

Learning Parameters of Gibbs Random Fields Using Unconditional and Conditional MLE of Potentials

Georgy Gimel'farb

CITR, Department of Computer Science, Tamaki Campus, The University of Auckland, Private Bag 92019, Auckland, New Zealand

Abstract. Two parameter learning schemes for Gibbs random field image models with translation invariant multiple pairwise pixel interactions are discussed. The schemes allow to estimate both the interaction structure and strengths (Gibbs potentials) from a given training sample. The first scheme is based on the unconditional MLE of the potentials. The estimates are specified in an implicit form and can be obtained in three steps: (i) an analytic first approximation of the potentials for a big many possible neighbours, (ii) a search for most characteristic neighbours, and (iii) a stochastic approximation refinement of the estimates for a chosen set of neighbours. The second scheme uses the conditional MLE suggesting that the training sample has the least upper bound (top rank) in its total Gibbs energy within the parent population. This scheme allows to deduce an explicit, to scaling factors, analytic form of the potentials. Then only the scaling factors have to be learnt using their MLE in a like three-step manner. The conditional MLE of the potentials seems to be close to the unconditional ones and extends capabilities of the Gibbs image models.

1 Introduction

Gibbs random field (GRF) models describe digital grayscale images presented on a finite 2D lattice in terms of an explicit geometric structure and quantitative strengths of pixel interactions (see, for instance, [2, 3, 10]). The interaction between several pixels indicates how the marginal probability distribution of gray levels in these pixels differs from the like distribution for the independent random field (IRF): the larger the difference, the stronger the interaction.

Usually, the interactions are represented by a neighbourhood graph connecting the interacting pixels called the neighbours. In the Markov/Gibbs models, basic elements of the interaction structure are cliques, or complete subgraphs of the neighbourhood graph [2]. For brevity, the same terms “clique” and “clique family” are used for the like elements, but not the cliques in the strict sense, in the non-Markov Gibbs models [8]. These latter models take account of admissible gray range changes in the images.

The interaction strength for a given clique is specified by a (Gibbs) *potential* being a scalar function of the gray levels in the clique: the stronger the interaction, the higher the potential value [2].

We restrict the consideration only to the Markov and non-Markov Gibbs models with the translation invariant multiple pairwise pixel interactions [6, 8]. These models are promising in simulating and retrieving specific spatially homogeneous image textures [6, 8, 9] and in segmenting piecewise-homogeneous textures [7]. Such textures, called stochastic textures in [6, 8], have the interaction structure given by a first-order clique family of the pixels themselves and by several second-order clique families of the translation invariant pixel pairs.

Traditional Gibbs models, in particular, auto-binomial and auto-normal ones [2, 3, 10], possess mostly pre-defined interaction structures and potentials. Thus they have rather restricted potentialities in modeling different natural and artificial image textures. Models with multiple pairwise pixel interactions are adapted more easily to various textures because both the interaction structure and the potentials are learnt from a given training sample.

The paper is organised as follows. Section 2 gives a brief overview of the Gibbs image models of homogeneous grayscale image textures. Section 3 describes the learning scheme based on the unconditional MLE of the potentials. The potentials are implicitly specified by a system of stochastic equations to be solved first by an analytic and then by a stochastic approximation. Section 4 presents the alternative scheme based on the conditional MLE. The additional condition suggests that the training sample attains the least upper bound in total Gibbs energy within the parent population. This allows to deduce an explicit, to scaling factors, analytic form of the potentials and then approximate only these factors. Discussion and concluding remarks are given in Section 5.

2 Gibbs Models of Stochastic Image Textures

2.1 Assumptions and Notation

Let $\mathbf{R} = \{(m, n) : m = 0, \dots, M-1; n = 0, \dots, N-1\}$ be a 2D finite rectangular lattice with $M \cdot N$ pixels $i = (m, n)$ supporting digital grayscale images $g : \mathbf{R} \rightarrow \mathbf{Q}$. Here, $\mathbf{Q} = \{0, \dots, q_{\max}\}$ is a finite set of gray levels (GL).

Let \mathbf{A} be an index set for the pairwise clique families and $\mathbf{C}_a = \{(i, j) : i, j \in \mathbf{R}; i - j = \text{const}_a\}$ denote a particular family with $a \in \mathbf{A}$. This family has a specific pixel arrangement in the pairs specified by a fixed 2D inter-pixel shift $\text{const}_a \equiv (\mu_a, \nu_a)$: if $i = (m, n)$ and $j = (m', n')$ then $m' - m = \mu_a$ and $n' - n = \nu_a$.

The interaction strength for the first-order clique family \mathbf{R} is given by a Gibbs potential function $V : \mathbf{R} \rightarrow \mathcal{R}$ of a gray level (GL) in the pixel. Here, \mathcal{R} denotes the set of real numbers.

The interaction strength over each second-order family \mathbf{C}_a is assumed to be a function $V_a : \mathbf{D} \rightarrow \mathcal{R}$ of a GL difference (GLD) $d = g(i) - g(j) \in \mathbf{D} = \{-q_{\max}, \dots, 0, \dots, q_{\max}\}$ in the clique $(i, j) \in \mathbf{C}_a$.

2.2 Markov and Non-Markov Gibbs Models

The Markov/Gibbs model introduced in [6] takes into account only arbitrary shifts of the gray ranges. More flexible non-Markov Gibbs model in [8] allows

for arbitrary changes of the gray ranges: grayscale images $g \in \{g_1, g_2, \dots\}$ that differ only by gray ranges should have the same Gibbs probability as their normalised reference image g^{rf} . The gray range normalisation maps initial gray ranges $[\min_{i \in \mathbf{R}} g(i), \max_{i \in \mathbf{R}} g(i)]$ onto the maximum range $[0, q_{\max}]$. The non-Markov model embeds this normalisation $g \rightarrow g^{\text{rf}}$ directly into the Gibbs potentials so that the the Gibbs probability distribution (GPD) of the images is as follows:

$$\Pr(g|\mathbf{V}) = \frac{1}{Z_{\mathbf{V}}} \cdot \exp(E(g^{\text{rf}}|\mathbf{V})) \quad (1)$$

where $E(g^{\text{rf}}|\mathbf{V}) = e(g^{\text{rf}}|V) + \sum_{a \in \mathbf{A}} e_a(g^{\text{rf}}|V_a)$ is a *total Gibbs energy* of pixel interactions in the image g under the potentials $\mathbf{V} = (V, V_a : a \in \mathbf{A})$ for all the clique families. For brevity, indices “rf” of the normalised images are omitted below. Here, $e(g|V) = \sum_{i \in \mathbf{R}} V(g(i))$ denotes a *partial energy* of the pixelwise interactions, $e_a(g|V_a) = \sum_{(i,j) \in \mathbf{C}_a} V_a(g(i) - g(j))$ is a partial energy of pairwise pixel interactions for the clique family \mathbf{C}_a , $Z_{\mathbf{V}} = \sum_{g \in \mathcal{G}} \exp(E(g|\mathbf{V}))$ denotes a scaling factor, and \mathcal{G} is the parent population of all grayscale images supported by the lattice \mathbf{R} .

The parameter estimation schemes for the model of Eq. (1) are just the same as for the above-mentioned Markov/Gibbs model in [6] except for the gray range normalisation of a given training sample.

3 Learning Based on the Unconditional MLE of Potentials

3.1 Sufficient Statistics

The partial interaction energy can be represented as a dot product of the *centered* potential vector and the vector of relative sample GL or GLD frequencies, that is, of normalised GL or GLD histogram (H) collected over the reference image g (see [6, 8] for more detail):

$$\begin{aligned} e(g|V) &= |\mathbf{R}| \cdot \sum_{q \in \mathbf{Q}} V(q) \cdot F(q|g); \\ e_a(g|V_a) &= |\mathbf{R}| \cdot \rho_a \cdot \sum_{d \in \mathbf{D}} V_a(d) \cdot F_a(d|g). \end{aligned} \quad (2)$$

Here, $F(q|g) = \frac{1}{|\mathbf{R}|} \cdot \sum_{i \in \mathbf{R}} \delta(q - g(i))$ and $F_a(d|g) = \frac{1}{|\mathbf{C}_a|} \cdot \sum_{(i,j) \in \mathbf{C}_a} \delta(d - (g(i) - g(j)))$ are the normalised GLH and GLDH, respectively, $|\dots|$ denotes the set cardinality, $\rho_a = \frac{|\mathbf{C}_a|}{|\mathbf{R}|}$, and $\delta(\cdot)$ is the Kronecker function.

The potentials in Eq. (1) are centered as follows [6]:

$$\sum_{q \in \mathbf{Q}} V(q) = 0; \quad \forall a \in \mathbf{A} \quad \sum_{d \in \mathbf{D}} V_a(d) = 0 \quad (3)$$

that can be deduced from the unique representation of the Gibbs probability distribution (GPD) by a relative Hamiltonian [4]. This centering implies the similar centering of the histograms in Eq. (2). Below, both the potentials and the histograms are assumed to be centered.

The resulting exponential family representation of the GPD [1] shows that the centered GLH and GLDHs for all the clique families form *sufficient statistics* for the model of Eq. 1. It can be proven that conditions imposed in [1] to ensure strict log-concavity of the GPD (or unimodality of the likelihood function) hold for this model.

3.2 Learning the Model Parameters

Both the characteristic clique families and the potentials are learnt from a given training sample g° using analytic and stochastic approximation of the MLE of the potentials. This learning scheme introduced in [6, 8] is as follows:

(1) Analytic first approximation:

$$\begin{aligned} \forall q \in \mathbf{Q} \quad & V_{[0]}(q) = \lambda_{[0]} \cdot F(q|g^\circ); \\ \forall a \in \mathbf{A}; \quad \forall d \in \mathbf{D} \quad & V_{a,[0]}(d) = \lambda_{[0]} \cdot (F_a(d|g^\circ) - M_{\text{dif}}(d)) \end{aligned} \quad (4)$$

where $M_{\text{dif}}(d)$ denotes the centered marginal probabilities of the GLD for the IRF (it is easily shown that $M_{\text{dif}}(d) = \frac{1+q_{\max}-|d|}{(1+q_{\max})^2} - \frac{1}{1+2 \cdot q_{\max}}$) and the factor $\lambda_{[0]}$ is computed from the same centered normalised GLH and GLDHs, too (see [6] for greater detail).

- (2) Search for most characteristic interaction structure using approximate partial Gibbs energies of Eq. (2) with the potentials of Eq. (4) for comparing a big many possible clique families $a \in \mathbf{W}$ within a large search window \mathbf{W} .
- (3) Refinement of the potential estimates for the chosen families by a stochastic approximation technique proposed in [11].

The search window \mathbf{W} is specified by a given range of the intra-clique pixel shifts $|\mu_a| \leq \mu_{\max}$, $|\nu_a| \leq \nu_{\max}$ to be exhausted during the search. The partial energies over the search window form a *model-based interaction map* that allows to choose most characteristic clique families using an appropriate thresholding technique (see [6, 9] for more detail).

The above learning scheme gives good results in simulating different natural textures [6, 8]. These results show that there exist such natural textures that can be considered as the stochastic ones described by the model of Eq. (1) with the learnt parameters. But, such a learning involves rather big number of the potential values to be refined individually by stochastic approximation: in total, $q_{\max} \cdot (2 \cdot |\mathbf{A}| + 1)$ scalar values. Below, we consider a somewhat different learning

approach based on an *explicit*, to scaling factors, form of the potentials. In this case the number of parameters to be computed and refined is reduced to only the number $|\mathbf{A}|$ of the clique families.

4 Learning Based on the Conditional MLE of Potentials

4.1 Explicit Form of the Potentials

The explicit form of the potentials is deduced under a specific ranking of the training sample g° within \mathcal{G} . The desired potential estimates are obtained using the MLE of the scaling factors or, what is the same, the *conditional* MLE of the potentials (CMLE) provided that the training sample g° attains the least upper bound (top rank) which is feasible for it within \mathcal{G} in the total Gibbs energy [5].

Let us rank the images $g \in \mathcal{G}$ in ascending order of the partial Gibbs energy (in particular, $e(g|V)$ for the pixelwise family \mathbf{R} or $e_a(g|V_a)$ for the pairwise family \mathbf{C}_a ; $a \in \mathbf{A}$). It is easily seen from Eqs. (1) and (2) that this ranking is *invariant* to potential (and energy) normalisation that reduces the corresponding potential vector $\mathbf{V} = (V(q) : q \in \mathbf{Q})$ or $\mathbf{V}_a = (V_a(d) : d \in \mathbf{D})$ to the unit vector $\mathbf{v} = \frac{\mathbf{V}}{|\mathbf{V}|}$ or $\mathbf{v}_a = \frac{\mathbf{V}_a}{|\mathbf{V}_a|}$, respectively.

Let $\mathbf{F}(g^\circ) = (F(q|g^\circ) : q \in \mathbf{Q})$ and $\mathbf{F}_a(g^\circ) = (F_a(d|g^\circ) : d \in \mathbf{D})$ denote the vectors of the *centered* marginal GL sample frequencies and of the marginal GLD sample frequencies for the clique family \mathbf{C}_a , respectively. Then it is readily shown that unit vectors $\mathbf{v}^\circ = \frac{\mathbf{F}(g^\circ)}{|\mathbf{F}(g^\circ)|}$ and $\mathbf{v}_a^\circ = \frac{\mathbf{F}_a(g^\circ)}{|\mathbf{F}_a(g^\circ)|}$ maximise the normalised partial energies $e(g^\circ|\mathbf{v})$ and $e_a(g^\circ|\mathbf{v}_a)$, respectively.

Each potential vector obtained by arbitrary scaling of such a unit vector places the training sample g° to the same least top rank in a corresponding partial energy among the samples $g \in \mathcal{G}$ as compared to any other potential vector. Let the *least top rank principle* be applied to the ranking in the partial energy for each the clique family of the model in Eq. (1). Then the potentials $\mathbf{V} = (V, V_a : a \in \mathbf{A})$ ranking the training sample to a feasible top place within \mathcal{G} in the *total* Gibbs energy have to possess the following explicit, to scaling factors, form:

$$\mathbf{V}^\circ(\Lambda) = (\lambda \cdot \mathbf{F}(g^\circ), \lambda_a \cdot \mathbf{F}_a(g^\circ) : a \in \mathbf{A}). \quad (5)$$

Here, $\Lambda = (\lambda, \lambda_a : a \in \mathbf{A})$ is a vector of arbitrary positive scaling factors.

4.2 Conditional MLE of the Potentials

Therefore, the CMLE of the Gibbs potentials $\mathbf{V}^* \equiv \mathbf{V}^\circ(\Lambda^*)$ for the image model of Eq. (1) is as follows:

$$\mathbf{V}^* = (\lambda^* \cdot \mathbf{F}(g^\circ), \lambda_a^* \cdot \mathbf{F}_a(g^\circ) : a \in \mathbf{A}) \quad (6)$$

where the desired scaling factors are computed by maximising the likelihood function $L(\Lambda|g^\circ) = \ln \Pr(g^\circ|\mathbf{V}^\circ(\Lambda))$:

$$\Lambda^* = \arg \max_{\Lambda} L(\Lambda|g^\circ). \quad (7)$$

Generally, this CMLE may differ from the unconditional MLE of the potentials. But, for the GPD of Eq. (1) both the estimates are supposedly fairly close if not equivalent. This conjecture that needs further theoretical investigations has some supporting considerations: in particular, (i) the CMLE of Eq. (6) and the analytic first approximation of the unconditional MLE of Eq. (4) have very similar forms, (ii) the refined potentials are usually close to their first approximations, and (iii) the samples possess fixed ranks not only under the uniform scaling of the potentials but also, by symmetry, under the like scaling of the centered sample histograms.

4.3 Learning the Factors

The desired factors of Eq. (7) are learnt in a similar way as the potentials themselves in [6, 8]: first, by the analytic first approximation and searching for a characteristic interaction structure and then by refining the factors for chosen clique families using the stochastic approximation.

The analytic first approximation of the factors is obtained by a truncated Taylor's series expansion of the likelihood function $L(A|g^\circ)$ about the zero point $A = \mathbf{0}$. This technique is quite similar to that proposed in [6, 8] and results in the following approximation:

$$\begin{aligned} \lambda_{[0]} &= \alpha_{[0]} \cdot \varepsilon_{[0]}; \\ \forall a \in \mathbf{A} \quad \lambda_{a,[0]} &= \alpha_{[0]} \cdot \varepsilon_{a,[0]}, \end{aligned} \quad (8)$$

where

$$\varepsilon_{[0]} = \sum_{q \in \mathbf{Q}} F^2(q|g^\circ)$$

and

$$\varepsilon_{a,[0]} = \rho_a \cdot \sum_{d \in \mathbf{D}} (F_a(d|g^\circ) - M_{\text{dif}}(d)) \cdot F_a(d|g^\circ)$$

are relative pixelwise and pairwise total Gibbs energies about the zero point, respectively. The scale factor $\alpha_{[0]}$ is computed from these energies as follows:

$$\alpha_{[0]} = \frac{\varepsilon_{[0]}^2 + \sum_{a \in \mathbf{A}} \varepsilon_{a,[0]}^2}{\varepsilon_{[0]}^2 \cdot U_{[0]} + \sum_{a \in \mathbf{A}} \varepsilon_{a,[0]}^2 \cdot U_{a,[0]}} \quad (9)$$

where

$$\begin{aligned} U_{[0]} &= \sum_{q \in \mathbf{Q}} \sigma_{\text{irf}} \cdot F^2(q|g^\circ); \\ U_{a,[0]} &= \rho_a \cdot \sum_{d \in \mathbf{D}} \sigma_{\text{dif}} \cdot F_a^2(d|g^\circ), \end{aligned} \quad (10)$$

and σ_{irf} and σ_{dif} denote variances of the marginal frequencies of the GL and GLD for the IRF, respectively.

Search for the interaction structure exploits in this case the weighted relative partial energies of pairwise pixel interactions: $\mathbf{e}_{[0]} = \{\omega_{a,[0]} \cdot \varepsilon_{a,[0]} : a \in \mathbf{W}\}$ where the weight $\omega_{a,[0]} = \rho_a \cdot \sum_{d \in \mathbf{D}} F_a^2(d|g^\circ)$.

Stochastic approximation refinement of the factors exploits the similar partial energies that depend on a proximity between the marginal GL and GLD frequencies for each clique family in the training image sample g° and a sample generated by pixelwise stochastic relaxation using the current factors. At each step t of the stochastic approximation, the current factors Λ_t are updated as follows:

$$\begin{aligned} \lambda_{[t+1]} &= \lambda_{[t]} + \alpha_{[t]} \cdot \varepsilon_{[t]}(g_{[t]}); \\ \forall a \in \mathbf{A} \quad \lambda_{a,[t+1]} &= \lambda_{a,[t]} + \alpha_{[t]} \cdot \varepsilon_{a,[t]}(g_{[t]}). \end{aligned} \quad (11)$$

Here, $g_{[t]}$ is an image sample generated at the step t , $\alpha_{[t]}$ denotes the current scaling factor decreasing from the starting value $\alpha_{[0]}$ in Eq. (9) as $\frac{c_0+1}{c_1+c_2 \cdot t}$ (see [11] for theoretical and empirical choices of the control values c_0 , c_1 , and c_2), and $\varepsilon_{[t]}(g_{[t]})$ and $\varepsilon_{a,[t]}(g_{[t]})$ are the current differential partial energies:

$$\begin{aligned} \varepsilon_{[t]}(g_{[t]}) &= \sum_{q \in \mathbf{Q}} \Delta(q|g_{[t]}) \cdot F(q|g^\circ); \\ \varepsilon_{a,[t]}(g_{[t]}) &= \rho_a \cdot \sum_{d \in \mathbf{D}} \Delta_a(d|g_{[t]}) \cdot F_a(d|g^\circ). \end{aligned} \quad (12)$$

where $\Delta(q|g_{[t]}) = F(q|g^\circ) - F(q|g_{[t]})$ and $\Delta_a(d|g_{[t]}) = F_a(d|g^\circ) - F_a(d|g_{[t]})$.

5 Concluding Remarks

Both the learning schemes approximate closely the model parameters using the sample GLH and GLDHs. But, the learning scheme based on the unconditional MLE of the potentials is notably less robust. It exploits the assumption that marginal relative sample frequencies, obtained by normalising the sample GLH and GLDHs, are consistent statistical estimates of the corresponding marginal probabilities under the given model of Eq. (1). But, the bigger the number of the gray levels $|\mathbf{Q}|$, the larger the size of the training sample to obtain such estimates. To simplify the choice of the training samples, for texture simulation experiments in [6, 8] this number has been restricted to $|\mathbf{Q}| = 16$.

In practice, there are much more GLs, usually, $|\mathbf{Q}| = 256$. Therefore the training samples are usually too small to contain all the possible GLs and GLDs. In this case one or another known robust approximation to the unobserved marginals from the obtained incomplete sample histograms may be implemented for getting valid potential estimates. This is quite possible if the learning scheme is based on the CMLE because errors in the differential partial energies

of Eq. (12) due to such approximation influence, during the refinement process of Eq. (11), only the scaling factors but not the overall form of the potentials specified by the approximated GLH and GLDHs.

But, the scheme based on the unconditional MLE of the potentials, updates independently each the potential value $V_{[t]}(q)$ for $q \in \mathbf{Q}$ and $V_{a,[t]}(d)$ for $d \in \mathbf{D}$ and $a \in \mathbf{A}$ using the distinctions between the corresponding marginal frequencies for the training and current generated samples. Therefore, the unobserved marginals have to be approximated at each stochastic approximation step t , that is, for each currently generated image $g_{[t]}$. In this case the computational complexity of the learning increases as compared to the alternative scheme and the approximation errors may result in unpredictable errors in the final potential estimates.

References

1. Barndorff-Nielsen, O.: Information and Exponential Families in Statistical Theory. Wiley (1978)
2. Besag, J. E.: Spatial interaction and the statistical analysis of lattice systems". J. Royal Statistical Soc. **B36** (1974) 192-236
3. Chellappa, R., Jain, A. (Eds): Markov Random Fields. Theory and Application. Academic Press (1993)
4. Dobrushin, R. L., Pigorov, S. A.: Theory of random fields. Proc. 1975 IEEE-USSR Joint Workshop on Information Theory, December 15-19, 1975, Moscow, USSR. IEEE (1976) 39-49
5. Gimel'farb, G. L.: Analytic approximation of Gibbs potentials to model stochastic textures. Proc. First Joint Australia & New Zealand Biennial Conf. on Digital Image & Vision Computing: Techniques and Applications (DICTA'97/IVCNZ'97), December 10-12, 1997, Albany, Auckland, New Zealand. Massey University (1997) 153-158
6. Gimel'farb, G. L.: Texture modeling by multiple pairwise pixel interactions. IEEE Trans. Pattern Anal. Machine Intell. **18** (1996) 1110-1114
7. Gimel'farb, G. L.: Gibbs models for Bayesian simulation and segmentation of piecewise-uniform textures. Proc. 13th IAPR Int. Conf. on Pattern Recognition, August 25-29, 1996, Vienna, Austria. TUWien (1996) **B** 591-595
8. Gimel'farb, G. L.: Non-Markov Gibbs texture model with multiple pairwise pixel interactions. *Ibid*, 760-764
9. Gimel'farb, G. L., Jain, A. K.: On retrieving textured images from an image database. Pattern Recognition **29** (1996) 1461-1483
10. Winkler, G.: Image Analysis, Random Fields, and Dynamic Monte Carlo Methods. Springer (1995)
11. Younes, L. Estimation and annealing for Gibbsian fields. Annales de l'Institut Henri Poincare. **24** (1988) 269-294