Bayesian approaches for subspace estimation

Olivier Besson\textsuperscript{1}, Nicolas Dobigeon\textsuperscript{2} and Jean-Yves Tourneret\textsuperscript{2}

\begin{itemize}
  \item \textsuperscript{1} University of Toulouse, ISAE
  \item \textsuperscript{2} University of Toulouse, IRIT/INP-ENSEEIHT – TéSA
  \end{itemize}

\url{http://www.enseeiht.fr/~dobigeon}

Séminaire Problèmes Inverses
Bordeaux, December 21th, 2012
Subspace modeling

- In many signal processing applications, the signal(s) of interest can be assumed to lie in a low dimensional subspace, hence the ubiquitous model

\[ Y = HS + N \]

where \( Y \) stands for the \( N \times K \) observation matrix, \( H \) is the \( N \times p \) matrix whose range space \( \mathcal{R}(H) \) is to be estimated, and \( N \) stands for noise.

- The SVD is one of the most widely used tool to retrieve \( \mathcal{R}(H) \) from \( Y \) since the \( p \) principal left singular vectors of \( Y \) are the maximum likelihood estimates of an orthonormal basis for \( \mathcal{R}(H) \), under the assumption that \( S \) is deterministic and unknown, and \( N \sim \mathcal{N}(0, \sigma_n^2 I) \).
Subspace modeling

- In many signal processing applications, the signal(s) of interest can be assumed to lie in a low dimensional subspace, hence the ubiquitous model

\[ Y = HS + N \]

where \( Y \) stands for the \( N \times K \) observation matrix, \( H \) is the \( N \times p \) matrix whose range space \( \mathcal{R}(H) \) is to be estimated, and \( N \) stands for noise.

- The SVD is one of the most widely used tool to retrieve \( \mathcal{R}(H) \) from \( Y \) since the \( p \) principal left singular vectors of \( Y \) are the maximum likelihood estimates of an orthonormal basis for \( \mathcal{R}(H) \), under the assumption that \( S \) is deterministic and unknown, and \( N \sim \mathcal{N}(0, \sigma_n^2 I) \).
Problem statement

SVD drawbacks
The SVD suffers from two main problems:

1. a performance breakdown when the signal to noise ratio is very low, due to a non-zero probability of subspace swap or subspace leakage;

2. estimation of $\mathcal{R}(H)$ may not reliable for small $K$ and is not possible when $K < p$: $Y$ is at most of rank $K$ and information is lacking about how to complement $\mathcal{R}(Y)$ in order to estimate $\mathcal{R}(H)$. 
Problem statement

Mean-square distance between $\mathcal{R}(\hat{H}_{SV})$ and $\mathcal{R}(H)$

$N = 20$, $p = 5$

$K = 5$
$K = 15$
$K = 30$
SVD drawbacks

The SVD suffers from two main problems:

1. a performance breakdown when the signal to noise ratio is very low, due to a non-zero probability of subspace swap or subspace leakage;

2. estimation of \( \mathcal{R}(\mathbf{H}) \) may not reliable for small \( K \) and is not possible when \( K < p \): \( \mathbf{Y} \) is at most of rank \( K \) and information is lacking about how to complement \( \mathcal{R}(\mathbf{Y}) \) in order to estimate \( \mathcal{R}(\mathbf{H}) \).

Bayesian approach

In order to help estimation, we investigate a Bayesian approach and assign to the unknown matrix \( \mathbf{H} \) an appropriate prior distribution, reflecting our prior knowledge about \( \mathbf{H} \).
Problem statement

**SVD drawbacks**
The SVD suffers from two main problems:

1. a performance breakdown when the signal to noise ratio is very low, due to a non-zero probability of subspace swap or subspace leakage;
2. estimation of $\mathcal{R}(H)$ may not reliable for small $K$ and is *not possible* when $K < p$: $Y$ is at most of rank $K$ and information is lacking about how to complement $\mathcal{R}(Y)$ in order to estimate $\mathcal{R}(H)$.

**Bayesian approach**
In order to help estimation, we investigate a *Bayesian approach* and assign to the unknown matrix $H$ an appropriate prior distribution, reflecting our prior knowledge about $H$. 
Outline

Problem statement

MMSD estimator

Prior modeling on the Stiefel manifold
  Bayesian model
  Numerical simulations

Prior modeling based on CS decomposition
  Bayesian model
  Gibbs sampler
  Numerical simulations

Application to hyperspectral imagery

Conclusions
Outline

Problem statement

**MMSD estimator**

Prior modeling on the Stiefel manifold
  - Bayesian model
  - Numerical simulations

Prior modeling based on CS decomposition
  - Bayesian model
  - Gibbs sampler
  - Numerical simulations

Application to hyperspectral imagery

Conclusions
Bayesian estimators

**MMSE estimator**
Defined as the minimizer of the average squared Euclidean distance:

$$\hat{H}_{\text{mmse}} \triangleq \arg\min_{\hat{H}} E \left\{ \| \hat{H} - H \|^2 \right\}$$

and computed as the mean of the posterior distribution

$$\hat{H}_{\text{mmse}} = E \{ H | Y \} = \int H p(H | Y) dH$$

with

$$p(H | Y) \propto p(Y | H) p(H)$$

One is not interested in $H$ per se but rather in its range space $\mathcal{R}(H)$, and thus we are operating in the Grassmann manifold $G_{N,p}$, i.e., the set of $p$-dimensional subspaces in $\mathbb{R}^N$.

$\rightarrow$ Euclidean distance not a natural metric on $G_{N,p}$
Operating on the Grassmann manifold

Natural distance between subspaces
A natural distance between two subspaces $\mathcal{R}(\mathbf{H}_1)$ and $\mathcal{R}(\mathbf{H}_2)$ is\(^1\)

$$d^2 (\mathbf{H}_1, \mathbf{H}_2) = \sum_{k=1}^{p} \theta_k^2$$

The $\theta_k$’s are the principal angles between these subspaces, which can be obtained by an SVD\(^2\) of $\mathbf{H}_2^T \mathbf{H}_1$:

$$\mathbf{H}_2^T \mathbf{H}_1 = \mathbf{X} \text{diag} (\cos \theta_1, \ldots, \cos \theta_p) \mathbf{Z}^T$$

where $\mathbf{X}$ and $\mathbf{Z}$ are two $p \times p$ unitary matrices.

The cos. of the angles (not the angles themselves) emerge from the SVD.

---


\(^2\)\(\mathbf{H}_1\) and \(\mathbf{H}_2\) are here orthonormal bases for the subspaces.
Another natural distance between subspaces

For sake of practicality, we introduce the distance between two subspaces $\mathcal{R}(H_1)$ and $\mathcal{R}(H_2)$ as

$$d^2(H_1, H_2) = \sum_{k=1}^{p} \sin^2 \theta_k$$

A natural cost function in $G_{n,p}$ since it corresponds to the Frobenius norm of the difference between the projection matrices:

$$d^2(H_1, H_2) = \left\| H_1 H_1^T - H_2 H_2^T \right\|_F^2$$
Minimum mean-square distance (MMSD) estimator

The MMSD estimator of $H$ is defined as

$$
\hat{H}_{mmsd} = \arg \min_{\hat{H}} E \left\{ \| \hat{H} \hat{H}^T - HH^T \|_F^2 \right\}.
$$

Given that $d^2(H_1, H_2) = 2 \left( p - \text{Tr} \left\{ H_1^T H_2 H_2^T H_1 \right\} \right)$, we have also

$$
\hat{H}_{mmsd} = \arg \max_{\hat{H}} E \left\{ \text{Tr} \left\{ \hat{H}^T HH^T \hat{H} \right\} \right\}.
$$
**Minimum mean-square distance (MMSD) estimator**

The MMSD estimator of $H$ is defined as

$$
\hat{H}_{mmsd} = \arg\min_{\hat{H}} E\left\{ \left\| \hat{H}\hat{H}^T - HH^T \right\|_F^2 \right\}.
$$

Given that $d^2(H_1, H_2) = 2\left(p - \text{Tr}\left\{ H_1^T H_2 H_2^T H_1 \right\} \right)$, we have also

$$
\hat{H}_{mmsd} = \arg\max_{\hat{H}} E\left\{ \text{Tr}\left\{ \hat{H}^T HH^T \hat{H} \right\} \right\}.
$$
Since

\[ E \left\{ \text{Tr} \left\{ \hat{H}^T HH^T \hat{H} \right\} \right\} = \int \left[ \int \text{Tr} \left\{ \hat{H}^T HH^T \hat{H} \right\} p(H|Y) \, dH \right] p(Y) \, dY \]

it follows that

\[ \hat{H}_{\text{mmsd}} = \arg \max_{\hat{H}} \text{Tr} \left\{ \hat{H}^T \left[ \int HH^T p(H|Y) \, dH \right] \hat{H} \right\} \]

Therefore, the MMSD estimate of the subspace spanned by \( H \) is given by the largest eigenvectors of the matrix \( \int HH^T p(H|Y) \, dH \), which we denote as

\[ \hat{H}_{\text{mmsd}} = P_p \left\{ \int HH^T p(H|Y) \, dH \right\} = P_p \left\{ E \left\{ HH^T | Y \right\} \right\}. \]

⇒ The MMSD estimator amounts to find the best \( p \)-rank approximation to the posterior mean of the projection matrix \( HH^T \).
MMSD estimator

Since
\[
E \left\{ \text{Tr} \left\{ \hat{H}^T \mathbf{H} \mathbf{H}^T \hat{H} \right\} \right\} = \int \left[ \int \text{Tr} \left\{ \hat{H}^T \mathbf{H} \mathbf{H}^T \hat{H} \right\} p(\mathbf{H}|\mathbf{Y}) \, d\mathbf{H} \right] p(\mathbf{Y}) \, d\mathbf{Y}
\]

it follows that
\[
\hat{\mathbf{H}}_{\text{mmsd}} = \arg \max_{\hat{\mathbf{H}}} \text{Tr} \left\{ \hat{\mathbf{H}}^T \left[ \int \mathbf{H} \mathbf{H}^T p(\mathbf{H}|\mathbf{Y}) \, d\mathbf{H} \right] \hat{\mathbf{H}} \right\}
\]

Therefore, the MMSD estimate of the subspace spanned by \( \mathbf{H} \) is given by the largest eigenvectors of the matrix \( \int \mathbf{H} \mathbf{H}^T p(\mathbf{H}|\mathbf{Y}) \, d\mathbf{H} \), which we denote as
\[
\hat{\mathbf{H}}_{\text{mmsd}} = \mathcal{P}_p \left\{ \int \mathbf{H} \mathbf{H}^T p(\mathbf{H}|\mathbf{Y}) \, d\mathbf{H} \right\} = \mathcal{P}_p \left\{ E \left\{ \mathbf{H} \mathbf{H}^T | \mathbf{Y} \right\} \right\}.
\]

\( \Rightarrow \) The MMSD estimator amounts to find the best \( p \)-rank approximation to the posterior mean of the projection matrix \( \mathbf{H} \mathbf{H}^T \).
Bayesian approaches for subspace estimation

MMSD estimator

\[ \hat{H}_{\text{mmse}} = E \{ H | Y \} = \int H p(H | Y) \, dH \]

\[ \hat{H}_{\text{mmsd}} = \mathcal{P}_p \left\{ E \left\{ H H^T | Y \right\} \right\} = \mathcal{P}_p \left\{ \int H H^T p(H | Y) \, dH \right\} \]

Remarks

▶ In some cases, \( p(H | Y) \) depends on \( H \) through \( H H^T \)

→ postmultiplication of \( H \) by any \( p \times p \) unitary matrix yields the same value of \( p(H | Y) \). Therefore averaging \( H \) over \( p(H | Y) \) does not make sense.

▶ \( \int H H^T p(H | Y) \, dH \) is not necessarily unitary but its range space can be used to estimate \( \mathcal{R}(H) \).
Outline

Problem statement

MMSD estimator

Prior modeling on the Stiefel manifold

Bayesian model

Numerical simulations

Prior modeling based on CS decomposition

Bayesian model

Gibbs sampler

Numerical simulations

Application to hyperspectral imagery

Conclusions
Distributions on the Stiefel manifold\(^3\)

(Matrix-)Bingham distribution

\[
p_B(H) = \frac{1}{\mathbf{1}_F\left(\frac{1}{2}p, \frac{1}{2}N; \mathbf{A}\right)} \exp\left\{ \mathbf{H}^T \mathbf{A} \mathbf{H} \right\}
\]

Remark: only depends on the projection matrix \(\mathbf{H} \mathbf{H}^T\)...

(Matrix-)von Mises-Fisher (vMF) distribution

\[
p_{vMF}(H) = \frac{1}{\mathbf{0}_F\left(\frac{1}{2}N; \frac{1}{4} \mathbf{F}^T \mathbf{F}\right)} \exp\left\{ \mathbf{F}^T \mathbf{H} \right\}
\]

Remark: explicitly depends on the matrix \(\mathbf{H}\)...

\(^3\)\(S_{n,p}\) is the set of \(n \times p\) orthonormal matrices.
Illustration with the linear model

Likelihood
We assume that \( Y = HS + N \) with \( N \sim \mathcal{N} (O, \sigma^2_n I) \) and \( \pi (S) \propto 1: \)

\[
p (Y|H) = \int p (Y|H, S) \pi (S) dS
\]

\[
\propto \exp \left\{ -\frac{1}{2\sigma^2_n} Y^T Y + \frac{1}{2\sigma^2_n} Y^T HH^T Y \right\}.
\]

Remarks

▶ Here, \( \sigma^2_n \) is assumed to be known. The case of unknown \( \sigma^2_n \) can be considered by assigning a prior distribution (e.g., a conjugate prior) to \( \sigma^2_n \) and modifying accordingly the posterior distributions to be derived next.

▶ The MMSD approach can be extended to the mixed case where a parameter vector \( \theta \in \mathbb{R}^q \) needs to be estimated jointly with \( H \). Under such circumstances, one can estimate \( H \) and \( \theta \) as

\[
(\hat{H}_{mmsd}, \hat{\theta}_{mmsd}) = \arg \min_{H, \theta} E \left\{ -\text{Tr} \left\{ \hat{H}^T HH^T \hat{H} \right\} + (\hat{\theta} - \theta)^T (\hat{\theta} - \theta) \right\}.
\]
Illustration with the linear model

Likelihood

We assume that $\mathbf{Y} = \mathbf{HS} + \mathbf{N}$ with $\mathbf{N} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2_n \mathbf{I}\right)$ and $\pi(\mathbf{S}) \propto 1$:

$$p(\mathbf{Y}|\mathbf{H}) = \int p(\mathbf{Y}|\mathbf{H}, \mathbf{S}) \pi(\mathbf{S}) d\mathbf{S}$$

$$\propto \text{etr} \left\{ -\frac{1}{2\sigma^2_n} \mathbf{Y}^T\mathbf{Y} + \frac{1}{2\sigma^2_n} \mathbf{Y}^T\mathbf{H}\mathbf{H}^T\mathbf{Y} \right\}.$$ 

Remarks

- Here, $\sigma^2_n$ is assumed to be known. The case of unknown $\sigma^2_n$ can be considered by assigning a prior distribution (e.g., a conjugate prior) to $\sigma^2_n$ and modifying accordingly the posterior distributions to be derived next.

- The MMSD approach can be extended to the mixed case where a parameter vector $\mathbf{\theta} \in \mathbb{R}^q$ needs to be estimated jointly with $\mathbf{H}$. Under such circumstances, one can estimate $\mathbf{H}$ and $\mathbf{\theta}$ as

$$\left(\hat{\mathbf{H}}_{\text{mmsd}}, \hat{\mathbf{\theta}}_{\text{mmsd}}\right) = \arg \min_{\hat{\mathbf{H}}, \hat{\mathbf{\theta}}} E \left\{ -\text{Tr} \left\{ \hat{\mathbf{H}}^T\mathbf{H}\mathbf{H}^T\hat{\mathbf{H}} \right\} + \left(\hat{\mathbf{\theta}} - \mathbf{\theta}\right)^T \left(\hat{\mathbf{\theta}} - \mathbf{\theta}\right) \right\}.$$
Illustration with the linear model

Likelihood

We assume that $Y = HS + N$ with $N \sim \mathcal{N}(0, \sigma_n^2 I)$ and $\pi(S) \propto 1$:

$$p(Y|H) = \int p(Y|H, S) \pi(S) dS$$

$$\propto \operatorname{etr} \left\{ -\frac{1}{2\sigma_n^2} Y^T Y + \frac{1}{2\sigma_n^2} Y^T HH^T Y \right\}.$$ 

Remarks

- Here, $\sigma_n^2$ is assumed to be known. The case of unknown $\sigma_n^2$ can be considered by assigning a prior distribution (e.g., a conjugate prior) to $\sigma_n^2$ and modifying accordingly the posterior distributions to be derived next.

- The MMSD approach can be extended to the mixed case where a parameter vector $\theta \in \mathbb{R}^q$ needs to be estimated jointly with $H$. Under such circumstances, one can estimate $H$ and $\theta$ as

$$\left( \hat{H}_{\text{mmsd}}, \hat{\theta}_{\text{mmsd}} \right) = \arg \min_{H, \theta} E \left\{ -\operatorname{Tr} \left\{ \hat{H}^T HH^T \hat{H} \right\} + \left( \hat{\theta} - \theta \right)^T \left( \hat{\theta} - \theta \right) \right\}.$$
Illustration with the linear model

Prior distributions
We consider Bingham or von Mises-Fisher (vMF) distributions for \( H \):

\[
\pi_B(H|\bar{H}, \kappa) \propto \text{etr} \left\{ \kappa H^T \bar{H} \bar{H}^T H \right\}
\]

\[
\pi_{\text{vMF}}(H|\bar{H}, \kappa) \propto \text{etr} \left\{ \kappa \bar{H}^T H \right\}
\]

i.e., we assume that \( \mathcal{R}(H) \) is “close” to \( \mathcal{R}(\bar{H}) \) where \( \bar{H} \) is an orthonormal matrix, and \( \kappa \) rules the distance between the two subspaces.

Remarks

- Under the Bingham prior distribution, only \( \mathcal{R}(H) \) and \( \mathcal{R}(\bar{H}) \) are close (at least for large values of \( \kappa \)).
- Under the vMF prior distribution, \( H \) and \( \bar{H} \) also are close.
Illustration with the linear model

Prior distributions
We consider Bingham or von Mises-Fisher (vMF) distributions for $\mathbf{H}$:

$$
\pi_B(\mathbf{H}|\bar{\mathbf{H}}, \kappa) \propto \text{etr} \left\{ \kappa \mathbf{H}^T \bar{\mathbf{H}} \bar{\mathbf{H}}^T \mathbf{H} \right\}
$$

$$
\pi_{vMF}(\mathbf{H}|\bar{\mathbf{H}}, \kappa) \propto \text{etr} \left\{ \kappa \bar{\mathbf{H}}^T \mathbf{H} \right\}
$$

i.e., we assume that $\mathcal{R}(\mathbf{H})$ is “close” to $\mathcal{R}(\bar{\mathbf{H}})$ where $\bar{\mathbf{H}}$ is an orthonormal matrix, and $\kappa$ rules the distance between the two subspaces.

Remarks

- Under the Bingham prior distribution, only $\mathcal{R}(\mathbf{H})$ and $\mathcal{R}(\bar{\mathbf{H}})$ are close (at least for large values of $\kappa$).
- Under the vMF prior distribution, $\mathbf{H}$ and $\bar{\mathbf{H}}$ also are close.
Bingham vs. vMF priors

Let define the average fraction of energy of $H$ in $\mathcal{R} (\bar{H})$

$$\text{AFE} (H, \bar{H}) = \mathbb{E} \left\{ \text{Tr} \left\{ H^T \bar{H} \bar{H}^T H / p \right\} \right\} .$$

Figure: Average fraction of energy of $H$ in $\mathcal{R} (\bar{H})$ versus $\kappa$. $N = 20, p = 5.$
Bingham vs. vMF priors

Figure: Distribution of the angles between $\mathcal{R}(H)$ and $\mathcal{R}(\tilde{H})$ for a Bingham distribution. $N = 20$, $p = 5$ and $\kappa = 20$. 
Bayesian approaches for subspace estimation
Prior modeling on the Stiefel manifold

Bingham vs. vMF priors

Figure: Distribution of the angles between $\mathcal{R}(\mathbf{H})$ and $\mathcal{R}(\bar{\mathbf{H}})$ for a von Mises-Fisher distribution. $N = 20$, $p = 5$ and $\kappa = 20$. 
The posterior distribution writes

\[ p(H|Y) \propto \exp \left\{ H^T \left[ \kappa \bar{H} \bar{H}^T + \frac{1}{2\sigma^2_n} YY^T \right] H \right\} \]

which is recognized as a Bingham distribution with parameter matrix \( \kappa \bar{H} \bar{H}^T + \frac{1}{2\sigma^2_n} YY^T \).

The MMSD estimator is obtained in closed-form as\(^4\)

\[ \hat{H}^{(B)}_{\text{mmsd}} = \mathcal{P}_p \left\{ \kappa \bar{H} \bar{H}^T + \frac{1}{2\sigma^2_n} YY^T \right\}. \]

MMSD estimator (Bingham prior)

The posterior distribution writes

\[
p(H|Y) \propto \text{etr} \left\{ H^T \left[ \kappa \bar{H} \bar{H}^T + \frac{1}{2\sigma_n^2} YY^T \right] H \right\}
\]

which is recognized as a Bingham distribution with parameter matrix \( \kappa \bar{H} \bar{H}^T + \frac{1}{2\sigma_n^2} YY^T \).

The MMSD estimator is obtained in closed-form as\(^4\)

\[
\hat{H}^{(B)}_{\text{mmsd}} = \mathcal{P}_p \left\{ \kappa \bar{H} \bar{H}^T + \frac{1}{2\sigma_n^2} YY^T \right\}.
\]

MMSD estimator (vMF prior)

When $H$ has a vMF prior, the posterior distribution writes

$$p(H|Y) \propto \etr \left\{ \kappa H^T \bar{H} + \frac{1}{2\sigma_n^2} H^T YY^T H \right\}$$

(1)

which is referred to as the Bingham-von-Mises-Fisher (BMF) distribution and denoted as $H|Y \sim \text{BMF}(YY^T, \frac{1}{2\sigma_n^2} I, \kappa \bar{H})$.

It appears that no analytic expression for $\int HH^T p(H|Y) dH$ exists in this case. A MCMC simulation method can be advocated to generate a large number of matrices $H^{(n)}$ drawn from (1), and to approximate the MMSD estimator as

$$\hat{H}_{\text{mmsd}}^{(\text{vMF})} \sim \mathcal{P}_p \left\{ \frac{1}{N_r} \sum_{n=N_{bi}+1}^{N_{bi}+N_r} H^{(n)} H^{(n)^T} \right\}.$$

---

MMSD estimator (vMF prior)

When $\mathbf{H}$ has a vMF prior, the posterior distribution writes

$$p(\mathbf{H}|\mathbf{Y}) \propto \exp \left\{ \kappa \mathbf{H}^T \tilde{\mathbf{H}} + \frac{1}{2\sigma^2_n} \mathbf{H}^T \mathbf{Y} \mathbf{Y}^T \mathbf{H} \right\}$$

which is referred to as the Bingham-von-Mises-Fisher (BMF) distribution and denoted as $\mathbf{H}|\mathbf{Y} \sim \text{BMF} \left( \mathbf{Y} \mathbf{Y}^T, \frac{1}{2\sigma^2_n} \mathbf{I}, \kappa \tilde{\mathbf{H}} \right)$.  

It appears that no analytic expression for $\int \mathbf{H} \mathbf{H}^T p(\mathbf{H}|\mathbf{Y}) d\mathbf{H}$ exists in this case. A MCMC simulation method can be advocated to generate\(^5\) a large number of matrices $\mathbf{H}^{(n)}$ drawn from (1), and to approximate the MMSD estimator as

$$\hat{\mathbf{H}}_{\text{mmsd}}^{(\text{vMF})} \sim \mathcal{P}_p \left\{ \frac{1}{N_r} \sum_{n=N_{bi}+1}^{N_{bi}+N_r} \mathbf{H}^{(n)} \mathbf{H}^{(n)T} \right\}.$$

---

Numerical simulations

Bingham vs. vMF priors

Influence of the number of snapshots

Bingham prior

vMF prior
Numerical simulations
Bingham vs. vMF priors

Influence of the number of the SNR

[Graphs showing the influence of the number of the SNR on different priors]
Drawbacks of this prior modeling

- From a user point of view, it is not obvious to set a value for the concentration parameter $\kappa$ since the latter is not an intuitively appealing parameter, in contrast to the angles between $\mathcal{R}(H)$ and $\mathcal{R}(\tilde{H})$ which are more directly meaningful.
- The Bingham and vMF distributions hold for the whole matrix: the choice of a distribution and a value for $\kappa$ will consequently induce a distribution for the angles, but this relation is not revealed in a straightforward and intelligible manner.
Outline

Problem statement

MMSD estimator

Prior modeling on the Stiefel manifold
  Bayesian model
  Numerical simulations

Prior modeling based on CS decomposition
  Bayesian model
  Gibbs sampler
  Numerical simulations

Application to hyperspectral imagery

Conclusions
CS decomposition

The model proposed herein is based on the CS decomposition of $H$

$$H = \begin{bmatrix} U_1 C \\ U_2 S \end{bmatrix} V^T$$

where

- $U_1$ and $V$ are orthogonal matrices, $U_1, V \in O(p)$,
- $U_2$ is an $N \times p$ semi-orthonormal matrix $U_2^T U_2 = I_p$,
- $C = \text{diag}(\cos \theta_1, \ldots, \cos \theta_p)$ and $S = \text{diag}(\sin \theta_1, \ldots, \sin \theta_p)$.

The $\theta_k$ correspond to the principal angles between $\mathcal{R}(H)$ and $\mathcal{R}(\bar{H})$ while the columns of $\begin{bmatrix} U_1 \\ 0 \end{bmatrix}$ and $HV$ are the associated principal vectors.

As requested, this representation has the nice property that the angles between $\mathcal{R}(H)$ and $\mathcal{R}(\bar{H})$ are directly revealed, and do not depend on the matrices $U_1, U_2$ and $V$, which can be arbitrary.
The model proposed herein is based on the CS decomposition of $H$

$$H = \begin{bmatrix} U_1 C \\ U_2 S \end{bmatrix} V^T$$

where

- $U_1$ and $V$ are orthogonal matrices, $U_1, V \in O(p)$,
- $U_2$ is an $N \times p$ semi-orthonormal matrix $U_2^T U_2 = I_p$,
- $C = \text{diag} (\cos \theta_1, \cdots, \cos \theta_p)$ and $S = \text{diag} (\sin \theta_1, \cdots, \sin \theta_p)$.

The $\theta_k$ correspond to the principal angles between $\mathcal{R}(H)$ and $\mathcal{R}(\bar{H})$ while the columns of $\begin{bmatrix} U_1 \\ 0 \end{bmatrix}$ and $HV$ are the associated principal vectors.

As requested, this representation has the nice property that the angles between $\mathcal{R}(H)$ and $\mathcal{R}(\bar{H})$ are directly revealed, and do not depend on the matrices $U_1, U_2$ and $V$, which can be arbitrary.
Bayesian approaches for subspace estimation

Prior modeling based on CS decomposition

Bayesian model

Likelihood

$$p(Y|H) \propto \exp \left\{ -\frac{1}{2\sigma_n^2} Y^T Y + \frac{1}{2\sigma_n^2} Y^T HH^T Y \right\}.$$  

It depends on $H$ only through the projection matrix $HH^T$ and the latter, under the CS decomposition, is independent of $V$. Therefore, we need to set prior distributions for $U_1$, $U_2$ and $\theta = [\theta_1 \ldots \theta_p]^T$ only.

Prior distributions

1. $U_1$ and $U_2$ have uniform prior distributions on the orthogonal group $O(p)$ and the Stiefel manifold $S_{p,N-p}$.

2. $\theta_k$ are i.i.d. random variables with $\theta_k \sim \mathcal{U}([0, \theta_{\text{max}}])$.

→ the model relies on rather mild assumptions, directly involves the $\theta_k$ and only one parameter ($\theta_{\text{max}}$) has to be set by the user.
Bayesian model

Likelihood

\[ p(Y|H) \propto \text{etr} \left\{ -\frac{1}{2\sigma_n^2} Y^T Y + \frac{1}{2\sigma_n^2} Y^T H H^T Y \right\} \]

It depends on \( H \) only through the projection matrix \( H H^T \) and the latter, under the CS decomposition, is independent of \( V \). Therefore, we need to set prior distributions for \( U_1, U_2 \) and \( \theta = [\theta_1 \cdots \theta_p]^T \) only.

Prior distributions

1. \( U_1 \) and \( U_2 \) have uniform prior distributions on the orthogonal group \( O(p) \) and the Stiefel manifold \( S_{p,N-p} \).
2. \( \theta_k \) are i.i.d. random variables with \( \theta_k \sim \mathcal{U}([0, \theta_{\text{max}}]) \).

↪ the model relies on rather mild assumptions, directly involves the \( \theta_k \) and only one parameter \( (\theta_{\text{max}}) \) has to be set by the user.
Computing the MMSD estimator

\[
\hat{H}_{\text{mmsd}} = \mathcal{P}_p \left\{ \int HH^T p(H|Y) dH \right\} = \mathcal{P}_p \left\{ \mathbb{E} \left\{ HH^T | Y \right\} \right\}.
\]

Posterior distribution of \( U_1, U_2, \theta | Y \) with \( Y = [Y_1^T \quad Y_2^T]^T \)

\[
p(U_1, U_2, \theta | Y) \propto p(Y|H) \pi(U_1) \pi(U_2) \pi(\theta)
\]
\[
\propto \text{etr} \left\{ \frac{1}{2\sigma_n^2} \left[ C^2 U_1^T Y_1 Y_1^T U_1 + S^2 U_2^T Y_2 Y_2^T U_2 \right] \right\}
\]
\[
\times \text{etr} \left\{ \frac{1}{\sigma_n^2} Y_2^T U_2 S \mathbf{C} U_1^T Y_1 \right\} \pi(U_1) \pi(U_2) \pi(\theta)
\]
Computing the MMSD estimator

Implementation

- **Approximation:**

\[
\int \mathbf{H} \mathbf{H}^T p(\mathbf{H} | \mathbf{Y}) \, d\mathbf{H} \simeq \frac{1}{N_r} \sum_{t=N_{bi}+1}^{N_{bi}+N_r} \mathbf{H}^{(t)} \mathbf{H}^{(t)^T}
\]

where \( \mathbf{H}^{(n)} \) is distributed according to \( p(\mathbf{H} | \mathbf{Y}) \).

- **Gibbs Sampling:** Since generating according to \( p(\mathbf{H} | \mathbf{Y}) \) is problematic, we propose to successively generate samples from

\[
\begin{align*}
\mathbf{U}_1^{(t)} &\sim p(\mathbf{U}_1 | \mathbf{U}_2, \theta, \mathbf{Y}) \\
\mathbf{U}_2^{(t)} &\sim p(\mathbf{U}_2 | \mathbf{U}_1, \theta, \mathbf{Y}) \\
\theta^{(t)} &\sim p(\theta | \mathbf{U}_1, \mathbf{U}_2, \mathbf{Y})
\end{align*}
\]
Bayesian approaches for subspace estimation
Prior modeling based on CS decomposition

Gibbs sampler

Conditional posterior distribution of $\mathbf{U}_1|\mathbf{U}_2, \theta, \mathbf{Y}$

$$
\mathbf{U}_1|\mathbf{U}_2, \theta, \mathbf{Y} \sim \text{BMF} \left( \mathbf{Y}_1 \mathbf{Y}_1^T, \frac{1}{2\sigma_n^2} \mathbf{C}^2, \frac{1}{\sigma_n^2} \mathbf{Y}_1 \mathbf{Y}_2^T \mathbf{U}_2 \mathbf{S} \mathbf{C} \right).
$$

Conditional posterior distribution of $\mathbf{U}_2|\mathbf{U}_1, \theta, \mathbf{Y}$

$$
\mathbf{U}_2|\mathbf{U}_1, \theta, \mathbf{Y} \sim \text{BMF} \left( \mathbf{Y}_2 \mathbf{Y}_2^T, \frac{1}{2\sigma_n^2} \mathbf{S}^2, \frac{1}{\sigma_n^2} \mathbf{Y}_2 \mathbf{Y}_1^T \mathbf{U}_1 \mathbf{S} \right).
$$

Conditional posterior distribution of $\theta|\mathbf{U}_1, \mathbf{U}_2, \mathbf{Y}$

Making the change of variables $x_k = \sin^2 \theta_k$ yields

$$
p(x_k|\mathbf{U}_1, \mathbf{U}_2, \mathbf{Y}) \propto x_k^{-1/2}(1 - x_k)^{-1/2} \times \exp \left\{ -(\alpha_k - \gamma_k)x_k + 2\beta_k x_k^{1/2}(1 - x_k)^{1/2} \right\} \mathbb{I}_{[0,x_{\text{max}}]}(x_k)
$$

which can be sampled through a Metropolis-Hastings scheme with a scaled beta distribution as the proposal distribution.
Numerical simulations

Scenario

- We consider a scenario with \( N = 20, \ p = 5 \).
- The SNR is defined as

\[
\text{SNR} = 10 \log_{10} \left( \frac{\mathbb{E} \left\{ \text{Tr} \left\{ \Psi^T H^T H \Psi \right\} \right\}}{\mathbb{E} \left\{ \text{Tr} \left\{ N^T N \right\} \right\}} \right) = 10 \log_{10} \left( \frac{p}{N \sigma_n^2} \right).
\]

- The figure of merit is the mean-square distance between \( H \) and \( \hat{H} \):

\[
\text{MSD} \left( \hat{H}, H \right) = \mathbb{E} \left\{ d^2 \left( \hat{H}, H \right) \right\} = \mathbb{E} \left\{ \sum_{k=1}^{p} \theta_k^2 \right\}.
\]

- The angles between \( \mathcal{R} (H) \) and \( \mathcal{R} (\bar{H}) \) are fixed to \( \theta = [15^\circ \ 25^\circ \ 35^\circ \ 45^\circ \ 55^\circ]^T \) which results in \( \text{MSD} \left( \bar{H}, H \right) = 2.1704 \).
- We compare MMSD, MAP and SVD estimators.
Numerical simulations
Influence of $\theta_{\text{max}}$

$MSD(\hat{H}, H)$

- MMSD, $K = 5$, $SNR = 3dB$
- MMSD, $K = 10$, $SNR = 0dB$
Numerical simulations
Influence of the number of snapshots

\begin{align*}
MSD(\hat{H}, H) \\
\text{SNR} = 0\text{dB} & \\
\text{SNR} = 3\text{dB}
\end{align*}
Numerical simulations
Influence of SNR

\[ MSD(\hat{H}, H) \]

\[ K = 5 \]

\[ K = 10 \]
Outline

Problem statement

MMSD estimator

Prior modeling on the Stiefel manifold
  Bayesian model
  Numerical simulations

Prior modeling based on CS decomposition
  Bayesian model
  Gibbs sampler
  Numerical simulations

Application to hyperspectral imagery

Conclusions
Hyperspectral imagery

Hyperspectral Images

- same scene observed at different wavelengths,
Hyperspectral imagery

Hyperspectral Images

- same scene observed at different wavelengths,

Hyperspectral Cube
Hyperspectral imagery

Hyperspectral Images

- same scene observed at different wavelengths,
- pixel represented by a vector of hundreds of measurements.
Hyperspectral imagery

Hyperspectral Images

- same scene observed at different wavelengths,
- pixel represented by a vector of hundreds of measurements.

Hyperspectral Cube
Spectral Mixture Analysis (SMA)

Linear Mixing Model (LMM):

\[ y_p = \sum_{r=1}^{R} m_{r,a_p,r} + n_p \]

Geometrical formulation of SMA
SMA = looking for a simplex enclosing the data
In the noise-free case, the data $\mathbf{Y}$ are represented in a lower-dimensional subset $\mathcal{V}_{R-1}$ of $\mathbb{R}^{R-1}$ without lost of information:

$$\mathcal{V}_{R-1} = \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{R-1}).$$
Clusters of observations generated according to several linear and nonlinear models (LMM, NM, FM and GBM (blue)) and the corresponding endmembers (red).
Bayesian approaches for subspace estimation
Application to hyperspectral imagery

Application to hyperspectral imagery
Evaluating the validity of the linear model

- The widely admitted model is the linear mixing model
  \[ y_\ell = \sum_{r=1}^{R} a_{\ell,r} m_r \]
  where \( m_r \) are the endmembers, \( a_{\ell,r} \) are the abundances satisfying \( a_{\ell,r} \geq 0 \) and \( \sum_{r=1}^{R} a_{\ell,r} = 1 \).

- The centered data belongs to a \( p \)-dimensional subspace: the latter can be obtained from PCA on the whole image, which provides \( \tilde{H} \).

- Our goal is to compare the local subspace around each pixel and check its distance to \( \tilde{H} \), so as to reveal zones where the linear mixing model could be questioned.

- We test our MMSD estimator on real data acquired by the NASA spectro-imager AVIRIS over Moffett Field, CA: the scene consists of a lake and a coastal area.
Bayesian approaches for subspace estimation

Application to hyperspectral imagery

Application to hyperspectral imagery
Bingham-based MMSD estimator
Bayesian approaches for subspace estimation
Application to hyperspectral imagery

Application to hyperspectral imagery
CS-based MMSD estimator

Outline

Problem statement

MMSD estimator

Prior modeling on the Stiefel manifold
  Bayesian model
  Numerical simulations

Prior modeling based on CS decomposition
  Bayesian model
  Gibbs sampler
  Numerical simulations

Application to hyperspectral imagery

Conclusions
Summary

- We consider knowledge-aided subspace estimations in low sample support or low SNR scenarios.

- Bayesian approaches were proposed based on
  - Bingham and von Mises-Fisher prior modeling: a concentration parameter $\kappa$ needs to be tuned.
  - the CS decomposition of $\mathbf{H}$: this model involves rather mild assumptions and depends directly on the angles $\{\theta_k\}_{k=1}^p$ between $\mathbf{H}$ and a prior subspace $\bar{\mathbf{H}}$. Only the maximum range $\theta_{\text{max}}$ of the angles is to be fixed.

- Monte Carlo computations are required to approximate the Bingham-based and CS-based MMSD estimators.

- The methods improve upon the SVD for low number of snapshots $K$ (and works even with $K < p$) and low SNR.

- They were applied to hyperspectral imagery to assess validity of the linear mixing model at the local pixel level.
Main results available in


Related works


Acknowledgments

- Prof. K. Bingham, University of Minnesota.
- Prof. G. Letac, University of Toulouse.
Bayesian approaches for subspace estimation

Olivier Besson\textsuperscript{1}, Nicolas Dobigeon\textsuperscript{2} and Jean-Yves Tourneret\textsuperscript{2}

\textsuperscript{1} University of Toulouse, ISAE
\textsuperscript{2} University of Toulouse, IRIT/INP-ENSEEIHT – TéSA
http://www.enseeiht.fr/~dobigeon

Séminaire Problèmes Inverses
Bordeaux, December 21th, 2012