Arxiv e-print

High-Dimensional Regression with Gaussian Mixtures and Partially-Latent Response Variables (Supplementary Materials)

Antoine Deleforge · Florence Forbes · Radu Horaud

This document contains supplementary materials for the paper:

A. Deleforge, F. Forbes, and R. Horaud. High-Dimensional Regression with Gaussian Mixtures and Partially-Latent Response Variables. *Statistics and Computing*. 2014.

1 The general hybrid GLLiM-EM algorithm

Considering the complete data, with $(\boldsymbol{Y}, \boldsymbol{T})_{1:N}$ being the observed variables and $(Z, \boldsymbol{W})_{1:N}$ being the missing ones, the corresponding EM algorithm consists of estimating the parameter vector $\boldsymbol{\theta}^{(i+1)}$ that maximizes the expected complete-data log-likelihood, given the current parameter vector $\boldsymbol{\theta}^{(i)}$ and the observed data:

$$\boldsymbol{\theta}^{(i+1)} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}[\log p((\boldsymbol{y}, \boldsymbol{t}, \boldsymbol{W}, Z)_{1:N}; \boldsymbol{\theta}) | (\boldsymbol{y}, \boldsymbol{t})_{1:N}; \boldsymbol{\theta}^{(i)}]. \tag{1}$$

Using that $\boldsymbol{W}_{1:N}$ and $\boldsymbol{T}_{1:N}$ are independent conditionally on $Z_{1:N}$ and that $\{\boldsymbol{c}_k^{\mathrm{w}}\}_{k=1}^K$ and $\{\boldsymbol{\Gamma}_k^{\mathrm{w}}\}_{k=1}^K$ are fixed, maximizing (1) is then equivalent to maximizing the following expression:

$$\mathbb{E}_{r_Z^{(i+1)}}[\mathbb{E}_{r_{W|Z}^{(i+1)}}[\log p(\boldsymbol{y}_{1:N} \mid (\boldsymbol{t}, \mathbf{W}, Z)_{1:N}; \boldsymbol{\theta})] + \log p((\boldsymbol{t}, Z)_{1:N}; \boldsymbol{\theta})] \tag{2}$$

where $r_Z^{(i+1)}$ and $r_{W|Z}^{(i+1)}$ denote the posterior distributions

$$r_Z^{(i+1)} = p(\mathbf{Z}_{1:N} | (\boldsymbol{y}, \boldsymbol{t})_{1:N}; \boldsymbol{\theta}^{(i)}) \tag{3}$$

$$r_{W|Z}^{(i+1)} = p(\mathbf{W}_{1:N} | (\boldsymbol{y}, \boldsymbol{t}, Z)_{1:N}; \boldsymbol{\theta}^{(i)}). \tag{4}$$

It follows that the E-step splits into an **E-W** step and an **E-Z** step. The subsequent M-step can also be divided into two steps referred to as the **M-GMM** step and the **M-mapping** step. In what follows details are given for the **E-W** step and the **M-mapping** step as the **E-Z** and **M-GMM** steps are straightforward as explained in the main paper (equations (27) to (30. For the sake of readability, the current iteration superscript $(i+1)$ is replaced with a tilde. Hence, $\boldsymbol{\theta}^{(i+1)} = \widetilde{\boldsymbol{\theta}}$ (the model parameter vector).

A. Deleforge, F. Forbes, R. Horaud
INRIA Grenoble Rhône-Alpes
655 avenue de l'Europe
38330 Montbonnot Saint-Martin, France
Tel.: +33 4 76 61 52 08
E-mail: firstname.lastname@inria.fr

1.1 E-W-step

This step consists of identifying the conditional probability distribution $\widetilde{r}_{W|Z}$ defined in (4). Given parameter estimates, it is fully defined by computing the distribution $p(\boldsymbol{w}_n|Z_n=k,\boldsymbol{t}_n,\boldsymbol{y}_n;\boldsymbol{\theta}^{(i)})$, denoted below by $\widetilde{r}^{\mathrm{w}}_{nk}$, for all n and all k. This density as a function of $\boldsymbol{w}_n$ is proportional to

$$\widetilde{r}^{\mathrm{w}}_{nk} = p(\boldsymbol{w}_n|Z_n=k,\boldsymbol{t}_n,\boldsymbol{y}_n;\boldsymbol{\theta}^{(i)}) \propto p(\boldsymbol{y}_n|\boldsymbol{t}_n,\boldsymbol{w}_n,Z_n=k;\boldsymbol{\theta}^{(i)})\; p(\boldsymbol{w}_n|Z_n=k;\boldsymbol{\theta}^{(i)}). \tag{5}$$

The second term in the right-hand side is equal to $\mathcal{N}(\boldsymbol{w}_n;\boldsymbol{c}^{\mathrm{w}}_k,\boldsymbol{\Gamma}^{\mathrm{w}}_k)$ by virtue of definitions (4) and (17) in the main paper, while from (18)

$$p(\boldsymbol{y}_n|\boldsymbol{t}_n,\boldsymbol{w}_n,Z_n=k;\boldsymbol{\theta}^{(i)}) = \mathcal{N}(\boldsymbol{y}_n;\mathbf{A}^{(i)}_k[\boldsymbol{t}_n;\boldsymbol{w}_n]+\boldsymbol{b}^{(i)}_k,\boldsymbol{\Sigma}^{(i)}_k). \tag{6}$$

Rewriting the expression in the exponential term of the right-hand side as a quadratic form in $\boldsymbol{w}_n$, it is easy to see that $p(\boldsymbol{y}_n|\boldsymbol{t}_n,\boldsymbol{w}_n,Z_n=k;\boldsymbol{\theta}^{(i)})$ induces a term over $\boldsymbol{w}_n$ proportional to a Gaussian density,

$$\mathcal{N}(\boldsymbol{w}_n;\boldsymbol{\mu}^{l(i)}_{nk},\boldsymbol{\Sigma}^{l(i)}_k),$$

with $\boldsymbol{\mu}^{l(i)}_{nk} = \boldsymbol{\Sigma}^{l(i)}_k\; {\mathbf{A}^{\mathrm{w}(i)}_k}^{\top}(\boldsymbol{\Sigma}^{(i)}_k)^{-1}(\boldsymbol{y}_n-\mathbf{A}^{\mathrm{t}(i)}_k\boldsymbol{t}_n-\boldsymbol{b}^{(i)}_k)$, and $\boldsymbol{\Sigma}^{l(i)}_k = ({\mathbf{A}^{\mathrm{w}(i)}_k}^{\top}(\boldsymbol{\Sigma}^{(i)}_k)^{-1}\mathbf{A}^{\mathrm{w}(i)}_k)^{-1}$.

It follows that $\widetilde{r}^{\mathrm{w}}_{nk}$ is proportional to the product of two Gaussian distributions over $\boldsymbol{w}_n$ which according to Lemma 1 in Section 3.1, implies that $\widetilde{r}^{\mathrm{w}}_{nk}$ is Gaussian, with mean $\widetilde{\boldsymbol{\mu}}^{\mathrm{w}}_{nk}$ and covariance matrix $\widetilde{\mathbf{S}}^{\mathrm{w}}_k$ given by:

$$\begin{aligned}
\widetilde{\boldsymbol{\mu}}^{\mathrm{w}}_{nk} &= \widetilde{\mathbf{S}}^{\mathrm{w}}_k\big({\mathbf{A}^{\mathrm{w}(i)}_k}^{\top}(\boldsymbol{\Sigma}^{(i)}_k)^{-1}(\boldsymbol{y}_n-\mathbf{A}^{\mathrm{t}(i)}_k\boldsymbol{t}_n-\boldsymbol{b}^{(i)}_k)+(\boldsymbol{\Gamma}^{\mathrm{w}(i)}_k)^{-1}\boldsymbol{c}^{\mathrm{w}(i)}_k\big)\\
\widetilde{\mathbf{S}}^{\mathrm{w}}_k &= \big((\boldsymbol{\Gamma}^{\mathrm{w}(i)}_k)^{-1}+{\mathbf{A}^{\mathrm{w}(i)}_k}^{\top}(\boldsymbol{\Sigma}^{(i)}_k)^{-1}\mathbf{A}^{\mathrm{w}(i)}_k\big)^{-1}
\end{aligned} \tag{7}$$

1.2 M-mapping-step

This step consists of updating parameters $\{\mathbf{A}_k,\boldsymbol{b}_k,\boldsymbol{\Sigma}_k\}_{k=1}^K$. Omitting the terms in (2) that do not depend on $\{\mathbf{A}_k,\boldsymbol{b}_k,\boldsymbol{\Sigma}_k\}_{k=1}^K$, the M-mapping-step is equivalent to maximizing

$$\mathbb{E}_{r_Z^{(i+1)}}[\mathbb{E}_{r_{W|Z}^{(i+1)}}[\log p(\boldsymbol{y}_{1:N}\mid(\boldsymbol{t},\mathbf{W},Z)_{1:N};\boldsymbol{\theta})]]\;. \tag{8}$$

Using $\widetilde{r}_{nk}=p(Z_n=k|\boldsymbol{t}_n,\boldsymbol{y}_n;\boldsymbol{\theta}^{(i)})$ and $\widetilde{r}^{\mathrm{w}}_{nk}=p(\boldsymbol{w}_n|Z_n=k,\boldsymbol{t}_n,\boldsymbol{y}_n;\boldsymbol{\theta}^{(i)})$, this writes

$$\sum_{n=1}^{N}\sum_{k=1}^{K}\widetilde{r}_{nk}\;\mathbb{E}_{\widetilde{r}^{\mathrm{w}}_{nk}}[\log p(\boldsymbol{y}_n\mid\boldsymbol{t}_n,\boldsymbol{W}_n,Z_n=k;\boldsymbol{\theta})]. \tag{9}$$

Using again the fact that $p(\boldsymbol{y}_n|\boldsymbol{t}_n,\boldsymbol{w}_n,Z_n=k;\boldsymbol{\theta})$ induces an expression over $\boldsymbol{w}_n$ proportional to $\mathcal{N}(\boldsymbol{w}_n;\boldsymbol{\mu}^l_{nk},\boldsymbol{\Sigma}^l_k)$, with $\boldsymbol{\mu}^l_{nk}=\boldsymbol{\Sigma}^l_k\;(\mathbf{A}^{\mathrm{w}}_k)^{\top}(\boldsymbol{\Sigma}_k)^{-1}(\boldsymbol{y}_n-\mathbf{A}^{\mathrm{t}}_k\boldsymbol{t}_n-\boldsymbol{b}_k)$, and $\boldsymbol{\Sigma}^l_k=(\mathbf{A}^{\mathrm{w}\top}_k(\boldsymbol{\Sigma}_k)^{-1}\mathbf{A}^{\mathrm{w}}_k)^{-1}$, it follows that

$$\log(\boldsymbol{y}_n|\boldsymbol{t}_n,\boldsymbol{w}_n,Z_n=k;\boldsymbol{\theta}) = \log C-\frac{1}{2}(\boldsymbol{w}_n-\boldsymbol{\mu}^l_{nk})^T(\boldsymbol{\Sigma}^l_k)^{-1}(\boldsymbol{w}_n-\boldsymbol{\mu}^l_{nk}),$$

where C is a term that does not depend on $\boldsymbol{w}_n$. Then, using Lemma 2 in Section 3.2,

$$\begin{aligned}
\mathbb{E}_{\widetilde{r}^{\mathrm{w}}_{nk}}[\log p(\boldsymbol{y}_n\mid\boldsymbol{t}_n,\boldsymbol{W}_n,Z_n=k;\boldsymbol{\theta})] &= \log C-\frac{1}{2}(\widetilde{\boldsymbol{\mu}}^{\mathrm{w}}_{nk}-\boldsymbol{\mu}^l_{nk})^{\top}(\boldsymbol{\Sigma}^l_k)^{-1}(\widetilde{\boldsymbol{\mu}}^{\mathrm{w}}_{nk}-\boldsymbol{\mu}^l_{nk})-\frac{1}{2}(\widetilde{\mathbf{S}}^{\mathrm{w}}_k\;\boldsymbol{\Sigma}^l_k)\\
&= \log p(\boldsymbol{y}_n\mid\boldsymbol{t}_n,\boldsymbol{W}_n=\widetilde{\boldsymbol{\mu}}^{\mathrm{w}}_{nk},Z_n=k;\boldsymbol{\theta})-\frac{1}{2}\operatorname{tr}(\widetilde{\mathbf{S}}^{\mathrm{w}}_k\;\boldsymbol{\Sigma}^l_k).
\end{aligned} \tag{10}$$

For the update of $\boldsymbol{b}_k$, it follows that the term to be maximized is only the first one and is the same as in a standard GMM where $\boldsymbol{b}_k$ would be the mean of component k for observed data that would be $\boldsymbol{y}_n - \widetilde{\mathbf{A}}_k\widetilde{\boldsymbol{x}}_{nk}$, with $\widetilde{\boldsymbol{x}}_{nk} = [\boldsymbol{t}_n; \widetilde{\boldsymbol{\mu}}^{\mathrm{w}}_{nk}] \in \mathbb{R}^L$. It comes the standard weighted mean formula,

$$\widetilde{\boldsymbol{b}}_k = \sum_{n=1}^{N} \frac{\widetilde{r}_{kn}}{\widetilde{r}_k}(\boldsymbol{y}_n - \widetilde{\mathbf{A}}_k\widetilde{\boldsymbol{x}}_{nk}). \tag{11}$$

For updating $\boldsymbol{\Sigma}_k$, using (10) and (6), as an expression over $\boldsymbol{\Sigma}_k$, the term in (9) to be maximized writes,

$$-\frac{1}{2}\sum_{n=1}^{N} \widetilde{r}_{nk}\ (\log|\boldsymbol{\Sigma}_k| + \mathrm{tr}(\mathbf{A}^{\mathrm{w}}_k\widetilde{\mathbf{S}}^{\mathrm{w}}_k(\mathbf{A}^{\mathrm{w}}_k)^\top\boldsymbol{\Sigma}_k^{-1}) + (\boldsymbol{y}_n - \mathbf{A}_k\widetilde{\boldsymbol{x}}_{nk} - \boldsymbol{b}_k)^\top\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{y}_n - \mathbf{A}_k\widetilde{\boldsymbol{x}}_{nk} - \boldsymbol{b}_k).$$

Denoting $\widetilde{r}_k = \sum_{n=1}^{N}\widetilde{r}_{nk}$, this is equivalent to minimize over $\boldsymbol{\Sigma}_k$

$$\widetilde{r}_k \log|\boldsymbol{\Sigma}_k| + \mathrm{tr}((\sum_{n=1}^{N}\widetilde{r}_{nk}\mathbf{M}_{nk})\boldsymbol{\Sigma}_k^{-1}),$$

where $\mathbf{M}_{nk} = \mathbf{A}^{\mathrm{w}}_k\widetilde{\mathbf{S}}^{\mathrm{w}}_k(\mathbf{A}^{\mathrm{w}}_k)^\top + (\boldsymbol{y}_n - \mathbf{A}_k\widetilde{\boldsymbol{x}}_{nk} - \widetilde{\boldsymbol{b}}_k)(\boldsymbol{y}_n - \mathbf{A}_k\widetilde{\boldsymbol{x}}_{nk} - \widetilde{\boldsymbol{b}}_k)^\top$ and $\sum_{n=1}^{N}\widetilde{r}_{nk}\mathbf{M}_{nk}$ are positive definite matrices.

Using the result recalled in Lemma 3 in Section 3.3, it follows straightforwardly that,

$$\widetilde{\boldsymbol{\Sigma}}_k = \widetilde{\mathbf{A}}^{\mathrm{w}}_k\widetilde{\mathbf{S}}^{\mathrm{w}}_k\widetilde{\mathbf{A}}^{\mathrm{w}\top}_k + \sum_{n=1}^{N}\frac{\widetilde{r}_{kn}}{\widetilde{r}_k}(\boldsymbol{y}_n - \widetilde{\mathbf{A}}_k\widetilde{\boldsymbol{x}}_{nk} - \widetilde{\boldsymbol{b}}_k)(\boldsymbol{y}_n - \widetilde{\mathbf{A}}_k\widetilde{\boldsymbol{x}}_{nk} - \widetilde{\boldsymbol{b}}_k)^\top. \tag{12}$$

As regards $\mathbf{A}_k$, it comes from (9) and (10), omitting the terms independent of $\mathbf{A}_k$, that we have to minimize the following expression over $\mathbf{A}_k$,

$$\mathrm{tr}((\sum_{n=1}^{N}\widetilde{r}_{nk}\widetilde{\mathbf{M}}_{nk})\boldsymbol{\Sigma}_k^{-1}), \tag{13}$$

where $\widetilde{\mathbf{M}}_{nk}$ is the part of $\mathbf{M}_{nk}$ that involves $\mathbf{A}_k$,

$$\begin{aligned}\widetilde{\mathbf{M}}_{nk} &= \mathbf{A}^{\mathrm{w}}_k\widetilde{\mathbf{S}}^{\mathrm{w}}_k(\mathbf{A}^{\mathrm{w}}_k)^\top + (\mathbf{A}_k\widetilde{\boldsymbol{x}}_{nk})(\mathbf{A}_k\widetilde{\boldsymbol{x}}_{nk})^\top - 2(\boldsymbol{y}_n - \boldsymbol{b}_k)(\mathbf{A}_k\widetilde{\boldsymbol{x}}_{nk})^\top \\ &= \mathbf{A}_k\widetilde{\mathbf{S}}^{\mathrm{x}}_k\mathbf{A}_k^\top + \mathbf{A}_k\widetilde{\boldsymbol{x}}_{nk}\widetilde{\boldsymbol{x}}_{nk}^\top\mathbf{A}_k^\top - 2(\boldsymbol{y}_n - \boldsymbol{b}_k)\widetilde{\boldsymbol{x}}_{nk}^\top\mathbf{A}_k^\top,\end{aligned}$$

where $\widetilde{\mathbf{S}}^{\mathrm{x}}_k$ is defined by

$$\widetilde{\mathbf{S}}^{\mathrm{x}}_k = \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \widetilde{\mathbf{S}}^{\mathrm{w}}_k\end{bmatrix}, \tag{14}$$

so that $\mathbf{A}^{\mathrm{w}}_k\widetilde{\mathbf{S}}^{\mathrm{w}}_k(\mathbf{A}^{\mathrm{w}}_k)^\top = \mathbf{A}_k\widetilde{\mathbf{S}}^{\mathrm{x}}_k\mathbf{A}_k^\top$. It follows that,

$$\sum_{n=1}^{N}\widetilde{r}_{nk}\widetilde{\mathbf{M}}_{nk} = \widetilde{r}_k\ \mathbf{A}_k\widetilde{\mathbf{S}}^{\mathrm{x}}_k\mathbf{A}_k^\top + \mathbf{A}_k(\sum_{n=1}^{N}\widetilde{r}_{nk}\widetilde{\boldsymbol{x}}_{nk}\widetilde{\boldsymbol{x}}_{nk}^\top)\mathbf{A}_k^\top - 2(\sum_{n=1}^{N}\widetilde{r}_{nk}(\boldsymbol{y}_n - \boldsymbol{b}_k)\widetilde{\boldsymbol{x}}_{nk}^\top)\mathbf{A}_k^\top \tag{15}$$

and

$$\frac{\sum_{n=1}^{N}\widetilde{r}_{nk}\widetilde{\mathbf{M}}_{nk}}{\widetilde{r}_k} = \mathbf{A}_k(\widetilde{\mathbf{S}}_k^{\mathrm{x}} + \sum_{n=1}^{N}\frac{\widetilde{r}_{nk}}{\widetilde{r}_k}\widetilde{\boldsymbol{x}}_{nk}\widetilde{\boldsymbol{x}}_{nk}^{\top})\mathbf{A}_k^{\top} - 2(\sum_{n=1}^{N}\frac{\widetilde{r}_{nk}}{\widetilde{r}_k}(\boldsymbol{y}_n - \boldsymbol{b}_k)\widetilde{\boldsymbol{x}}_{nk}^{\top})\mathbf{A}_k^{\top} \tag{16}$$

Deriving (see derivatives in Section 3.4) the trace in (13) with regard to $\mathbf{A}_k$ leads to

$$2(\widetilde{\mathbf{S}}_k^{\mathrm{x}} + \sum_{n=1}^{N}\frac{\widetilde{r}_{nk}}{\widetilde{r}_k}\widetilde{\boldsymbol{x}}_{nk}\widetilde{\boldsymbol{x}}_{nk}^{\top})\mathbf{A}_k^{\top}\boldsymbol{\Sigma}_k^{-1} - 2(\sum_{n=1}^{N}\frac{\widetilde{r}_{nk}}{\widetilde{r}_k}(\boldsymbol{y}_n - \boldsymbol{b}_k)\widetilde{\boldsymbol{x}}_{nk}^{\top})\boldsymbol{\Sigma}_k^{-1}$$

Equating this to 0, leads to

$$\widetilde{\mathbf{A}}_k = (\sum_{n=1}^{N}\frac{\widetilde{r}_{nk}}{\widetilde{r}_k}(\boldsymbol{y}_n - \widetilde{\boldsymbol{b}}_k)\widetilde{\boldsymbol{x}}_{nk}^{\top})\ (\widetilde{\mathbf{S}}_k^{\mathrm{x}} + \sum_{n=1}^{N}\frac{\widetilde{r}_{nk}}{\widetilde{r}_k}\widetilde{\boldsymbol{x}}_{nk}\widetilde{\boldsymbol{x}}_{nk}^{\top})^{-1}$$

where $\widetilde{\boldsymbol{b}}_k$ satisfies equation (11). Replacing $\widetilde{\boldsymbol{b}}_k$ in the expression above further leads to a closed-form solution for $\mathbf{A}_k$,

$$\widetilde{\mathbf{A}}_k = (\sum_{n=1}^{N}\frac{\widetilde{r}_{nk}}{\widetilde{r}_k}(\boldsymbol{y}_n - \widetilde{\boldsymbol{y}}_k)\widetilde{\boldsymbol{x}}_{nk}^{\top})\ (\widetilde{\mathbf{S}}_k^{\mathrm{x}} + \sum_{n=1}^{N}\frac{\widetilde{r}_{nk}}{\widetilde{r}_k}\widetilde{\boldsymbol{x}}_{nk}\widetilde{\boldsymbol{x}}_{nk}^{\top} - \widetilde{\boldsymbol{x}}_k\widetilde{\boldsymbol{x}}_k^{\top})^{-1}$$

where

$$\widetilde{\boldsymbol{x}}_k = \sum_{n=1}^{N}\frac{\widetilde{r}_{kn}}{\widetilde{r}_k}\widetilde{\boldsymbol{x}}_{nk}, \tag{17}$$

$$\widetilde{\boldsymbol{y}}_k = \sum_{n=1}^{N}\frac{\widetilde{r}_{kn}}{\widetilde{r}_k}\boldsymbol{y}_n. \tag{18}$$

which can be also written,

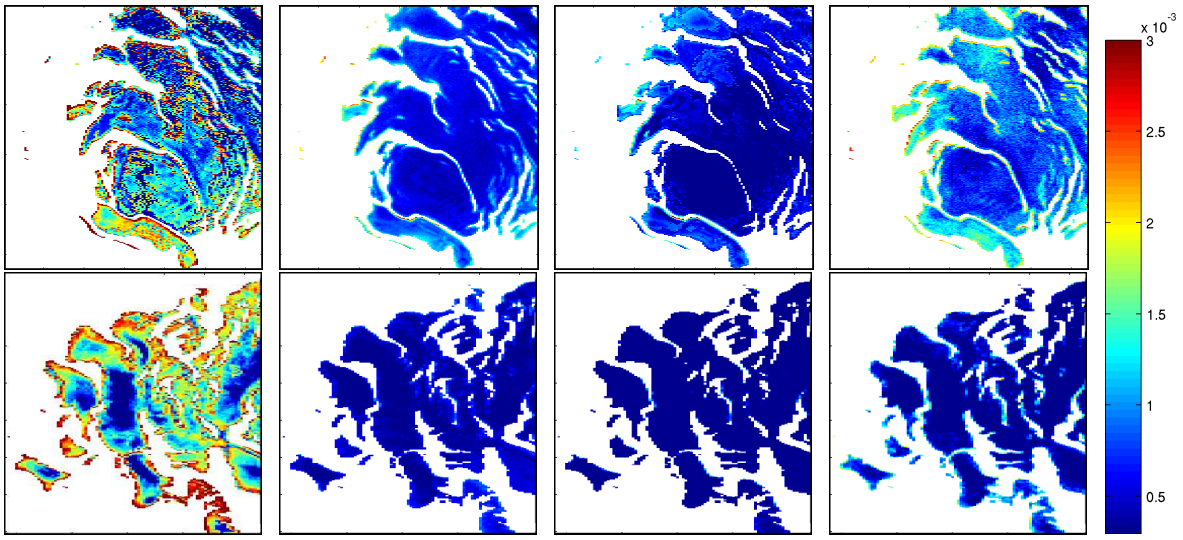
$$\widetilde{\mathbf{A}}_k = (\sum_{n=1}^{N}\frac{\widetilde{r}_{nk}}{\widetilde{r}_k}(\boldsymbol{y}_n - \widetilde{\boldsymbol{y}}_k)(\widetilde{\boldsymbol{x}}_{nk} - \widetilde{\boldsymbol{x}}_k)^{\top})\ (\widetilde{\mathbf{S}}_k^{\mathrm{x}} + \sum_{n=1}^{N}\frac{\widetilde{r}_{nk}}{\widetilde{r}_k}(\widetilde{\boldsymbol{x}}_{nk} - \widetilde{\boldsymbol{x}}_k)(\widetilde{\boldsymbol{x}}_{nk} - \widetilde{\boldsymbol{x}}_k)^{\top})^{-1},$$

since

$$\sum_{n=1}^{N}\frac{\widetilde{r}_{nk}}{\widetilde{r}_k}(\boldsymbol{y}_n - \widetilde{\boldsymbol{y}}_k)\widetilde{\boldsymbol{x}}_k^{\top} = 0$$

and

$$\sum_{n=1}^{N}\frac{\widetilde{r}_{nk}}{\widetilde{r}_k}(\widetilde{\boldsymbol{x}}_{nk} - \widetilde{\boldsymbol{x}}_k)(\widetilde{\boldsymbol{x}}_{nk} - \widetilde{\boldsymbol{x}}_k)^{\top} = \sum_{n=1}^{N}\frac{\widetilde{r}_{nk}}{\widetilde{r}_k}\widetilde{\boldsymbol{x}}\widetilde{\boldsymbol{x}}_{nk}^{\top} - \widetilde{\boldsymbol{x}}_k\widetilde{\boldsymbol{x}}_k^{\top}.$$



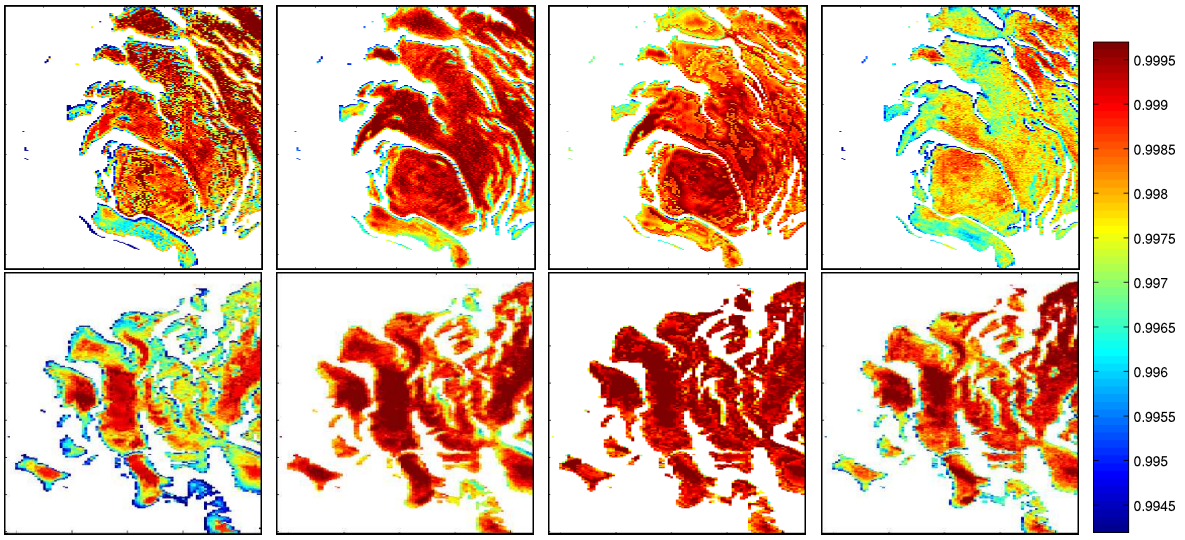
(a) hGLLiM-2 (b) RVM (c) MLE (d) JGMM

Fig. 1 Proportion of dust obtained with our method (hGLLiM-2) and with three other methods. The data correspond to hyperspectral images grabbed from two different viewpoints of the South polar cap of Mars. First row: orbit 41, second row: orbit 61. White areas correspond to unexamined regions, where the synthetic model does not apply.

2 Retrieval of Mars surface physical properties from hyperspectral images

Visible and near infrared imaging spectroscopy is a key remote sensing technique used to study and monitor planets. It records the visible and infrared light reflected from the planet in a given wavelength range and produces cubes of data where each observed surface location is associated with a spectrum. Physical properties of the planets' surface, such as chemical composition, granularity, texture, etc, are some of the most important parameters that characterize the morphology of spectra. In the case of Mars, radiative transfer models have been developed to numerically evaluate the link between these parameters and observable spectra. Such models allow to simulate spectra from a given set of parameter values. In practice, the goal is to scan the Mars ground from an orbit in order to observe gas and dust in the atmosphere and look for signs of specific materials such as silicates, carbonates and ice at the surface. We are thus interested in solving the associate inverse problem which is to deduce physical parameter values from the observed spectra.

The Hybrid GLLiM model with a 2-dimensional latent variable (hGLLiM-2) and three other methods, namely RVM, MLE and JGMM (see Section 6.3 in the main paper for details) were used to retrieve the physical properties of the South polar cap using two hyperspectral images of approximately the same area from different view points (orbit 41 and orbit 61). Since we are looking for proportions between 0 and 1, values smaller than 0 or higher than 1 are not acceptable and hence they were set to one of the bounds. As it can be seen in Fig. 1 and Fig. 2, hybrid GLLiM outputs proportion maps with similar characteristics for the two view points, which suggests good consistency. Such a consistency is not observed using the other tested methods. In addition, RVM and and MLE output a much higher number of values falling outside the interval $[0, 1]$. Moreover, hGLLiM-2 is the only method featuring less dust at the South pole cap center and higher concentrations of dust at the boundaries of the CO2 ice, which matches expected results from planetology [Douté et al., 2005]. Finally, note that the proportions of CO2 ice and dust clearly seem to be complementary using hGLLiM-2, while this complementarity is less obvious using other methods.



(a) hGLLiM-2 (b) RVM (c) MLE (d) JGMM

Fig. 2 Proportion of CO2 ice obtained from hyperspectral images of two different viewpoints of the South polar cap of Mars. First row: orbit 41, second row: orbit 61. White areas correspond to unexamined regions, where the synthetic model does not apply.

3 Some properties of Gaussian distributions and useful results

3.1 Product of two multivariate Gaussian densities

Lemma 1 *The product of two multivariate Gaussian densities* $\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)$ *and* $\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$ *is proportional to the multivariate Gaussian density with mean* $\boldsymbol{\mu}_3$ *and covariance matrix* $\boldsymbol{\Sigma}_3$ *given by:*

$$\begin{aligned}\boldsymbol{\mu}_3 &= \boldsymbol{\Sigma}_3\big(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\big)\\ \boldsymbol{\Sigma}_3 &= \big(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\big)^{-1}\end{aligned} \tag{19}$$

Proof: The result follows easily by developing the terms in the sum of the two Gaussians quadratic forms and rearranging them into a single quadratic form over the variable $\boldsymbol{x}$. Note however that the product of two Gaussian densities is not normalized.

3.2 Mean of square forms

Lemma 2 *If* $\boldsymbol{X}$ *is a variable with mean* $\boldsymbol{\mu}$ *and covariance matrix* $\boldsymbol{\Sigma}$*, then,*

$$E[\boldsymbol{X}^\top\mathbf{M}\boldsymbol{X}] = \boldsymbol{\mu}^\top\mathbf{M}\boldsymbol{\mu} + \mathrm{tr}(\mathbf{M}\boldsymbol{\Sigma})$$

Proof: The proof is omitted as this is a very standard result. See for instance [Schott, 1997] p.391

3.3 Maximum likelihood for a Gaussian sample

As shown in [Schott, 1997] p.347, when estimating using maximum likelihood principle, the mean and variance from a normally distributed sample, it comes a maximization problem similar to that in the following lemma.

Lemma 3 *The* $(m \times m)$ *matrix* $\mathbf{A}$ *minimizing* $\mathrm{tr}(\mathbf{M}\mathbf{A}^{-1}) + \alpha \log |\mathbf{A}|$ *where* $\mathbf{M}$ *is a* $m \times m$ *symmetric definite positive matrix and* α *is a positive real number is* $\mathbf{A} = \frac{\mathbf{M}}{\alpha}$.

Proof: See [Schott, 1997] p.347 for a similar proof.

3.4 Derivatives of traces

It can be found for instance in [Schott, 1997] the following results regarding the derivatives of the following two expressions.

Lemma 4

$$\frac{\partial}{\partial \mathbf{A}} \mathrm{tr}(\mathbf{M}\mathbf{A}^\top) = \mathbf{M}$$
$$\frac{\partial}{\partial \mathbf{A}} \mathrm{tr}(\mathbf{A}\mathbf{B}\mathbf{A}^\top\mathbf{C}) = \mathbf{C}\mathbf{A}\mathbf{B} + \mathbf{C}^\top\mathbf{A}\mathbf{B}^\top$$

References

S. Douté, B. Schmitt, J-P. Bibring, Y. Langevin, F. Altieri, G. Bellucci, B. Gondet, and the MEX OMEGA team. Nature and composition of the icy terrains of the south pole of Mars from MEX OMEGA observations. In *The 36th Lunar and Planetary Science Conference, (Lunar and Planetary Science XXXVI)*, March 2005.

J. R. Schott. *Matrix Analysis for Statistics*. Wiley series in probability and statistics, 1997.