1}{z}$ is the *R transform* of the distribution μ.

$$G_{\mu_{\mathbf{w}}}^{-1}(z) - \frac{1}{z} = \sum_{i=1}^{q} \left(\mathbf{w_i} G_{\mu_i}^{-1}(\mathbf{w_i} z) - \frac{1}{z} \right) \quad (14)$$

Taking derivatives w.r.t z we get ,

$$G_{\mu_{\mathbf{w}}}^{-1'}(z) + \frac{1}{z^2} = \sum_{i=1}^{q} \left(\mathbf{w_i}^2 G_{\mu_i}^{-1'}(\mathbf{w_i} z) + \frac{1}{z^2} \right)$$

$$-z^2 G_{\mu_{\mathbf{w}}}^{-1'}(z) = - \left(\sum_{i=1}^{q} (\mathbf{w_i} z)^2 G_{\mu_i}^{-1'}(\mathbf{w_i} z) \right) - q + 1 \quad (15)$$

which gives the desired result.

□

Corollary 2. *For $\mathbf{w} \in \mathbb{R}^q$, the following holds ,*

$$|\langle \tilde{\mathbf{u}}(\mathbf{w}), \mathbf{u} \rangle|^2 \xrightarrow{a.s.} - \left(\sum_{i=1}^{q} \left(\frac{\mathbf{w_i}}{\theta(\mathbf{w})} \right)^2 G_{\mu_i}^{-1'} \left(\frac{\mathbf{w_i}}{\theta(\mathbf{w})} \right) \right) - q + 1 \quad (16)$$

where $G_\mu(.)$ is the Cauchy Transform of the distribution μ, if each sample SNR obeys

$$\theta_i > \lim_{h \downarrow 0} \frac{1}{G_{\mu_i}(b_i + h)}$$

or

$$\theta_i < \lim_{h \downarrow 0} \frac{1}{G_{\mu_i}(a_i - h)}$$

, *where* $b_i = \sup supp(\mu_i) \quad a_i = \inf supp(\mu_i)$.

Proof. This is a direct consequence of (16) and (9), y replacing $z = \dfrac{1}{\theta(\mathbf{w})}$. □

In (16), we have assumed that each sample should itself be above phase transition. This requirement ensures that convex combination of samples themselves will be above phase transition. The algorithm *OptFuse* (described in 3), also requires this technicality so that the parameter estimation for θ_i's is accurate. The necessity of this requirement will be demonstrated later on in simulations in section III.

We propose the optimization of the following objective function as an alternative to (12),

$$\mathbf{w}_{\mathsf{opt}} = \underset{\mathbf{w}\in\mathbb{R}^q, ||\mathbf{w}||_1=1}{\arg\max} - \left(\sum_{i=1}^{q} \left(\frac{\mathbf{w_i}}{\theta(\mathbf{w})} \right)^2 G_{\mu_i}^{-1'} \left(\frac{\mathbf{w_i}}{\theta(\mathbf{w})} \right) \right) \tag{17}$$

The objective function is itself only a function of the individual limiting noise spectral measures, μ_i's. The objective function itself only depends on the measures μ_i's through $G_\mu(.)$ and the function $G_{\mu_i}(.)$ for each $i \in \{1, \ldots, q\}$, could be approximated from the spectra of $\tilde{\mathbf{X}}_i$.

This result (16) is useful because any iterative scheme for maximization will require a module for computation of the function $-z^2 G_{\mu_{\mathbf{w}}}^{-1'}(z)$. This result gives an explicit method for evaluating the objective function without computing $\mu_{\mathbf{w}}$ at each iteration. Instead, the spectrum of each sample $\mathbf{X}_i$ can be precomputed.

For $\mathbf{\Lambda} \in \mathbb{R}^N$ define, $\widehat{G}_{\mathbf{\Lambda}}(z) = \dfrac{1}{N} \sum_{i=1}^{N} \dfrac{1}{(z - \mathbf{\Lambda}_i)}$.

Algorithm 1 : For finding $\widehat{G}_{\mathbf{\Lambda}}^{-1}(z)$

Require: $\Lambda \in \mathbb{R}^N$, $z \in \mathbb{R}$
Compute x such that

$$\frac{1}{N} \sum_{1}^{N} \frac{1}{x - \mathbf{\Lambda}_i} = z$$

return $x = \widehat{G}_{\mathbf{\Lambda}}^{-1}(z)$.

Algorithm 2 : For finding $\widehat{G}'_{\mathbf{\Lambda}}(z)$

Require: $\Lambda \in \mathbb{R}^N$, $z \in \mathbb{R}$
Compute

$$-\frac{1}{N} \sum_{1}^{N} \frac{1}{(z - \mathbf{\Lambda}_i)^2}.$$

Since the objective function is not guaranteed to be concave, we used a multiple iterate random initialization based gradient descent routine using MATLAB's *fmincon*.

IV. Simulations

For benchmarking the performance of *OptFuse*, we considered the same experimental setup as before. We illustrate the cosine similarity squared between the estimated eigenvector $\tilde{\mathbf{u}}$ and $\mathbf{u}$. We also compare the optimal linear coefficients obtained from *OptFuse* with $\bar{\mathbf{w}}_{\mathsf{opt}}$.

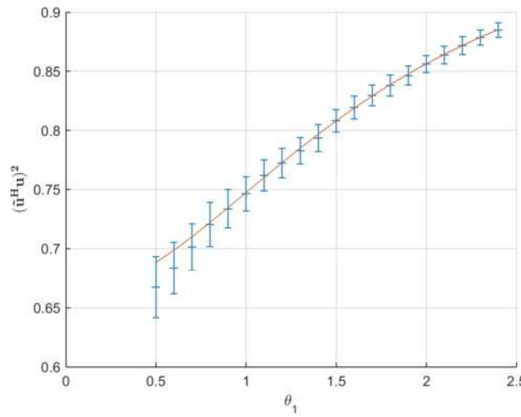


Fig. 4: Simulations illustrating the inner Product Squared of the recovered eigenvector estimate with the original eigenvector for *OptFuse*. In this experiment, we took three sample matrices of size 1000×1000 of rank one perturbations of GOE matrices. Two SNR parameters θ_2 and θ_3 were fixed at 1 and 1.4. The solid lines represent the theoretical predictions. Note that for each sample the phase transition occurs at $\theta_c = 1$. We can see that *OptFuse* is able to predict the cosine similarity between $\tilde{\mathbf{u}}$ and $\mathbf{u}$ accurately.

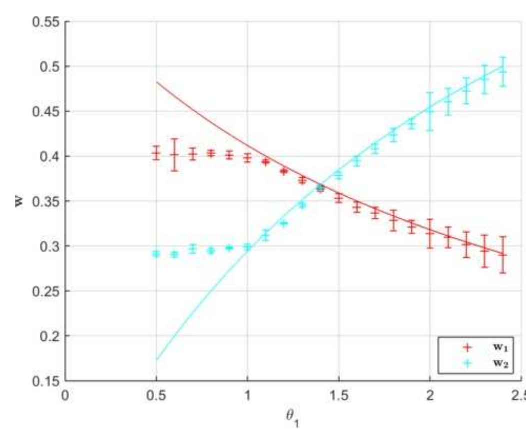


Fig. 5: Simulations illustrating the linear coefficients maximizing the inner Product Squared of the recovered eigenvector estimate with the original eigenvector for *OptFuse*. In this experiment, we took three sample matrices os size 1000×1000 of rank one perturbations of GOE matrices. Two SNR parameters θ_2 and θ_3 were fixed at 1 and 1.4. The solid lines represent the theoretical predictions. We can see that *OptFuse* is able to predict the coefficients predicting cosine similarity between $\tilde{\mathbf{u}}$ and $\mathbf{u}$ accurately.

Algorithm 3 : OptFuse for estimating the j^{th} signal eigenvector

Require: k independent $\{\tilde{\mathbf{X}}_i\}_{i=1}^{k}$ signal-plus-noise matrices and the rank of perturbation, r

for $i = 1, \ldots, k$ **do**

$\mathbf{\Lambda_i} \longleftarrow$ **EIG** $\left(\tilde{\mathbf{X}}_\mathbf{i}\right)$, spectral decompostion of samples.

$\hat{\mathbf{\Sigma}}_i \longleftarrow \mathbf{\Lambda}_i(r+1:N)$, approximate noise spectra.

$\hat{\theta}^{\mathbf{i}}_{\mathbf{j}} \longleftarrow \dfrac{1}{\hat{G}_{\hat{\mathbf{\Sigma}}_i}(\mathbf{\Lambda_i}(\mathbf{j}))}$, estimate of the SNR parameters for each sample.

end for

$\mathbf{w^j}_{\mathsf{opt}} \longleftarrow \arg\max_w \left[-\sum_{i=1}^{q} \left(\frac{w_i}{\hat{\theta_w}}\right)^2 \frac{1}{\hat{G}'_{\mathbf{\Sigma_i}}\left(\hat{G}^{-1'}_{\hat{\mathbf{\Sigma}}_i}\left(\frac{w_i}{\hat{\theta_w}}\right)\right)} \right] - q + 1$ with $\hat{\theta_w} = \sum_{i=1}^{q} w_i \hat{\theta}^i_j, w^T\mathbf{1} = 1, w \succeq \mathbf{0}$, using 1 and 2

return Optimal Linear Combining Coefficients $\mathbf{w^j}$ for estimating the j^{th} signal eigenvector.

V. Conclusion

We considered a problem where we are given multiple similarity matrices and the objective is to find an optimal linear weighting of them so that the weighted sum has the most accurate eigenvectors. We utilized recent results from random matrix theory to recast this as a constrained data-driven optimization problem and developed an efficient algorithm (OptFuse) for solving it. We demonstrated the improved performance of the algorithm relative to an equal weighting scheme.

The objective function for the 3 is provably concave for certain distributions (eg. Wigner Semicircular Law). However , it still is an open question to prove the concavity of the objective function for any arbitrary set of spectral noise measures. In case the concavity is established for any general noise distritbution, convex optimization routines could be leveraged in the algorithm to make it more efficient.

VI. Acknowledgements

This work was supported by an ONR Young Investigator Award N000141110660, an AFOSR Young Investigator Award FA9550-12-1-0266, a NSF award CCF-1116115 and a ARO MURI grant W911NF-11-1-0391.

References

[1] F. Benaych-Georges and R. R. Nadakuditi, "Eigenvalues and Eigenvectors of finite, low rank perturbations of large random matrices," *Advances in Mathematics*, vol. 111, pp. 494–521, 2011.

[2] Pastur, L and Vasilchuk, V, "On the law of addition of random matrices," *Communications in Mathematical Physics*, vol. 214, pp. 249–286, 2000.

[3] Pastur, L and Vasilchuk, V, "Free random variables,No.1," *American Mathematical Soc.*, 1992.

[4] Hiai, Fumio and Petz, Dénes, V, "The semicircle law, free random variables and entropy" *American Mathematical Society Providence*, vol. 77, 2000.

[5] Ng, Andrew Y and Jordan, Michael I and Weiss, Yair and others, "On spectral clustering: Analysis and an algorithm" *Advances in neural information processing systems*, vol. 2,pp. 849–856, 2002.

[6] Von Luxburg, Ulrike, "A tutorial on spectral clustering" *Statistics and computing*, vol. 17,pp. 395–416, 2007.

[7] Von Luxburg, Ulrike, "A Spectral Clustering Approach To Finding Communities in Graph." *SDM*, vol. 5,pp. 76–84, 2005.

[8] Verma, Deepak and Meila, Marina, "A comparison of spectral clustering algorithms" *University of Washington Tech Rep UWCSE030501*, vol. 1,pp. 1–18, 2003.

[9] Shi, Jianbo and Malik, Jitendra, "Normalized cuts and image segmentation" *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22,pp. 888–905, 2000.

[10] Zhang, Xiangrong and Jiao, Licheng and Liu, Fang and Bo, Liefeng and Gong, Maoguo, "Spectral clustering ensemble applied to SAR image segmentation" *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46,pp. 2126–2136, 2008.

[11] Fortunato, Santo, "Community detection in graphs" *Physics Reports*, vol. 486,pp. 75–174, 2010.

[12] Newman, Mark, "Networks: an introduction" *Oxford University Press*, 2010.

[13] Atrey, Pradeep K and Hossain, M Anwar and El Saddik, Abdulmotaleb and Kankanhalli, Mohan S, "Multimodal fusion for multimedia analysis: a survey" *Multimedia systems*, vol. 16,pp. 345–379, 2010.

[14] Wu, Yi and Chang, Edward Y and Chang, Kevin Chen-Chuan and Smith, John R, "Optimal multimodal fusion for multimedia data analysis" *Proceedings of the 12th annual ACM international conference on Multimedia*,pp. 572–579, 2004.

[15] Wright, John and Arvind Ganesh and Shankar Rao and Yigang Peng and Ma, Yi, "Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization" *Advances in Neural Information Processing Systems 22*,pp. 2080–2088, 2009.

[16] Cai, Jian-Feng and Candès, Emmanuel J and Shen, Zuowei, "A singular value thresholding algorithm for matrix completion" *SIAM Journal on Optimization*,pp. 1956–1982, 2010.

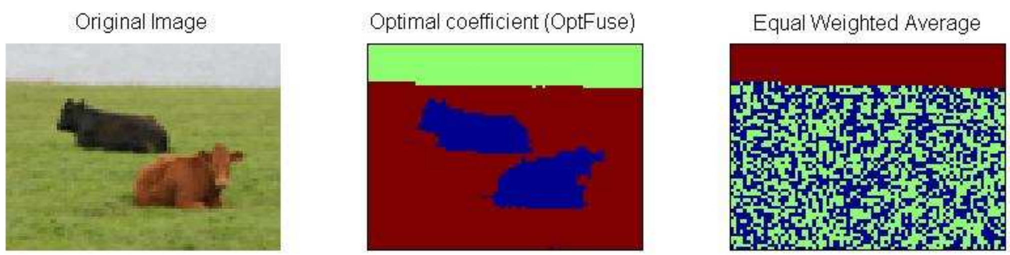


Fig. 6: Image segmentation application on an image taken from the MSR Object Classification dataset. The leftmost image is a original image resized by a factor of $1/8$. From this image a dissimilarity matrix was made using the squared euclidean metric using each one of the R,G and B layers. We added a symmetric GOE noise to each matrix with scaling 10. The objective is to classify the cow objects in the original image using these three dissimilarity matrices. The eigenvalue decomposition of each layer showed that each layer has three outliers in the spectrum. Hence, we used OptFuse to extract the best possible estimate of the eigenvectors of the perturbation. These estimated were later used for finding the clusters in the image using *kmeans* based clustering. The coefficients were $\mathbf{w}^1_{\mathsf{opt}} = [0.4192\ , -0.0072\ , 0.5736]$, $\mathbf{w}^2_{\mathsf{opt}} = [-0.0301\ , 0.9560\ , -0.0139]$ and $\mathbf{w}^3_{\mathsf{opt}} = [0.3512\ , 0.3756\ , 0.2732]$