FAST BILINEAR EXTRAPOLATION OF 3D ISING FIELD PARTITION FUNCTION.
APPLICATION TO fMRI IMAGE ANALYSIS.

Laurent Risser¹,²,³, Jérôme Idier¹, Philippe Ciuciu²,³ and Thomas Vincent²,³

¹ IRCCyN/CNRS, 1 rue de la Noë, 44300 Nantes, France
² NeuroSpin/CEA, 91191 Gif-sur-Yvette, France
³ IFR 49, Institut d’Imagerie Neurofonctionnelle, Paris, France

firstname.lastname@irccyn.ec-nantes.fr, ² firstname.lastname@cea.fr

ABSTRACT

Symmetric Ising models define the simplest discrete Markov random fields that can be used for segmentation purpose. Unsupervised segmentation requires an automatic setting of the temperature parameter of Ising fields. To this end, partition function (PF) estimation becomes a key issue. In this paper, we present a bilinear extrapolation technique for a fast PF estimation of 3D Ising field. The proposed method is a two-step procedure that applies to the context where multiple 3D Ising fields are involved over different objects (eg, brain regions) of different size and topology. First, a small set of reference PFs is accurately estimated using path sampling. Second, the large remaining set of PFs is computed using a temperature-dependent bilinear extrapolation technique. It is shown that our approach is accurate and computationally efficient to account for topological fluctuations of Ising fields on regular and irregular graphs. A convincing application to joint detection-estimation of brain activity in functional MRI is also presented.

Index Terms— Ising field, Partition function, Fast scheme, Path sampling, 3D images.

1. INTRODUCTION

In medical image analysis, one is often interested in recovering spatial structures. A simple but suboptimal approach to enhance the signal-to-noise ratio consists in spatially filtering the datasets at the expense of a loss of spatial resolution. A more challenging approach proceeds by introducing some prior knowledge on the sought spatial structure and therefore deals with the original unsmoothed dataset. Spatial information is usually embedded in local interaction models such as Markov Random Fields (MRFs). Discrete MRFs, whose definition relies on one or several hyper-parameters, are of particular interest for segmentation or classification purposes.

For instance, temperature level controls the amount of spatial correlation in symmetric Ising models. To obtain user-independent results, one has to automatically tune the MRF parameters. For Ising fields, this amounts to estimating a single temperature level. This requires a precise estimation of the so-called Partition Function (PF), i.e., the normalization constant of the MRF. In this paper, we focus on fast schemes for estimating PF of 3D Ising models. More precisely, the present work brings a computationally efficient solution to the multiple PF estimation problem, where a small subset of reference PFs has been already estimated and a large set of PFs remains to approximate. This situation typically arises in functional Magnetic Resonance Imaging (fMRI) to recover activation clusters in different regions of the brain [1], each region being of particular size and topology and thus involving a 3D Ising model on a specific graph.

The rest of the paper is organized as follows. Section 2 is dedicated to single 3D Ising PF estimation. The main contribution is presented in Section 3, where the extension to the multiple PF estimation problem is formulated and addressed using a bilinear extrapolation scheme. An application of this work to the recovering of brain activations in fMRI is finally presented in Section 4.

2. PROBLEM STATEMENT

Let us consider a grid characterized by a set of sites \( s = (s_i)_{i=1,n} \). A binary label \( q_i \in \{0,1\} \) is associated to each site \( s_i \). A pair of adjacent sites \( s_i \) and \( s_j \) (\( i \neq j \)) is denoted \( i \sim j \) and is called a clique \( c \). The set of all cliques allows us to define an undirected graph denoted \( G \). Let \( q = (q_1,q_2,\cdots,q_n) \in \{0,1\}^n \) be the set of binary labels associated to \( s \). In what follows, we assume \( q \) to be distributed according to a symmetric Ising model:

\[
\Pr(q|\beta) = Z(\beta)^{-1} \exp(\beta U(q)),
\]

where

\[
U(q) = \sum_{i\sim j} I(q_i=q_j)
\]

is the global “negative energy” and \( I(A) = 1 \) whenever \( A \) is true and 0 otherwise. The inverse temperature \( \beta \geq 0 \) controls the amount of spatial correlation between the components of \( q \) according to \( G \).
At $\beta = 0$, the field $q$ is made up of spatially independent labels while it defines a MRF whenever $\beta > 0$: the larger the $\beta$-value the more likely configurations containing clusters of identical labels $q_i$ are. The partition function $Z(\beta)$ reads:

$$Z(\beta) = \sum_{q \in \{0, 1\}^n} \exp(\beta U(q))$$

and depends on the geometry of $G$. Its exact evaluation in a reasonable amount of time is impossible except on tiny grids. For instance, this already takes several hours on a modern computer for the small grid depicted in Fig. 1, which contains 27 sites and 54 cliques. It requires the evaluation of $U(q)$ for all configurations of $\{0, 1\}^n$, whose number is exponentially related to $n$. Robust and fast estimation of $Z(\beta)$ is thus a key issue for numerous 3D medical imaging problems involving Ising models and more generally discrete MRFs.

![Fig. 1](image)

Regular grid of size $3 \times 3 \times 3$ with a 6-connectivity system. The points and lines respectively represent the sites and the cliques.

3. EXTRAPOLATION SCHEMES FOR PARTITION FUNCTION ESTIMATION

3.1. Single partition function estimation

Several approaches have been designed to estimate a single PF [2, 3]. Path sampling is an efficient method deriving from importance sampling [3], which consists in sampling unusual distributions using an easy-to-sample instrumental distribution and a correction factor. In this work, robust PF estimates were obtained using path sampling; see [4] for details. Time dedicated to a single PF estimation using this method is acceptable but becomes penalizing when numerous PFs need to be estimated as required when dealing with several hundreds of grids of variable size and shape. Since this typical situation occurs in our fMRI application, a fast compromisise consists in resorting to path sampling to get log-scale estimates $(\log Z_{G_p}(\beta))_{p=1:P}$ in a small subset of reference graphs $(G_p)_{p=1:P}$ and then in using extrapolation formulas to obtain $\log Z_T(\beta)$ for the large remaining set of brain regions to be analyzed. In what follows, for the sake of notational simplicity this large set of graphs is referenced by a test graph $T$. All tests were performed on 3D Ising fields with a 6-connectivity system. However, they may be easily extended to other neighborhood systems as well as to Potts fields.

3.2. Partition function extrapolation schemes

**Linear extrapolation.** In [4], the authors have developed the following linear interpolation procedure: 1) For a set of reference grids or graphs $(G_p)_{p=1:P}$, path sampled log-PF estimates $(\log Z_{G_p}(\beta_k))$ are computed on a discrete $\beta$-grid $(\beta_k = k\Delta \beta)$. 2) At each $\beta_k$-level, the linear regression coefficient $\vec{a}_k$ is estimated from all reference log-PF estimates as the minimizer of $\sum_{p=1}^P \| \log Z_{T_p}(\beta_k) - \vec{a}_k c_{G_p}\|^2$ where $c_{G_p}$ defines the number of cliques in $G_p$. 3) $\beta_k$-dependent linear extrapolation formula then applies to any Ising field defined on a test grid $T$: $\log Z_T(\beta) = \vec{a}_k c_T$. In [4], this method was shown to be efficient for accurately estimating the log-PFs of Ising fields defined on large regular grids (more than $15^3$ sites).

**Bilinear extrapolation.** Here, we also adopt a three-step procedure in which the second stage is improved since a bilinear regression that accounts for both the number of cliques $c_{G_p}$ and sites $s_{G_p}$ is introduced. The regression coefficients $(\vec{a}_k, \vec{b}_k$ and $\vec{c}_k)$ are obtained by minimizing the least square criterion $\sum_{p=1}^P \| \log Z_{T_p}(\beta_k) - a_k c_{G_p} - b_k s_{G_p} - c_k \|^2$. Then, $\beta_k$-dependent bilinear extrapolation formula applies to any 3D Ising field defined on a test grid $T$: $\log Z_T(\beta) = \vec{a}_k c_T + \vec{b}_k s_T + \vec{c}_k$. Log-PF estimates for $\beta$ values outside the $\beta$-grid are obtained using linear interpolation between its two closest values on the $\beta$-grid. This extension provides accurate extrapolation results for small and irregular graphs\(^1\) such as those appearing in our fMRI application. Intersect values ($\vec{c}_k$) are very close to 0 whenever the reference graphs $(G_p)_{p=1:P}$ are large, such as those in [4]. Their values are however not negligible for small reference graphs. Moreover, as illustrated in Fig. 2, at a fixed number of cliques, the higher the number of sites, the higher the PF whatever the regularity of the reference graphs. This dependence of $Z_{G_p}$ on $s_{G_p}$ at constant values of $c_{G_p}$ becomes much more important at small $\beta$ values. Hence, our bilinear extension of [4] significantly improves estimation performance for small and irregular grids.

![Fig. 2](image)

Examples of log-PF values for $\beta = 0.5$ as functions of the number of cliques $c$ and sites $s$. The projection onto a plane $c = constant$ is depicted to show the dependence on $s$.

---

\(^1\)Here, irregular refers to equally spaced Cartesian graphs of non-parallelepipedic shape.
3.3. Validation

For validation purpose, we compared log-PF estimates computed using the linear and bilinear extrapolation techniques with those obtained using path sampling, considered as the ground truth. *Reference* and *test* graphs are either regular or irregular. *Regular reference* graphs are of size $I \times J \times K$ with $I, J, K \in [1, \cdots, 7]$. Irregular graphs were extracted from regular bounding boxes in which Ising field configurations were drawn using the temperature dependent Swendsen-Wang algorithm [5]. In each bounding box, we considered the largest connected component of sites having the same label as an irregular graph. *Irregular reference* graphs were then computed using 170 bounding boxes of increasing size (from $10^3$ to $15^3$ sites) and regularization levels $\beta$ within the range $[0.2, 0.7]$. Regular *test* graphs form three subsets: 30 of them are small (less than $10^3$ sites), 30 are medium size (between $10^3$ and $15^3$ sites) and 30 are large (more than $15^3$ sites). Finally, irregular *test* graphs also form three subsets. Each contain 30 graphs obtained from bounding boxes of $10^3$ sites, for $\beta = 0.2, 0.4$ and 0.5, respectively.

Table 1 displays the mean maximal approximation error (in percents) of the log-PF estimates in comparison to the ground truth. The bilinear scheme is clearly more accurate than the linear one. However, the linear technique is slightly more efficient for large, weakly irregular fields (i.e., $\beta = 0.5$). It is also important to notice that the most suitable set of reference graphs depends on the shape of the test graphs. For instance, in the first row, the best results are obtained using a regular reference graph, which is not the case in other rows. Not surprisingly, further tests confirm that our algorithm achieves higher performance when test and reference graphs have similar topology and levels of regularity. Note also that each log-PF estimation required around 60 sec. using path sampling and less than 0.1 sec. using the linear and bilinear techniques, occuring a gain of two orders of magnitude.

### Table 1. Mean maximal approximation error over regular and irregular test graphs. Both linear and bilinear extrapolation techniques are tested. Errors are given in percents.

<table>
<thead>
<tr>
<th>Test grid</th>
<th>Scheme / Reference grid</th>
<th>B=bilinear, L=linear</th>
<th>R=regular, I=irregular</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B / R</td>
<td>B / I</td>
<td>L / R</td>
</tr>
<tr>
<td>Regular</td>
<td>0.747</td>
<td>3.84</td>
<td>5.55</td>
</tr>
<tr>
<td>medium</td>
<td>1.30</td>
<td>0.991</td>
<td>7.27</td>
</tr>
<tr>
<td>large</td>
<td>1.59</td>
<td>1.31</td>
<td>9.18</td>
</tr>
<tr>
<td>$\beta = 0.2$</td>
<td>6.85</td>
<td>1.29</td>
<td>23.6</td>
</tr>
<tr>
<td>$\beta = 0.4$</td>
<td>0.984</td>
<td>0.264</td>
<td>7.71</td>
</tr>
<tr>
<td>$\beta = 0.5$</td>
<td>1.73</td>
<td>1.27</td>
<td>1.64</td>
</tr>
</tbody>
</table>

4. APPLICATION TO fMRI TIME SERIES ANALYSIS

4.1. Problem statement

Our extrapolation algorithm was applied to the spatially adaptive regularization of the region-based Joint Detection - Estimation (JDE) of brain activity introduced in [1, 6]. The JDE approach relies on a prior parcellation of the brain into $\mathcal{P} = (P_i)_{i=1,...,I}$ functionally homogeneous and connected parcels [7]. Every parcel $P_i$ comprising voxels $(V_j)_{j=1,...,J}$ is characterized by a single haemodynamic filter $h_i$. Within a given $P_i$, voxel-dependent and stimulus-related fluctuations of the BOLD signal magnitude are encoded by the response levels $a = (a_{ij})_{j=1,...,J, m=1,...,M}$ (m stands for the stimulus type index). The fMRI time course measured in voxel $V_j$ then reads: $y_j = \sum_{m=1}^{M} a_{jm} x^{m} * h + b_j$, where $x^{m}$ stands for the binary stimuli vector, whose non-zero entries encode the arrival times of the stimulus of type $m$ and $b_j$ stands for the noise component [6]. Within the Bayesian framework, prior probability density functions (pdfs) are introduced on every sought object i.e., $(a, h)$. A Gaussian prior on $h$ with a smooth constraint is introduced as well as non-informative priors on every hyper-parameter [1]. Spatial Gaussian mixture models are expressed on $a$ through the introduction of hidden variables $q = (q_{jm}^{m})_{j=1,...,J, m=1,...,M}$ that encode whether voxel $V_j$ is activating in response to stimulus $m$ ($q_{jm}^{m} = 1$) or not ($q_{jm}^{m} = 0$). Hence, stimulus-dependent hidden Ising fields are introduced on these states such that the global prior pdf reads $p(a | \Theta_a) = \prod_m p(a^m | \theta_m)$ with:

$$p(a^m | \theta_m) = \sum_{q^m} \left( \prod_j f(a_{jm}^{m} | q_{jm}^{m}, \theta_m) \Pr(q^m | \beta_m) \right)$$

and $f(a_{jm}^{m} | q_{jm}^{m} = i) \sim N(\mu_{i,m}, v_{i,m})$. Parameters $\mu_{i,m}$ and $v_{i,m}$ define the prior mean and variance of class $i = 0, 1$, respectively for the stimulus type $m$. The set $\theta_m$ comprises four prior mixture parameters $\theta_m = (\mu_{1,m}, v_{1,m}, \mu_{0,m}, \beta_m)$ since non-activating voxels are modeled using a zero-mean Gaussian density ($\mu_{0,m} = 0$). Samples of the full posterior pdf $p(h, a, q, \Theta | \gamma)$ are simulated using a Gibbs sampler algorithm and inner Metropolis-Hastings (MH) steps in case of inability to draw samples from the full conditional posterior probability. Posterior mean estimates are then computed from these samples. Here, we focus on the sampling of parameter $\beta_m$, which is achieved using a symmetric random walk MH step: At iteration $k$, a candidate $\beta_m^{(k+1/2)} \sim N(\beta_m^{(k)}, \sigma^2)$ is generated. It is accepted (i.e., $\beta_m^{(k+1)} = \beta_m^{(k+1/2)}$) with probability:

$$\alpha(\beta_m^{(k)} \rightarrow \beta_m^{(k+1/2)}) = \min(1, A_{k,k+1/2})$$

where the acceptance ratio $A_{k,k+1/2}$ follows from Eq. (1):

$$A_{k,k+1/2} = \frac{Z(\beta_m^{(k+1)})}{Z(\beta_m^{(k)})} \exp \left( (\beta_m^{(k+1/2)} - \beta_m^{(k)})U(q_{jm}^{m}) \right)$$
using Bayes’ rule and considering a uniform prior for $\beta_m$. The JDE approach then requires to estimate ratios of $Z(.)$ or log-PF differences for all $\mathcal{P}_\gamma$ parcels prior to exploring the full posterior pdf.

4.2. Results on real fMRI data

We applied the JDE procedure to real fMRI data recorded during an event-related experiment designed to quickly map main sensory cortices (auditory, visual, motor) as well as higher cognitive functions (reading, computation). It consisted of a single session of $N = 125$ scans lasting $TR = 2.4$ s each, yielding a 3-D volume composed of $64 \times 64 \times 32$ voxels. The paradigm comprised a total of sixty stimuli, declined in 10 experimental conditions (auditive phrase, visual phrase, left auditory clic, left visual clic ...). We compare the three versions of the JDE procedure: Independent Mixture Models (IMM), Supervised SMM (SSSM, $\beta = 0.8$) and unsupervised SMM (USMM), in order to assess the impact of the adaptive spatial correlation model. Fig. 3 shows normalized contrast maps $\tilde{a}^{LAC} - \tilde{a}^{RAC}$ of auditory induced left clic versus auditory induced right clic, where the posterior mean estimates $\tilde{a}$ have been computed over 5000 realizations of the Gibbs sampler after a burn-in period of $10^3$ iterations. As expected, the activations lie in the contralateral right motor cortex. Here, only USMM is more sensitive illustrating therefore the advantage of an adaptive spatial correlation model. Indeed, estimated $\beta_{PM}$ with USMM for the left auditory clic was 0.56 so that the supervised setting of SSSM with $\beta = 0.8$ gives too much correlation and less sensitive results. On these real fMRI data, using the irregular reference grids presented in subsection 3.3, our bilinear extrapolation scheme provides reliable log-PF’s estimate for a brain parcellation $\mathcal{P} = (\mathcal{P}_\gamma)_{\gamma=1:1}$ containing $\Gamma = 500$ parcels. Note that the linear extrapolation scheme would have lead to non-reliable log-PF estimates here. In terms of computational complexity, these log-PF estimates were computed in about 10 seconds, which makes this approach very appealing. In comparison, path sampling requires about 1 hour to estimate the whole set of log-PF, for a negligible gain in accuracy (less than 3%). Finally, we did not observe any significant difference between the USSM activation map derived using path sampling (results not shown) and our extrapolation scheme (shown in Fig. 3(c)).

5. CONCLUSION

To achieve fast PF estimation of 3D Ising field, a bilinear extrapolation algorithm that exploits pre-computed accurate log-PF estimates on reference graphs has been proposed. Every log-PF estimate is computed as a function of both the number of cliques and sites in the corresponding graphs. This algorithm has been shown much more robust than a similar algorithm only exploiting the number of cliques, for small size and/or irregular graphs like those appearing in our fMRI application. This technique was integrated in a Bayesian joint detection-estimation algorithm of brain activity from fMRI data in order to make spatially adaptive regularization feasible. Results obtained on real fMRI data have shown a gain in statistical sensitivity for the unsupervised approach. Using our fast extrapolation technique, the computational burden remains acceptable for a whole brain analysis since the latter takes about one hour and the partition functions about 10 seconds. Future work will address the question of the robustness of the algorithm to the choice of the reference graphs.

6. REFERENCES


