AGU PUBLICATIONS

Journal of Geophysical Research – Solid Earth

Supporting Information for

Rapid Discriminative Variational Bayesian Inversion of Geophysical Data for the Spatial Distribution of Geological Properties

M. A. Nawaz[1] and A. Curtis[1,2]

1 School of Geosciences, Grant Institute, University of Edinburgh, Edinburgh, United Kingdom.

2 Exploration and Environmental Geophysics Group, ETH Zurich, Switzerland.

Corresponding author: Muhammad Atif Nawaz (Muhammad.AtifNawaz@ed.ac.uk)

Contents of this file

S1 – Mathematical derivation of equation (8) in the main text
S2 – Mathematical derivation of equation (13) in the main text

Introduction

The supplementary information provides mathematical derivations of equations (8) and (13) in the main text.

Text S1 – Mathematical derivation of equation (8) in the main text

The KL divergence between the approximate distribution $\mathcal{Q}(\boldsymbol{m}|\boldsymbol{d})$ and the true distribution $\mathcal{P}(\boldsymbol{m}|\boldsymbol{d};\boldsymbol{w})$ as defined in equation (7) in the main text may be written as

$$KL\big(\mathcal{Q}(\boldsymbol{m}|\boldsymbol{d})||\mathcal{P}(\boldsymbol{m}|\boldsymbol{d};\boldsymbol{w})\big)$$

$$= \int_{\boldsymbol{m}} \mathcal{Q}(\boldsymbol{m}|\boldsymbol{d}) \log \frac{\mathcal{Q}(\boldsymbol{m}|\boldsymbol{d})}{\mathcal{P}(\boldsymbol{m}|\boldsymbol{d};\boldsymbol{w})}\, d\boldsymbol{m}$$

$$= \int_{\boldsymbol{m}} \mathcal{Q}(\boldsymbol{m}|\boldsymbol{d}) \log \mathcal{Q}(\boldsymbol{m}|\boldsymbol{d})\, d\boldsymbol{m} - \int_{\boldsymbol{m}} \mathcal{Q}(\boldsymbol{m}|\boldsymbol{d}) \log \mathcal{P}(\boldsymbol{m}|\boldsymbol{d};\boldsymbol{w})\, d\boldsymbol{m} \qquad \text{S-1}$$

The first term here represents negative of the entropy $\mathcal{S}(\mathcal{Q})$ of the variational distribution $\mathcal{Q}(\boldsymbol{m}|\boldsymbol{d})$. Now, substituting for $\mathcal{P}(\boldsymbol{m}|\boldsymbol{d};\boldsymbol{w})$ from equation (5) in the main text we get

$$= -\mathcal{S}(\mathcal{Q}) - \int_{\boldsymbol{m}} \mathcal{Q}(\boldsymbol{m}|\boldsymbol{d}) \left(\sum\nolimits_{\hat{c}\in\hat{\mathcal{C}}} \boldsymbol{w}^{\mathrm{T}} \boldsymbol{f}(\boldsymbol{m}_{\hat{c}}, \boldsymbol{d}) - \log \mathcal{Z}(\boldsymbol{d};\boldsymbol{w})\right) d\boldsymbol{m} \qquad \text{S-2}$$

Defining $\mathcal{L}(\boldsymbol{w}; d) \equiv \log \mathcal{Z}(\boldsymbol{d};\boldsymbol{w})$ as the log-evidence, which is independent of $\mathcal{Q}(\boldsymbol{m}|\boldsymbol{d})$, we get

$$= -\mathcal{S}(\mathcal{Q}) - \mathbb{E}_{\mathcal{Q}}\left[\sum\nolimits_{\hat{c}\in\hat{\mathcal{C}}} \boldsymbol{w}^{\mathrm{T}} \boldsymbol{f}(\boldsymbol{m}_{\hat{c}}, \boldsymbol{d})\right] + \mathcal{L}(\boldsymbol{w};\boldsymbol{d}) \qquad \text{S-3}$$

where $\mathbb{E}_{\mathcal{Q}}[\cdot]$ represents the expectation operator with respect to $\mathcal{Q}(\boldsymbol{m}|\boldsymbol{d})$.

Now define the variational free energy $\mathcal{F}(\mathcal{Q},\boldsymbol{w})$ as (equation (9) in the main text)

$$\mathcal{F}(\mathcal{Q},\boldsymbol{w}) = \mathbb{E}_{\mathcal{Q}}\left(\sum\nolimits_{\hat{c}\in\hat{\mathcal{C}}} \boldsymbol{w}^{\mathrm{T}} \boldsymbol{f}(\boldsymbol{m}_{\hat{c}}, \boldsymbol{d})\right) + \mathcal{S}(\mathcal{Q}) \qquad \text{S-4}$$

Substituting equation S-4 into equation S-3 we get

$$\mathcal{L}(\boldsymbol{w};\boldsymbol{d}) = \mathcal{F}(\mathcal{Q},\boldsymbol{w}) + KL\big(\mathcal{Q}(\boldsymbol{m}|\boldsymbol{d})||\mathcal{P}(\boldsymbol{m}|\boldsymbol{d};\boldsymbol{w})\big) \qquad \text{S-5}$$

which is same as equation (8) in the main text.

Text S2 – Mathematical derivation of equation (13) in the main text

To derive equation (13) from equation (12) in the main text we substitute equation (12) into equation (9) in the main text which gives

$$\mathcal{F}(\mathcal{Q},\boldsymbol{w}) = \sum_{\hat{c}\in\hat{\mathcal{C}}} \mathbb{E}_{\mathcal{Q}}[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{m}_{\hat{c}},\boldsymbol{d})] + \sum_{c} \mathcal{S}(\mathcal{Q}_c)$$

$$= \sum_{\hat{c}\in\hat{\mathcal{C}}} \int_{\boldsymbol{m}_{\hat{c}}} \left(\prod_{c\subset\hat{c}} \mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d})\right) \boldsymbol{w}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{m}_{\hat{c}},\boldsymbol{d}) d\boldsymbol{m}_{\hat{c}} - \sum_{c} \int_{\boldsymbol{m}_c} \mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d}) \log \mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d})\, d\,\boldsymbol{m}_c \qquad \text{S-6}$$

The maximum computational complexity of the expectation term (the 1st term in equation S-6) is $O\left(|\hat{\mathcal{C}}| * |\boldsymbol{m}_{\hat{c}}|\right)$ and of the entropy term (the 2nd term in equation S-6) is $O(|\mathcal{C}| * |\boldsymbol{m}_c|)$, where $|\hat{\mathcal{C}}|$ and $|\mathcal{C}|$ are respectively the number of maximal cliques $\hat{c}$ and the approximating cliques c in the graph, and $|\boldsymbol{m}_{\hat{c}}|$ and $|\boldsymbol{m}_c|$ are the maximum dimensionality of model parameters in the maximal and the approximating cliques in the graph, respectively. Thus evaluation of the free energy functional in the above equation can be performed in time that is linear in the maximum dimensionality of factors in $\mathcal{Q}(\boldsymbol{m}|\boldsymbol{d})$ (or cliques in the graph).

In order to optimize the marginal distributions $\mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d})$ of any restricted variational distribution $\mathcal{Q}(\boldsymbol{m}|\boldsymbol{d}) \in \mathbb{Q}$, we iteratively maximize the variational free energy $\mathcal{F}(\mathcal{Q},\boldsymbol{w})$ within the family $\mathbb{Q}$ of factorizable distributions by successively optimizing each of the marginal distributions at a time. The factorizable form of the mean-field (MF) variational distribution allows successive optimization of each factor (marginal distribution) while keeping others fixed in an iterative fashion. We may characterize stationary points of the marginal distribution $\mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d})$ in terms of the rest of the marginals $\mathcal{Q}_{\backslash c}(\boldsymbol{m}_{\backslash c}|\boldsymbol{d})$ by restricting the energy functional $\mathcal{F}(\mathcal{Q},\boldsymbol{w})$ to the terms involving $\mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d})$, which gives

$$\mathcal{F}(\mathcal{Q}_c,\boldsymbol{w}) = \sum_{\hat{c}\in\hat{\mathcal{C}}} \mathbb{E}_{\mathcal{Q}}[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{m}_{\hat{c}},\boldsymbol{d})] + \mathcal{S}(\mathcal{Q}_c) + \text{ constant} \qquad \text{S-7}$$

We seek to derive update equations for a higher-order MF approximation by characterizing the stationary points of the free energy functional using Lagrange multipliers. Since $\mathcal{F}(\mathcal{Q}_c,\boldsymbol{w})$ is concave in $\mathcal{Q}_c$, we can maximize it by forming a Lagrangian as

$$L_c(\mathcal{Q}) = \sum_{\hat{c}\in\hat{\mathcal{C}}} \mathbb{E}_{\mathcal{Q}}[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{m}_{\hat{c}},\boldsymbol{d})] + \mathcal{S}(\mathcal{Q}_c) + \gamma\left(1 - \int_{\boldsymbol{m}_c} \mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d})\, d\boldsymbol{m}_c\right) \qquad \text{S-8}$$

where the Lagrange multiplier γ enforces the constraint that the marginal $\mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d})$ is a proper distribution. Differentiating $L_c(\mathcal{Q})$ with respect to $\mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d})$ and setting it equal to zero gives:

$$\mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d}) = \frac{1}{\mathcal{Z}_c(\boldsymbol{d})} \exp\left\{\sum_{\hat{c}\in\hat{\mathcal{C}}} \mathbb{E}_{\mathcal{Q}_{\backslash c}}[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{m}_{\hat{c}},\boldsymbol{d})|\, \boldsymbol{m}_c]\right\} \qquad \text{S-9}$$

as the necessary and sufficient condition for $\mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d})$ to be a local maximum of $\mathcal{F}(\mathcal{Q}_c, \boldsymbol{w})$ given the rest of the marginal distributions $\mathcal{Q}_{\backslash c}(\boldsymbol{m}_{\backslash c}|\boldsymbol{d})$. In this equation $\mathcal{Z}_c(\boldsymbol{d})$ is the local normalization constant that ensures that $\mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d})$ is a valid distribution, and the conditional expectation in the argument of the exponential function is given by

$$\mathbb{E}_{\mathcal{Q}_{\backslash c}}[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{m}_{\hat{c}}, \boldsymbol{d})|\, \boldsymbol{m}_c] = \int_{\boldsymbol{m}_{c'}} \left(\prod\nolimits_{c'} \mathcal{Q}_{c'}(\boldsymbol{m}_{c'}|\boldsymbol{d})\right) \boldsymbol{w}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{m}_{\hat{c}}, \boldsymbol{d})\, d\boldsymbol{m}_{c'} \quad \text{S-10}$$

where $c' \in \hat{c}\backslash\{c\} \wedge |c'| = q$. The above conditional expectation in equation S-10 is independent of the variational marginal distribution $\mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d})$, but is a function of $\boldsymbol{m}_c$ as a conditioning variable. It may be restricted to terms that involve c by exploiting the conditional independence of $\boldsymbol{m}_c$ given $\boldsymbol{d}$ under the MF approximation, resulting in a closed-form update for the marginal distribution $\mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d})$:

$$\mathcal{Q}_c(\boldsymbol{m}_c|\boldsymbol{d}) \leftarrow \frac{1}{\mathcal{Z}_c(\boldsymbol{d})} \exp\left\{\sum\nolimits_{\hat{c}\in\hat{\mathcal{C}}:c\subset\hat{c}} \mathbb{E}_{\mathcal{Q}_{\backslash c}}[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{m}_{\hat{c}}, \boldsymbol{d}) \mid \boldsymbol{m}_c]\right\} \quad \text{S-11}$$

which is same as equation (13) in the main text.