Supplementary Material of the paper: "Multi-Channel Stochastic Variational Inference for the Joint Analysis of Heterogeneous Biomedical Data in Alzheimer's Disease"

Luigi Antelmi[1], Nicholas Ayache[1], Philippe Robert[2,3], and Marco Lorenzi[1] for the Alzheimer's Disease Neuroimaging Initiative *

[1] University of Côte d'Azur, Inria Sophia Antipolis, Epione Research Project, France
[2] University of Côte d'Azur, CoBTeK, France
[3] Centre Memoire, CHU of Nice, France

Derivation of the Lower Bound

In the following derivation we will interchangeably use $\mathbf{x}$ and $\mathbf{x}_1, \ldots, \mathbf{x}_C$ to leave the notation uncluttered. For the same reason we will omit the variational and generative parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$.

Variational inference is carried out by introducing a set of probability density functions $q(\mathbf{z}|\mathbf{x}_c)$, belonging to a distribution family $\mathcal{Q}$, that are on average as close as possible to the true posterior over the latent variable $p(\mathbf{z}|\mathbf{x})$. In other words we aim to solve the following minimization problem:

$$\underset{q \in \mathcal{Q}}{arg\,min}\ \mathbb{E}_c \left[\mathcal{D}_{KL}\big(q(\mathbf{z}|\mathbf{x}_c) \,||\, p(\mathbf{z}|\mathbf{x}_1, \ldots, \mathbf{x}_C)\big) \right] \tag{1}$$

Given the intractability of $p(\mathbf{z}|\mathbf{x})$ for most complex models, we cannot solve directly this optimization problem. We look then for an equivalent problem, by rearranging the objective:

$$\begin{aligned} &\mathbb{E}_c \left[\mathcal{D}_{KL}\big(q(\mathbf{z}|\mathbf{x}_c) \,||\, p(\mathbf{z}|\mathbf{x})\big) \right] = \mathbb{E}_c \left[\int_{\mathbf{z}} q(\mathbf{z}|\mathbf{x}_c) \left(\ln q(\mathbf{z}|\mathbf{x}_c) - \ln p(\mathbf{z}|\mathbf{x}) \right) d\mathbf{z} \right] \\ &= \mathbb{E}_c \left[\int_{\mathbf{z}} q(\mathbf{z}|\mathbf{x}_c) \left(\ln q(\mathbf{z}|x_c) - \ln p(\mathbf{x}|\mathbf{z}) - \ln p(\mathbf{z}) + \ln p(\mathbf{x}) \right) d\mathbf{z} \right] \\ &= \ln p(\mathbf{x}) + \mathbb{E}_c \left[\mathcal{D}_{KL}\big(q(\mathbf{z}|\mathbf{x}_c) \,||\, p(\mathbf{z})\big) - \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c)} \left[\ln p(\mathbf{x}|\mathbf{z}) \right] \right] \end{aligned} \tag{2}$$

where in the middle line we use the Bayes' theorem to factorize the true posterior $p(\mathbf{z}|\mathbf{x})$. Now, we can reorganize the terms, such that:

---

* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. A detailed list of funding actors can be found in the *Acknowledgments* section of the *Supplementary Material* .

$$\begin{aligned} &\ln p\left(\mathbf{x}\right) - \underbrace{\mathbb{E}_c\left[\mathcal{D}_{KL}\big(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}|\mathbf{x}\right)\big)\right]}_{\geq 0} = \\ &= \underbrace{\mathbb{E}_c\left[\mathbb{E}_{q(\mathbf{z})}\left[\ln p\left(\mathbf{x}_1,\ldots,\mathbf{x}_C|\mathbf{z}\right)\right] - \mathcal{D}_{KL}\big(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}\right)\big)\right]}_{\text{lower bound } \mathcal{L}} \end{aligned} \tag{3}$$

Since the KL term in the left hand side is always non-negative, the right hand side lower bounds the log evidence. By maximizing the lower bound we obtain the result of maximizing the data log evidence while solving the minimization problem in (1).

The hypothesis that every channel is conditionally independent from all the others given $\mathbf{z}$, allows to factorize the data likelihood as $p\left(\mathbf{x}_1,\ldots,\mathbf{x}_C|\mathbf{z}\right) = \prod_{i=1}^{C} p\left(\mathbf{x}_i|\mathbf{z}\right)$, so that the lower bound becomes:

$$\mathcal{L} = \mathbb{E}_c\left[\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c)}\left[\sum_{i=1}^{C}\ln p\left(\mathbf{x}_i|\mathbf{z}\right)\right] - \mathcal{D}_{KL}\big(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}\right)\big)\right] \tag{4}$$

Finally, assuming every channel is equally likely to be observed with probability $1/C$, we can rewrite equation (3) as:

$$\ln p\left(\mathbf{x}_1,\ldots,\mathbf{x}_C\right) \geq \frac{1}{C}\sum_{c=1}^{C}\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c)}\left[\sum_{i=1}^{C}\ln p\left(\mathbf{x}_i|\mathbf{z}\right)\right] - \mathcal{D}_{KL}\big(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}\right)\big) \tag{5}$$

Data generation procedure

Datasets $\mathbf{x} = \{\mathbf{x}_c\}$ with $c = 1..C$ channels where created according to the following model:

$$\begin{aligned} &\mathbf{z} \sim \mathcal{N}\left(\mathbf{0};\mathbf{I}_l\right) \\ &\boldsymbol{\epsilon} \sim \mathcal{N}\left(\mathbf{0};\mathbf{I}_{d_c}\right) \\ &\mathbf{G}_c = diag\left(\mathbf{R}_c\mathbf{R}_c^T\right)^{-1/2}\mathbf{R}_c \\ &\mathbf{x}_c = \mathbf{G}_c\mathbf{z} + snr^{-1/2}\cdot\boldsymbol{\epsilon} \end{aligned} \tag{6}$$

where for every channel c, $\mathbf{R}_c \in \mathbb{R}^{d_c\times l}$ is a random matrix with l orthonormal columns (*i.e.,* $\mathbf{R}_c^T\mathbf{R}_c = \mathbf{I}_l$), $\mathbf{G}_c$ is the linear generative law, and snr is the signal-to-noise ratio. It's easy to demonstrate that the diagonal elements of the covariance matrix of $\mathbf{x}_c$ are inversely proportional to snr, *i.e.,*$diag\left(\mathbb{E}\left[\mathbf{x}_c\mathbf{x}_c^T\right]\right) = (1+snr^{-1})\mathbf{I}_{d_c}$. Scenarios where generated by varying one-at-a-time the dataset attributes, as listed in Tab. 1.

Acknowledgments

This work has been supported by:

– the French government, through the UCA$^{\text{JEDI}}$ Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-01;

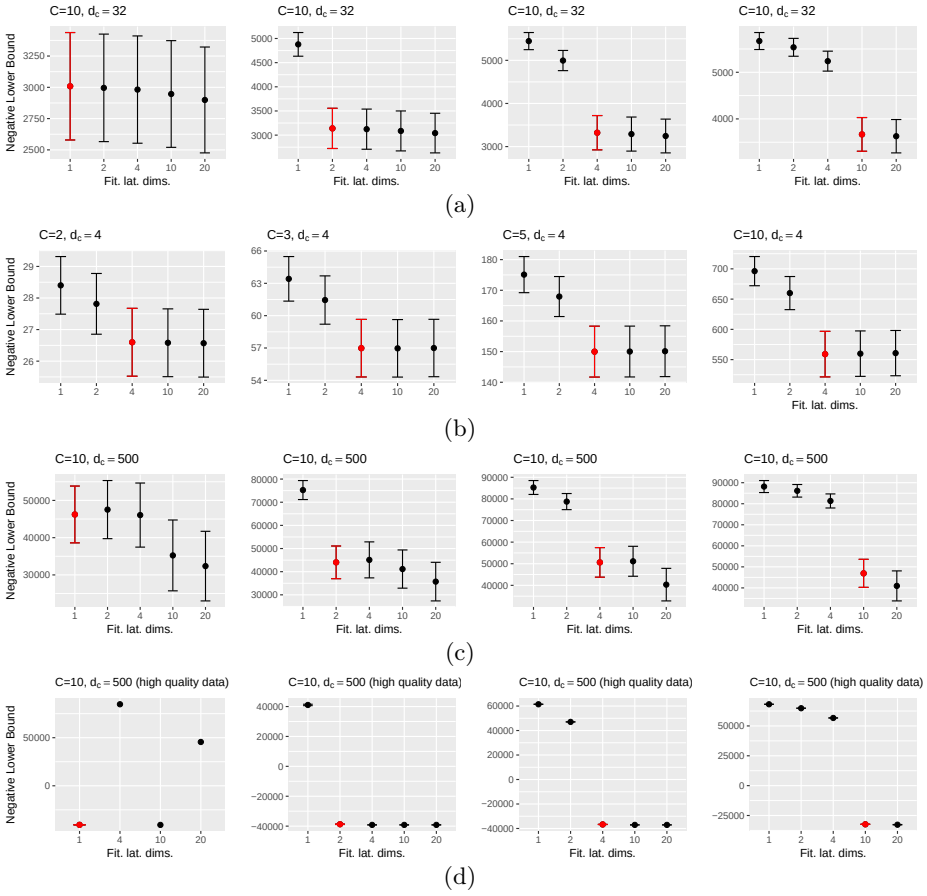
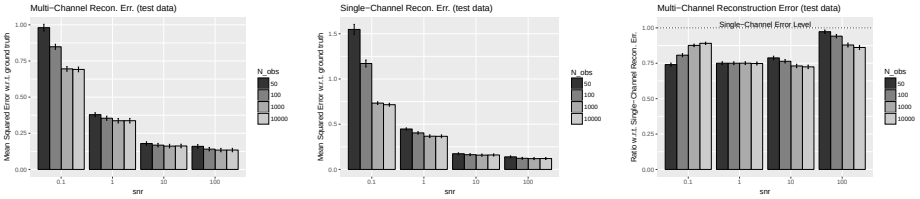


Fig. 1: Negative lower bound (NLB) on the synthetic training set computed at convergence for all the scenarios. Each bar shows mean $\pm$ std.err. of $N = 80$ total experiments as a function of the number of fitted latent dimensions. Red bars represents experiments where the number of true and fitted latent dimensions coincide. (top) Experimental setup $C = 10$, $d_c = 32$: NLB stops decreasing when the number of fitted latent dimension coincide with the generated ones; notable gap between the under-fitted and over-fitted experiments (elbow effect). (2^{nd} row) Experimental setup $d_c = 4$, $l = 4$: increasing the number of channels C makes the elbow effect more pronounced. (3^{rd} row) Experimental setup $C = 10$, $d_c = 500$: with high dimensional data ($d_c = 500$) using the lower bound as a model selection criteria to assess the true number of latent dimensions may end up in overestimation. (bottom) Restricted ($N = 5$ total experiments) high quality experimental setup $C = 10$, $d_c = 500$, $S = 10000$, $snr = 100$: the risk to overestimate the true number of latent dimensions can be mitigated by increasing the snr and S of the observations in the dataset.

Table 1: Dataset attributes, varied one-at-a-time in the prescribed ranges, and used to generate scenarios according to Eq. (6).

Attribute description	range / iteration list	symbol
Total channels	2, 3, 5, 10	C
Channel dimension	4, 8, 16, 32, 500	d_c
Latent space dimension	1, 2, 4, 10, 20	l
Number of samples/observations	50, 100, 1000, 10000	S
Signal-to-noise ratio	100, 10, 1, 0.1	snr
Replication number (re-initialize $\mathbf{R}_c$)	1, 2, 3, 4, 5	



(a) (b) (c)

Fig. 2: Reconstruction error on synthetic test data reconstructed with the multi-channel model specified in the main paper, Section 2.2. The reconstruction is better for high snr and high training data sample size. Scenarios where generated by varying one-at-a-time the dataset attributes listed in Tab. 1 for a total of 8 000 experiments. (a) Mean squared error from the ground truth test data using the Multi-Channel reconstruction: $\hat{\mathbf{x}} = \mathbb{E}_c \left[\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c)} \left[p\left(\mathbf{x}|\mathbf{z}\right) \right] \right]$ (b) Mean squared error from the ground truth test data using the Single-Channel reconstruction: $\hat{\mathbf{x}} = \mathbb{E}_{q(\mathbf{z}|\mathbf{x})} \left[p\left(\mathbf{x}|\mathbf{z}\right) \right]$ (c) Ratio between Multi- *vs* Single-Channel reconstruction errors: we notice that the error made in ground truth data recovery with multi-channel information is systematically lower than the one obtained with a single-channel decoder.

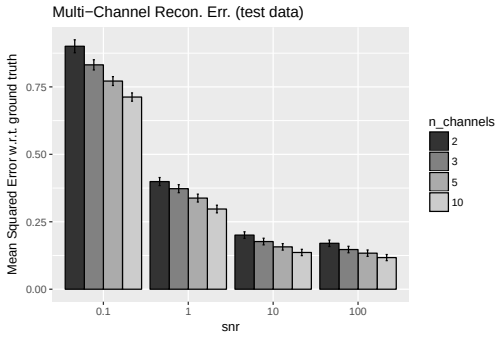
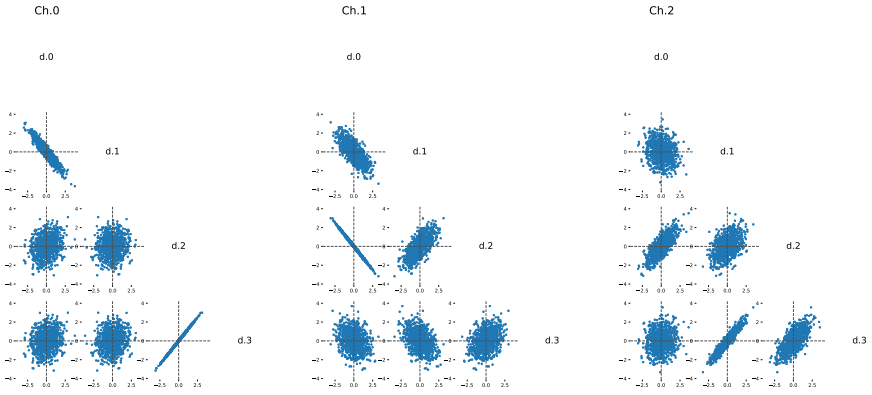
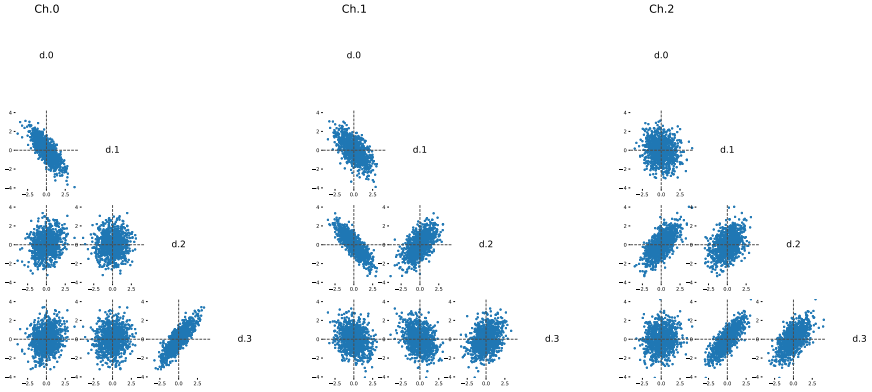


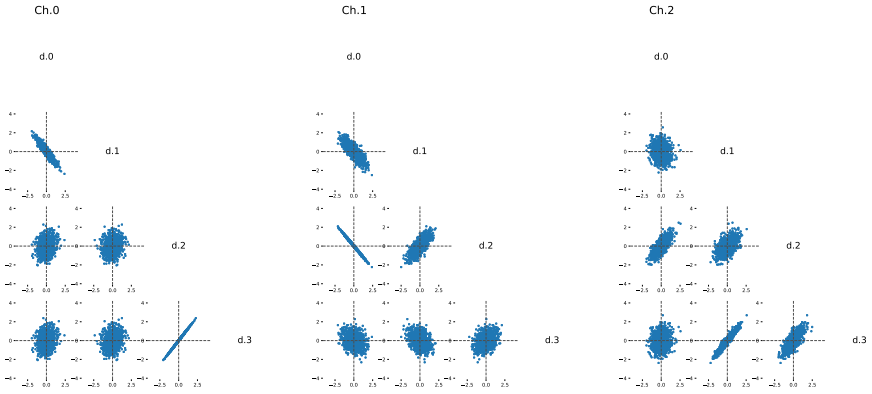
Fig. 3: Incrasing the snr and the number of channels ameliorate the ground truth data recovery. Reconstruction done as in Fig. 2a.



(a) Ground truth



(b) Noisy observations ($snr = 5$)



(c) Reconstruction

Fig. 4: Pairwise representation of the four dimensions d of three data channels Ch, generated from a two-dimensional latent dimension $\mathbf{z} \sim \mathcal{N}(\mathbf{0}; \mathbf{I})$, according to Eq. (6). Noisy data was fitted with our model with a linear reparameterization. (a) Ground truth ($snr = 0$). (b) Observations used to fit the multi-channel model ($snr = 5$). (c) Data generated from the latent variable inferred from the noisy data.

– the grant AAP Santé 06 2017-260 DGA-DSH, and by the Inria Sophia Antipolis - Méditerranée, "NEF" computation cluster.
– Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (`www.fnih.org`). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for NeuroImaging at the University of Southern California.