ECAI 2010
H. Coelho et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-606-5-687

Decision-theoretic Optimal Sampling in Hidden Markov Random Fields

N. Peyrard[1], **R. Sabbadin**[1] and **U. Farrokh Niaz**[2]

Abstract. Computation of the *Most Probable Explanation* (MPE) when probabilistic knowledge is expressed as a factored distribution is a classical AI reasoning problem: complete *evidence* is available about the values of some of the variables which are observed, and the problem consists in finding the most probable assignment of the remaining variables given the evidence. However, optimising the choice of the variables to observe (the sample) in order to maximise the MPE probability is a less classical and more difficult problem. In this article we tackle this question of optimal sampling in structured problems under limited budget, within the framework of Hidden Markov Random Fields (HMRF). The *value* of a sample (which we seek to optimise) is the expectation, over all possible sample outputs (observations), of the MPE probability. The contributions of this article are: i) an original probabilistic model for optimal sampling in HMRF ii) computational complexity results about this problem, leading in particular to approximability/inapproximability results and iii) an exact solution algorithm and two approximate solution algorithms of decreasing time complexity, which we empirically evaluate on a problem of spatial sampling for occurrence map restoration.

1 INTRODUCTION

The computation of the *most probable assignment* of a subset of variables given complete evidence on the complement of that set when probabilistic knowledge is expressed as a factored distribution is a classical (but still difficult) generic AI reasoning problem [17]. In the *graphical models* community it corresponds to the *Most Probable Explanation* (MPE) problem: the values of a subset of the variables are observed and the most probable value of the remaining ones is computed. It is also known as the computation of the *Maximum a Posteriori* (MAP) problem in image analysis, when variables are organised as pixels and one seeks for image restoration given noisy observations. In many real-life situations, the set of variables on which information is acquired is not fixed, even though its size is limited by a budget constraint. The question is then to select this set in order to achieve a high confidence in the values of the remaining variables, i.e. high MPE/MAP value. The variable assignment with highest probability is a restoration of the set of unobserved variables. The optimisation of the sample set is a less classical and more difficult problem than MPE/MAP, especially when observations are noisy and costly. A generic formulation as well as a study of its computational complexity and development of solution algorithms are still to be done. In this article we tackle this question of optimal sampling under limited budget for MPE/MAP maximisation, within the framework of Hidden Markov Random Fields (HMRF) [4, 9, 14]. A HMRF is a non-oriented graphical model, classically used in image restoration, where variables are of two types: hidden variables (to reconstruct) and observed variables (which are noisy observations of the hidden variables). The sampling set must be chosen among the set of observed variables. The *value* of a sample (which we seek to optimise) is defined as the expectation, over all possible observations of the sample set, of the conditional probability of the most probable hidden variables assignment. The advantage is that selection of the sample set and restoration of the unobserved variables are achieved using the same criterion. We describe an original framework for optimal sampling in HMRF aiming at MPE/MAP restoration (Section 2). Then we establish computational complexity results about this problem, leading in particular to approximability/inapproximability results (Section 3). An exact solution algorithm and two approximate solution algorithms of decreasing complexity are proposed (Section 4). We empirically evaluate their performance on the problem of spatial sampling for occurrence map restoration: the graph of the HMRF is a regular grid and all variables are binary (presence/absence and detection of the phenomenon). Since false negative observations are possible, observations are noisy (Section 5).

[1] INRA-BIA, Toulouse, France, email: {peyrard,sabbadin}@toulouse.inra.fr
[2] ISIR-UPMC, Paris, France . usman.niaz@isir.upmc.fr. U. Niaz contributed to this work during his Master's thesis at INRA.

2 OPTIMAL SAMPLING FOR HMRF RESTORATION

2.1 Hidden Markov Random Fields

Markov Random Fields (MRF) are a family of undirected *probabilistic graphical models* [12] allowing to concisely model probability distributions over factored state spaces where dependencies between variables are *local*. The dependence structure is represented by a graph, $G = (V, E)$, with n vertices representing variables and edges help showing the conditional dependence between the variables: any variable in V is conditionally independent of other variables given those sharing an edge with it.

If a positive factored probability distribution exhibits such independence property, it can be expressed as a product of positive *clique functions*, i.e. functions of subsets of variables which indices form cliques of the graph G (Hammersley-Clifford theorem) [4]:

Definition 1 (Markov Random Fields) *Let $X = \{X_1, \ldots, X_n\}$ be random variables taking values in $\{0, ..., k-1\}$ each and let P be a probability distribution over $\mathcal{X} = \{0, ..., k-1\}^n$. Let also $G = (V, E)$ with $|V| = n$ and $\mathcal{C}$ be the set of* cliques *of graph G.*

P is a Markov Random Field *defined by $(G, \mathcal{X}, \Psi)$ if Ψ is a set of positive functions $\Psi = \{\psi_c\}_{c\in\mathcal{C}}, \psi_c(x_c) > 0, \forall x_c$, such that:*

$$P(X = x) = P_\Psi(x) = \frac{1}{Z}\prod_{c\in\mathcal{C}} \psi_c(x_c), \forall x \in \mathcal{X}, \qquad (1)$$

where Z is a normalising constant ensuring that P sums to 1.

(Here and in the following, $\forall A \subset V$, notation x_A represents $\{x_i, i \in A\}$)

Now, let us consider the case where the variables X are not observed directly, but random variables $Y = \{Y_1, \ldots, Y_n\}$ taking values in $\mathcal{Y} = \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_n$ are observed instead, and each Y_i is conditionally independent of all other variables given X_i. The conditional parameterised probability is known and takes the following form (with $\Theta = \{\theta_1, \ldots, \theta_n\}$ a set of parameters):

$$P(Y = y|X = x) = P_\Theta(y|x) = \prod_{i \in V} P_{\theta_i}(y_i|x_i), \forall x, \forall y. \quad (2)$$

A *Hidden Markov Random Field (HMRF)* is defined by the tuple $(G, \mathcal{X}, \mathcal{Y}, \Psi, \Theta)$. A typical problem in the HMRF framework is to "restore" the hidden variables values $\{x_i\}$ given the observed variables values $\{y_i\}$ [9, 14].

However, observations may be incomplete, for example if observing requires to consume limited resources (time, people, etc.). In the following, we will be interested in the problem of choosing a *sample*, i.e. a subset $\sigma \subseteq V$ of sites which will be actually observed in order to restore the hidden field x. The result of a sample σ is a a *sample output* y_σ. From y_σ, we can define a *posterior probability conditioned on a sample output*:

Definition 2 (Posterior probability) *Let $P_\Psi(x)$ and $P_\Theta(y|x)$ be defined from a HMRF $(G = (V, E), \mathcal{X}, \mathcal{Y}, \Psi, \Theta)$. Let $\sigma \subseteq V$ be a sample and y_σ be a sample outpout. Then, the posterior probability conditionned on sample output y_σ is defined as:*

$$P_{\Psi,\Theta}(x|y_\sigma) \propto P_\Theta(y_\sigma|x)P_\Psi(x), \forall x \in \mathcal{X}. \quad (3)$$

2.2 Decision-theoretic HMRF sample optimisation

So, the problem we face is that of, given a HMRF and limited sampling resource, choosing a good sample $\sigma \subseteq V$. This choice should be based on the posterior probability distribution over hidden variables, $P_{\Psi,\Theta}(x|y_\sigma)$, defined in equation (3). However, since the sample output y_σ resulting from a sample σ is unknown when the choice of σ is made, the "quality" of the sample, denoted $U_{\Psi,\Theta}(\sigma)$, has to be computed as the expectation of the information values resulting from all potential sample outputs. In fact, we will consider a slightly generalised measure, in order to account for possibly different costs associated to the different sample locations. So, the decision criterion will include an additive sample cost function, $\gamma : 2^V \to \mathbb{R}^+$: $\gamma(\sigma) = \sum_{i \in \sigma} \gamma_i$, where the γ_i are positive.

Definition 3 (Decision-theoretic sample quality measure)

$$U_{\Psi,\Theta}(\sigma) = -\gamma(\sigma) + \sum_{y_\sigma} P_{\Psi,\Theta}(y_\sigma) V_{\Psi,\Theta}(\sigma, y_\sigma), \quad (4)$$

where $V_{\Psi,\Theta}$ measures how informative $P_{\Psi,\Theta}(x|y_\sigma)$ is.

$\gamma(\sigma)$ measures the "cost" of σ, and its unit is homogeneous to $V_{\Psi,\Theta}(\sigma, y_\sigma)$. The sample quality measure (4) corresponds to a *decision-theoretic information value* criterion [11], often used when the sampling problem precedes an actual decision choice, which utility depends on the hidden state of the process. This kind of criterion is thus used in problems of fault diagnosis and repair, detection and eradication of invasive species, etc...

In this article, our goal is to find the sample set which maximises (in expectation) the probability of the *most probable assignment* of X according to $P_{\Psi,\Theta}(x|y_\sigma)$. This maximisation problem is known as the *Most Probable Explanation* problem (MPE). In image analysis, it is also known as the *Maximum a Posteriori (MAP)*[3] criterion [14]. It has the advantage to be associated with a restoration procedure (the assignment maximising the conditional distribution is returned).

Definition 4 (Most Probable Explanation (MPE) criterion)

$$V^{MPE}_{\Psi,\Theta}(\sigma, y_\sigma) = \max_{x \in \mathcal{X}} P_{\Psi,\Theta}(x|y_\sigma), \forall y_\sigma. \quad (5)$$

The MPE criterion involves a maximisation over x and in this way provides a way to compute a restoration, x^{MPE}, of the hidden process:

$$x^{MPE}(y_\sigma) = \arg\max_{x \in \mathcal{X}} P_{\Psi,\Theta}(x|y_\sigma), \quad (6)$$

$V^{MPE}_{\Psi,\Theta}(\sigma, y_\sigma)$ is thus the probability of the most likely restoration, $x^{MPE}(y_\sigma)$, of the hidden process given the observations.

Several other choices can be classically made for $V_{\Psi,\Theta}$: the *entropy criterion* [6] or the *Maximum Posterior Marginals (MPM)* criterion [4]. Note that the solution algorithms which we will propose in the following can be easily adapted to the MPM criterion, eventhough the theoretic complexity class of the problem would be different. Regarding entropy, this criterion is not associated with a restoration procedure.

Now, we are able to describe the HMRF sample optimisation problem. It consists in finding a sample $\sigma^* \subseteq W \subseteq V$, of size bounded by a constant K, which maximises $U^{MPE}_{\Psi,\Theta}$, associated to $V^{MPE}_{\Psi,\Theta}$. The set $W \subseteq V$ is the subset of sites which we are allowed to sample (sites in $V \setminus W$ are not accessible to sampling).

Definition 5 (Decision-theoretic sample optimisation problem) *Let HMRF $(G = (V, E), \mathcal{X}, \mathcal{Y}, \Psi, \Theta)$ be given, together with an integer $K \geq 0$, a set $W \subseteq V$ of sites available for sample and costs $\gamma_i \geq 0, \forall i \in W$. Then, the HMRF sample optimisation problem $\mathcal{O}(U^{MPE}_{\Psi,\Theta})$ is defined as:*

$$\textit{Find } \sigma^* = \arg \max_{\sigma \subseteq W, |\sigma| \leq K} U^{MPE}_{\Psi,\Theta}(\sigma). \quad (7)$$

In the following section, we will study the computational complexity of this problem, as well as its approximability properties in the "small sample" case. For this, it is useful to note that $U^{MPE}_{\Psi,\Theta}(\sigma)$ takes the following form, which can be easily shown, using equations (5) and (4) and applying Bayes' theorem:

Proposition 1 (MPE sample value criterion)

$$U^{MPE}_{\Psi,\Theta}(\sigma) = -\gamma(\sigma) + \sum_{y_\sigma} \max_x P_\Theta(y_\sigma|x)P_\Psi(x). \quad (8)$$

In the following, we note $Q^{MPE}_{\Psi,\Theta}(\sigma, y_\sigma) = \max_x P_\Theta(y_\sigma|x)P_\Psi(x)$.

3 COMPUTATIONAL COMPLEXITY CONSIDERATIONS

3.1 Tree representation of the MPE sample optimisation problem

Solving the sample optimisation problem (7) intuitively amounts to explore the three-level tree shown in Figure 1. The top level is a $|W|$-depth binary tree, where vertices correspond to the vertices in W

[3] In the *graphical models* community, "MAP" stands for a more general problem, where one is looking for a maximum probability assignment of a subset of variables, given a *partial assignment* of the remaining variables [17].

and outgoing edges determine whether the vertex is selected into the sample or not. A branch in the top tree thus represents a sample σ. Then, from each valid branch (i.e. containing at most K selected vertices), a second l-ary tree (l is the size of the domains of the Y_i) of depth $|\sigma| \leq K$ is issued, where a branch is an assignment y_σ. Finally, a k-ary tree (k is the size of the domains of the X_i) of depth $|V|$ is issued from any branch corresponding to a valid assignment (σ, y_σ). A branch of this third level tree corresponds to an assignment x.

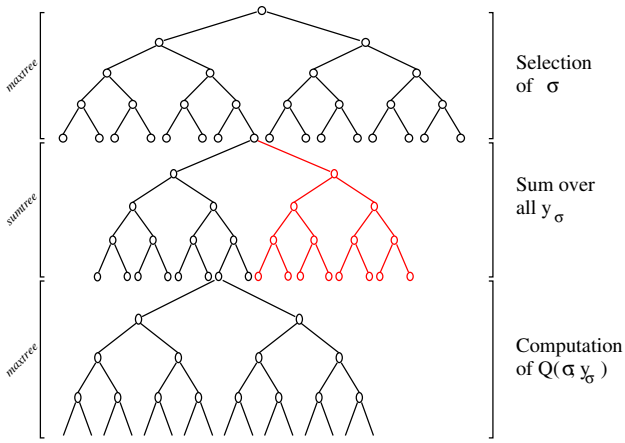

Figure 1. Tree representation of the MPE sample optimisation problem, with $W = V$, $|V| = 4$, $K = 2$ and $l = k = 2$.

It is easy to see that for any valid assignment (σ, y_σ), $x^{MPE}(y_\sigma)$ and $Q^{MPE}_{\Psi,\Theta}(\sigma, y_\sigma)$ can be computed through exploration of the bottom max-tree. Then, once $Q^{MPE}_{\Psi,\Theta}(\sigma, y_\sigma)$ is known for all valid assignments, $U^{MPE}_{\Psi,\Theta}(\sigma)$ can be computed for any valid sample, by exploring the second level sum-tree. Finally, σ^* can be obtained through exploration of the upper-level max-tree.

Using this tree description of the problem, we can show that the sample optimisation problem (7) belongs[4] to $\mathsf{NP}^{\mathsf{PP}^{\mathsf{NP}}}$. However, the computational complexity of the sampling problem depends on the size of the intermediate "sum tree", which depends on K. We are now going to show that, when K is "small", i.e. independent of the problem instance size, the decision problem associated to the sample optimisation problem is NP-complete.

3.2 Small sample optimisation problem

The computation of Z in $P_\Psi(x)$ is a difficult problem (#P-hard), necessary to compute the value of $U^{MPE}_{\Psi,\Theta}(\sigma)$. However, this computation becomes unnecessary when cost function γ is null and we are only interested in finding the best sample, not its exact value. Furthermore Z only has to be computed once for a given Ψ and does not have to be recomputed when σ or y_σ changes. From now on we will assume that the partition function Z is known.

Definition 6 (Decision problem associated to $U^{MPE}_{\Psi,\Theta}$) *Let us define $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$, the decision problem associated to $U^{MPE}_{\Psi,\Theta}$:*
Input*:*

- *HMRF $(G = (V, E), \mathcal{X}, \mathcal{Y}, \Psi, \Theta)$, where Z* **is known***,*
- *$\emptyset \subseteq W \subseteq V$ is the subset of sites available for sample, $K \leq |W|$ is a fixed upper bound on the sample size, γ is an additive cost function, $\mu \geq 0$.*

Question*: Does there exist $\sigma \subseteq W, |\sigma| \leq K$ s. t. $U^{MPE}_{\Psi,\Theta}(\sigma) \geq \mu$?*

Proposition 2 (NP-hardness of $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$)
The decision problem $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$ is NP-hard.

Proof: We show that the decision problem for MAX 2-SAT [8] can be polynomially reduced to $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$. Let $\phi = \{Cl_k\}$ be a set of 2-clauses over boolean variables $\{x_1, \ldots, x_n\}$ and ν be a positive integer. We define $\#SAT_\phi(x) = |\{Cl_k \in \phi, x_{Scope(Cl_k)} \models Cl_k\}|$. The decision problem for MAX 2-SAT is the following:

$$\text{"Given } \nu > 0, \text{ is } \max_x \#SAT_\phi(x) \geq \nu\text{?"}$$

Let us define the corresponding problem $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$ as:

- $V = \{1, \ldots, n\}, \mathcal{C} = \{(i,j) \in V^2, \exists Cl_k \in \phi, Scope(Cl_k) = (i,j)\}$.
- $\psi_{ij}(x_i, x_j) = exp(m_{ij}), \forall (i,j) \in \mathcal{C}$, where $m_{ij} = |\{Cl_k \in \phi, Scope(Cl_k) = (i,j) \text{ and } (x_i, x_j) \models Cl_k\}|$.
- $K = 0$, $W = V$, $\gamma = 0$, Θ arbitrary and $\mu = e^\nu / Z$.

Then, we can easily check that:

$$U^{MPE}_{\Psi,\Theta}(\emptyset) \geq \mu \Leftrightarrow \max_x \#SAT_\phi(x) \geq \nu, \quad (9)$$

$$\text{since } U^{MPE}_{\Psi,\Theta}(\emptyset) = \frac{1}{Z} \exp\left(\max_x \#SAT_\phi(x)\right).$$

From this, it results that MAX 2-SAT can be polynomially reduced to $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$, which is thus NP-hard. □

Now, we are going to show that the decision problem for $U^{MPE}_{\Psi,\Theta}$ belongs to NP when the sample size bound is "small", where "small" means "independent of the problem instance size".

Proposition 3 ($\mathcal{D}(U^{MPE}_{\Psi,\Theta})$ is in NP)
When K is independent of the problem instance size, the decision problem associated to the sample optimisation problem $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$ belongs to NP .

Proof: To show the result, it is enough to exhibit a certificate for $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$ of polynomial size and which can be checked in polynomial time. Note that to compute $U^{MPE}_{\Psi,\Theta}(\sigma)$, we have to sum over $l^{|\sigma|}$ terms, one for each y_σ. Thus, a certificate for $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$ will be a tuple $(\sigma, x^1, \ldots, x^{l^{|\sigma|}})$, where $x^j = \arg\max_x P_\Theta(y^j|x) P_\Psi(x) = x^{MPE}(y^j)$, if $y^j \in \{0, \ldots, l-1\}^{|\sigma|}$ is the j^{th} possible value of y_σ.

Since x^i is a length-$|V|$ vector of integers between 0 and $k-1$ and σ is a binary vector of $|W|$ bits, the certificate size is bounded by $|W| + l^K |V| \log_2 k$, polynomial in the size of the problem (if K is not allowed to vary with the problem instance size).

Now, computing $U^{MPE}_{\Psi,\Theta}(\sigma)$ for a given certificate requires to sum $l^{|\sigma|}$ terms which are products of $(|W| + |\mathcal{C}|)$ terms, then adding $|W|$ terms to the result. So, $O(l^K(|V| + |\mathcal{C}|))$ elementary operations (additions, multiplications) have to be performed. □

From propositions 2 and 3, we get:

Proposition 4 ($\mathcal{D}(U^{MPE}_{\Psi,\Theta})$ is NP-complete)
When K is fixed, the decision problem associated to the sample optimisation problem $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$ is NP-complete.

[4] See below for a definition of the complexity class PP and a proof of this statement

3.3 Approximability of the small sample optimisation problem

The NP-completeness of the decision problem being established, the next question that can be raised concerns the approximability of the optimisation problem. Once again, building on approximability/inapproximability results concerning MAX 2-SAT, we obtain corresponding results about the HMRF sample optimisation problem $\mathcal{O}(U^{MPE}_{\Psi,\Theta})$ (Definition 5).

First, recall that NPO is the class of optimisation problems analogous to NP [2]. The results of the preceding section imply that:

Proposition 5 (NPO-completeness)
The optimisation problem $\mathcal{O}(U^{MPE}_{\Psi,\Theta})$ is NPO-complete when Z is known, $\gamma = 0$ and K is not allowed to vary with the problem instance size.

Proof: $\mathcal{O}(U^{MPE}_{\Psi,\Theta})$ belongs to NPO since i) an instance is recognisable in polynomial time, ii) valid solutions σ are polynomially recognisable and iii) $U^{MPE}_{\Psi,\Theta}(\sigma)$ can be computed in polynomial time (if Z is known). Now, hardness is a direct consequence of the fact that $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$ is NP-hard. □

Recall that APX is the set of optimisation problems belonging to NPO which admit a polynomial-time approximation algorithm with approximation ratio bounded by a constant $r \geq 1$ (i.e. $\frac{m^*}{m(x)} \leq r$) [2]. A natural question is to ask whether the optimisation problem $\mathcal{O}(U^{MPE}_{\Psi,\Theta})$ belongs to APX. Unfortunately, we will show now that it is not the case. Thus, it does not admit a polynomial-time approximation algorithm with constant approximation ratio, unless P = NP.

Proposition 6 (Non approximability)
$\mathcal{O}(U^{MPE}_{\Psi,\Theta})$ *does not belong to* APX.

Proof: Assume that $\mathcal{O}(U^{MPE}_{\Psi,\Theta})$ is polynomially approximable with constant ratio ρ. Then, in particular, $U^{MPE}_{\Psi,\Theta}(\emptyset)$ is approximable with ratio ρ. But then, using the same kind of transformation from $\#SAT$ to $U^{MPE}_{\Psi,\Theta}(\emptyset)$ as in Proposition 2, we can associate to any 2-SAT formula ϕ a pairwise HMRF $(G, \mathcal{X}, \mathcal{Y}, \Psi, \Theta)$, such that

$$U^{MPE}_{\Psi,\Theta}(\emptyset) = \frac{1}{Z}\exp(\max_x \#SAT_\phi(x)).$$

Now, assume that $\tilde{U} = \frac{1}{Z}\exp(\#SAT_\phi(\tilde{x}))$ is a ρ-approximation of $U^{MPE}_{\Psi,\Theta}(\emptyset)$. Then, $\tilde{U} \geq \frac{1}{\rho}U^{MPE}_{\Psi,\Theta}(\emptyset) \Leftrightarrow \exp(\#SAT_\phi(\tilde{x})) \geq \frac{1}{\rho}\exp(\max_x \#SAT_\phi(x)) \Leftrightarrow \#SAT_\phi(\tilde{x}) \geq \max_x \#SAT_\phi(x) - \log\rho$.

So, any polynomial-time approximation algorithm for $U^{MPE}_{\Psi,\Theta}(\emptyset)$ with constant ratio ρ provides an approximation for MAX 2-SAT with *absolute precision* $\log\rho$. But, [10] has shown that MAX 2-SAT is not absolutely approximable. So, $U^{MPE}_{\Psi,\Theta}(\emptyset)$ does not admit a polynomial-time approximation algorithm with constant ratio and, a fortiori, $\mathcal{O}(U^{MPE}_{\Psi,\Theta})$ does not belong to APX. □

Eventhough $\mathcal{O}(U^{MPE}_{\Psi,\Theta})$ does not belong to APX, we still can show the following positive approximability result:

Proposition 7 ($U^{MPE}_{\Psi,\Theta}(\sigma)$ approximation)
Let us consider a HMRF $(G = (V, E), \mathcal{X}, \mathcal{Y}, \Psi, \Theta)$, with **binary variables** *and* **null cost function** *(Z known) and $\sigma \subseteq V$. Furthermore, assume $P_{\theta_i}(y_i|x_i) > 0, \forall i, x_i, y_i$. Then, $U^{MPE}_{\Psi,\Theta}(\sigma)$ can be approximated in polynomial time with an* **instance-dependent** *approximation ratio $\rho_{\Psi,\Theta,\sigma} = \kappa^{0.23}_{\Psi,\Theta,\sigma}$, where*

$$\kappa_{\Psi,\Theta,\sigma} = \frac{\prod_{c\in\mathcal{C}}\max_{x_c}\psi_c(x_c)\prod_{i\in\sigma}\max_{x_i,y_i}P_{\theta_i}(y_i|x_i)}{\prod_{c\in\mathcal{C}}\min_{x_c}\psi_c(x_c)\prod_{i\in\sigma}\min_{x_i,y_i}P_{\theta_i}(y_i|x_i)}.$$

If furthermore the HMRF is pairwise, $\rho_{\Psi,\Theta,\sigma} = \kappa^{0.069}_{\Psi,\Theta,\sigma}$.

Proof: Let us define

$$\begin{aligned} f_{\Psi,\Theta,\sigma,y_\sigma}(x) &= \sum_{i\in\sigma}\left(\log P_\Theta(y_i|x_i) - \min_{x'_i,y_i}\log P_\Theta(y_i|x'_i)\right) \\ &+ \sum_{c\in\mathcal{C}}\left(\log\psi_c(x_c) - \min_{x'_c}\log\psi_c(x'_c)\right). \end{aligned}$$

The subscripts of f will be omitted for sake of simplicity. Then, we use approximability results for WEIGHTED MAX SAT [1, 7], which provide approximation ratios applicable to f. Note $q(x) = \frac{1}{Z}\exp(f(x))$, then $Q^{MPE}_{\Psi,\Theta}(\sigma, y_\sigma) \propto \max_x q(x)$.

If f is ρ-approximable, there is a polynomial time algorithm returning for any instance an approximate solution $\tilde{x}$ such that $f(x^*) \leq \rho f(\tilde{x})$. Now,

$$\frac{q(x^*)}{q(\tilde{x})} = \frac{\exp(f(x^*))}{\exp(f(\tilde{x}))} \leq \frac{\exp(f(x^*))}{\exp(f(x^*))^{\frac{1}{\rho}}} = \exp(f(x^*))^{1-\frac{1}{\rho}}, \text{ but}$$

$$\begin{aligned} f(x^*) &\leq \sum_{i\in\sigma}\left(\max_{x_i}\log P_\Theta(y_i|x_i) - \min_{x'_i,y_i}\log P_\Theta(y_i|x'_i)\right) \\ &+ \sum_{c\in\mathcal{C}}\left(\max_{x_c}\log\psi_c(x_c) - \min_{x'_c}\log\psi_c(x'_c)\right) \\ &\leq \log(\kappa_{\Psi,\Theta,\sigma}). \end{aligned}$$

So, $q(x^*)$ is approximable with $\kappa^{1-\frac{1}{\rho}}_{\Psi,\Theta,\sigma}$. And since this ratio is independent of y_σ, it also holds for $Q^{MPE}_{\Psi,\Theta}(\sigma, y_\sigma)$ and $U^{MPE}_{\Psi,\Theta}(\sigma)$.

In the case of binary variables (and arbitrary constraints), [1] provided a constant approximation ratio $\rho = 1.2987$ for WEIGHTED MAX SAT. When furthermore constraints are pairwise, ρ is decreased to $\rho = 1.0741$ [7]. These two ratios translate to the instance-dependent ones mentioned in the proposition. □

3.4 Large sample optimisation problem

We have shown that when the sample size bound is independent of the problem size, $\mathcal{O}(U^{MPE}_{\Psi,\Theta})$ belongs to NPO. In this section we are going to show that, in the general case, the decision problem $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$ is much harder: it is PP-hard and this complexity bound is likely to be a strict lower bound.

Recall first that PP is the probabilistic complexity class which emblematic complete problem is MAJSAT:

Let ϕ be a 3-CNF over $\{x_1, \ldots, x_n\}$. Is $|\{x, x \models \phi\}| > 2^{n-1}$?

In other terms, one asks whether more than half of the assignments satisfy ϕ. Then, using reduction from MAJSAT to $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$, we show the following proposition:

Proposition 8 (PP-hardness of $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$)
$\mathcal{D}(U^{MPE}_{\Psi,\Theta})$, where Z is assumed to be known is PP-hard.

Sketch of proof: Assuming that $\phi = Cl_1 \wedge \ldots \wedge Cl_m$ is a 3-CNF over $\{X_1, \ldots, X_n\}$, we consider the following problem $\mathcal{D}(U^{MPE}_{\Psi,\Theta})$:

- $V = \{1, \ldots, n\}$, $W = V$, $\mathcal{X} = \mathcal{Y} = \{0, 1\}^n$, $\mathcal{C} = \{Scope(Cl_k)\}_{k=1\ldots,m}$ and E is defined accordingly,
- $\psi_{Scope(Cl_k)}(x_{Scope(Cl_k)}) = 1$ if $x_{Scope(Cl_k)} \models Cl_k$ and 2^{-n-1} else, $\forall k = 1, \ldots, m$,

- $K = n$, $P_{\theta_i}(y_i|x_i) = 1$ if $x_i = y_i$ and 0 else, $\mu = \frac{2^{n-1}}{Z}$, $\gamma(\sigma) = 0, \forall \sigma \subseteq V$.

  Note that $U^{MPE}_{\Psi,\Theta}(V) = \frac{1}{Z}\sum_x \prod_{Cl_k} \psi_{Scope(Cl_k)}(x_{Scope(Cl_k)})$.

But, $\prod_{Cl_k} \psi_{Scope(Cl_k)}(x_{Scope(Cl_k)}) = 1$ if $x \models \phi$ and is less than 2^{-n-1} else. So,

$$U^{MPE}_{\Psi,\Theta}(V) = \frac{1}{Z}\left(|\{x, x \models \phi\}| + \varepsilon\right) \text{ with } 0 \leq \varepsilon \leq \frac{1}{2}.$$

$$\text{So, } \frac{1}{Z}|\{y, y \models \phi\}| \leq U^{MPE}_{\Psi,\Theta}(V) \leq \frac{1}{Z}(|\{y, y \models \phi\}| + \frac{1}{2})$$

and $|\{y, y \models \phi\}| > 2^{n-1} \Leftrightarrow U^{MPE}_{\Psi,\Theta}(V) > \frac{2^{n-1}}{Z}$.

Since $U^{MPE}_{\Psi,\Theta}$ is increasing with inclusion, we get the result. □

4 SAMPLING OPTIMISATION ALGORITHMS

From now on, we limit ourselves to the case where the cost function γ is null. This allows to avoid to compute Z (sample values are thus determined up to a constant coefficient). However, the family of *variable elimination* subroutines which are used in the "exact" and "greedy" algorithm also apply to the computation of Z and the "belief propagation" algorithm does not require to know Z.

4.1 "Exact" variable elimination

In the exact approach, we use a classical *variable elimination algorithm* (e.g. [12]) to compute the MPE solution $x^{MPE}(y_\sigma) = \arg\max_x P_\Theta(y_\sigma|x)\prod_{c\in\mathcal{C}}\psi_c(x_c)$. Then, an optimal sample $\sigma^*(K) = \arg\max_{\sigma\subseteq W, |\sigma|\leq K} U^{MPE}_{\Psi,\Theta}(\sigma)$ is computed recursively, requiring to compute $x^{MPE}(y_\sigma)$ for all valid pairs (σ, y_σ).

Thus, it is easy to see that the time complexity of the algorithm grows exponentially with the sample size bound K (note that since $U^{MPE}_{\Psi,\Theta}$ is increasing with inclusion, $|\sigma^*(K)| = K$).

4.2 "Greedy" algorithm

The greedy algorithm is a variation of the exact algorithm in which a suboptimal sample is built incrementally:

- First, an optimal sample of size 1 is computed by the exact variable elimination method.
- Then, given a current sample σ of size $|\sigma| < K$, all samples $\sigma_i = \sigma \cup \{i\}, \in W \setminus \sigma$ are evaluated exactly and σ is replaced with the best augmented sample.
- When the sample size has reached K, the corresponding "greedy" suboptimal sample σ_G is returned.

Applying the greedy algorithm allows to replace a $C^K_{|W|}$ factor (linked to the exhaustive exploration of $\{\sigma, |\sigma| = K\}$) with a polynomial one (the number of samples evaluated by the greedy algorithm is less than $|W|K$). Of course, the greedy algorithm still takes exponential time to run: the variable elimination algorithm takes exponential time to run and, for a given σ, the set $\{y_\sigma\}$ also has exponential size.

Now, one could ask whether the greedy algorithm provides bounds on the loss of quality of the solution returned. Indeed, as far as *submodular set functions* [16] are concerned, the existence of such a bound on the quality of greedy solutions is known. However, eventhough $U^{MPE}_{\Psi,\Theta}$ is an increasing set function, *it is not submodular in the general case*. We will not develop here on this negative result.

4.3 "Belief propagation" algorithm

Belief propagation (BP) algorithms [20] use message passing updates to compute approximate marginal probabilities $b_i(x_i)$ for all vertices of a Markov Random Field $(G, \mathcal{X}, \Psi)$. The belief propagation approximation for the sample optimisation problem consists in:

- Computing BP approximate marginals $b_i(x_i)$ for the MRF $(G, \mathcal{X}, \Psi)$ and
- choosing the K most "uncertain" sites belonging to W to sample, i.e. the sites with the smaller values of $\max_{x_i} b_i(x_i)$.

Note that this approximation algorithm does not take into account sample outputs for choosing the sample. Thus, its time complexity is roughly independent of K.

5 EXPERIMENTS

We have tested the performance of the exact variable elimination algorithm as well as the greedy and belief propagation algorithms on a "map reconstruction problem" over a regular grid. Namely, we considered a pairwise HMRF as represented in Figure 2, with $\mathcal{X}_i = \mathcal{Y}_i = \{0, 1\}$. The singleton potential values are specified on the figure. The pair potential function are independent of i and have value $\exp(1) = e$ for assignments $(1, 1)$ and $(0, 0)$ and 1 otherwise. The sizes of the grid problems increase from 3×2 to 6×6.

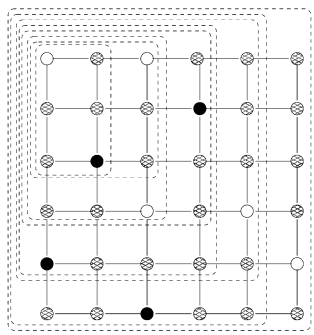

Figure 2. HMRF examples (from 3×2 to 6×6 grids). White dots: $\psi_i(1) = 0.3$. Grey dots: $\psi_i(1) = 1$. Black dots: $\psi_i(1) = 3$. $\psi_i(0) = 1$, $\forall i$.

We have computed the exact values of (i) optimal sample, V^* (ii) "greedy" sample, V^G and (iii) "belief propagation" sample, V^{BP}, for sample size bounds 1 to 4, for the eight subproblems of size 3×2 to 6×6. The value ratios $\frac{V^G}{V^*}$ and $\frac{V^{BP}}{V^*}$, as well as the logarithms of the execution times of all three methods are shown in Figure 3. Note that the exact method did not solve the 6×6 problem after 9 days ($\exp(13)$ seconds).

Eventhough the greedy algorithm is far faster than the exact one (up to 10 times faster for $K = 2$, 100 times faster for $K = 3$ and 500 times faster for $K = 4$), we observed that the computed solutions are quite close to optimality for all configurations (at least 90% of the optimal). The computational gain when running the belief propagation algorithm is from 10 to 50, compared to the greedy algorithm. However, the ratio to optimality decreases below 80%.

Now, for larger problems, the exact optimal solution is out of reach, and the greedy algorithm itself is limited in size, due to its exponential time complexity. The belief propagation algorithm complexity is only limited by the number of updates allocated and its

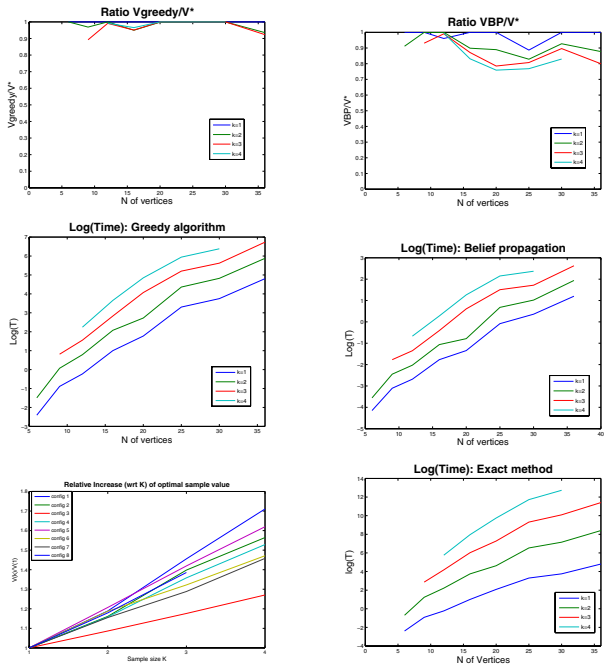

Figure 3. Top: Ratio of values, greedy (left), belief propagation (right). Middle and bottom right: Log-times of greedy, belief propagation and exact algorithms. Bottom left: relative increase of optimal value wrt sample size.

time complexity is independent on K. The largest part of the time needed to execute BP is devoted to the final exact computation of the returned solution value.

6 CONCLUDING REMARKS, RELATED AND FUTURE WORK

In this article, we propose an approach for optimal sampling in structured problems under limited budget, within the framework of Hidden Markov Random Fields (HMRF). The *value* of a sample is the expectation, over all possible observation outputs, of the MPE probability. We have obtained computational complexity results about this problem, leading in particular to approximability/inapproximability results. We have designed exact and approximate solution algorithms for this problem, which we have empirically evaluated on a problem of spatial sampling for occurrence map restoration.

Several recent works have addressed the question of decision-theoretic observation selection in graphical models. [13] have considered reliable observations and simple problem structures (chain model, naive Bayes model or polytree). They have provided complexity results and exact and approximate solution algorithms. In this work, rewards are *local*. [18, 19] have considered the noisy observations case, with simple problem structures (hidden Markov chains, tree-shaped Bayesian networks) and easily computable local rewards. On their side, [15] have considered general Bayesian networks and noisy observations, but with a very specific and easy to compute non-local reward function.

The originality of our work is to consider a general HMRF model structure, noisy observations and a non-trivially (NP-hard) computable global reward function (based on MPE). Furthermore, we focus on sample size, rather than on problem structure, in order to exhibit "easier" sub-classes and to provide approximability results.

Future works will be turned towards the search for *adaptive sampling policies*, breaking the sample set choice into successive steps, where the next sample is chosen on the basis of the outputs of previous samples. This generalised problem is likely to be PSPACE-complete. The similarities of this problem with the *Partially Observed Markov Decision Processes* [5] are worth exploring in order to design exact and approximate solution algorithms. In particular, simulation-based *Reinforcement Learning* approximate solution algorithms [3] will be investigated for this problem. These developements will be exploited to solve optimal spatial sampling problems in ecology/epidemiology.

REFERENCES

[1] T. Asano, K. Hori, T. Ono, and T. Hirata, 'A theoretical framework of hybrid approaches to MAX SAT', in *Proc. 8th Annual Int. Symposium on Algorithms and Computation*, LNCS 1350, pp. 153–162. Springer Verlag, (1997).

[2] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protasi, *Complexity and Approximation: Combinatorial optimization problems and their approximability properties*, Springer-Verlag, 1999.

[3] J. Baxter and P. L. Bartlett, 'Reinforcement learning in POMDP's via direct gradient ascent', in *Proc. 17th International Conf. on Machine Learning*, pp. 41–48. Morgan Kaufmann, (2000).

[4] J. Besag, 'On the statistical analysis of dirty pictures', *Journal of the Royal Statistical Society B*, **48**(3), 259–302, (1986).

[5] A. R. Cassandra, L. P. Kaelbling, and M. L. Littman, 'Acting optimally in partially observable stochastic domains', in *Proc. 11th Nat. Conf. on Artificial Intelligence (AAAI'94)*, pp. 1023–1028. AAAI Press, (1994).

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Interscience, 1991.

[7] U. Feige and M. X. Goemans, 'Approximating the value of two prover proof systems, with application to MAX 2-SAT and MAX DICUT', in *Proc. 3rd Israel Symposium on Theory of Computing and Systems*, ed., IEEE Computer Society, pp. 182–189, (1995).

[8] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-completeness*, Freeman, San Francisco, 1979.

[9] S. Geman and D. Geman, 'Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741, (1984).

[10] J. Hastad, 'Some optimal inapproximability results', in *Proc. of the 29th annual ACM Symposium on Theory Of Computing (STOC'97)*, pp. 1–10, New-York, NY, (1997).

[11] R. A. Howard, 'Information value theory', *IEEE Transactions on Systems Science and Cybernetics*, **SSC-2**(1), 22–26, (1966).

[12] D. Koller and N. Friedman, *Probabilistic Graphical Models – Principles and techniques*, The MIT Press, Cambridge, MA, 2009.

[13] A. Krause and C. Guestrin, 'Optimal value of information in gaphical models', *Journal of Artificial Intelligence Research*, **35**, 557–591, (2009).

[14] S. Z. Li, *Markov random field modeling in computer vision*, Springer, 1995.

[15] M. Munie and Y. Shoham, 'Optimal testing of structured knowledge', in *Proc. of 23rd AAAI Conference*, (2008).

[16] G. Nemhauser, L. Wosley, and M. Fisher, 'An analysis of the approximations for maximising submodular set functions', *Mathematical Programming*, **14**, 265–294, (1978).

[17] J. D. Park and A. Darwiche, 'Complexity results and approximation strategies for MAP explanations', *JAIR*, **21**, 101–133, (2004).

[18] Y. Radovilsky, G. Shattah, and S. Shimony, 'Efficient deterministic approximation algorithm for nonmyopic value of information in graphical models', in *SMC conference*, Taipei, Taiwan, (2006).

[19] Y. Radovilsky and S. Shimony, 'Observation subset selection as local compilation of performance profiles', in *Proc. of the 24th Annual Conf. on Uncertainty in Artificial Intelligence (UAI-08)*, pp. 460–467, Corvallis, Oregon, (2008). AUAI Press.

[20] J. S. Yedidia, W. T. Freeman, and Y. Weiss, 'Constructing free-energy approximations and generalized belief propagation algorithms', *IEEE transactions on information theory*, **51**(7), 2282–2312, (2005).