HIERARCHICAL ANNEALING FOR SCIENTIFIC MODELS

S. K. Alexander[1] *P. Fieguth*[2] *E. R. Vrscay*[1]

[1]Dept. of Applied Mathematics [2]Dept. of Systems Design Engineering
University of Waterloo, Waterloo, Ontario, Canada, N2L-3G1
{sk2alexa|pfieguth|ervrscay}@uwaterloo.ca

ABSTRACT

The computational complexity of simulated annealing makes it an impractical tool in many applications, particularly for complex, non-local models on very large 2D and 3D domains as desired in many scientific contexts. In particular, it is very difficult to produce large scale structure from a fine, pixellated lattice. Thus a hierarchical approach is intuitively attractive. However, existing approaches are few and limited. Motivated by a current problem in porous media, we develop a hierarchical approach to complex model sampling. In experiments, this approach results in 1–2 orders of magnitude computational gain, and significant gains in convergence as well.

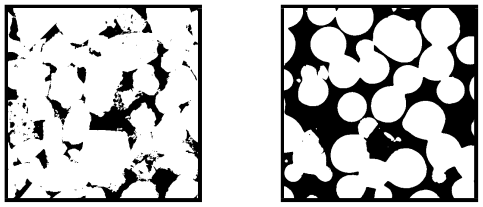
1. INTRODUCTION

In this work we are interested in the problem of computational practicality for simulated annealing in large phase spaces. We choose a motivational application: generation of samples of binary porous media images (e.g. Fig.1), the importance of which is discussed in [1].

There are two central ideas discussed in this paper. The first key idea is that insights from renormalization approaches in Markov Random Field (MRF) sampling suggest a hierarchical approach to related problems. The second key point is that while the scaling behaviour of local-interaction models is difficult to analyze (e.g. difficult to renormalize), there are other models involving non-local quantities that are inherently rescaleable.

We describe how our problem relates to various parts of the annealing literature, and propose a hierarchical annealing approach. Several models are discussed, and numerical experiments reported. These empirical results show significant improvement over approaches in the porous media literature [1, 2].

Supported in part by NSERC research grants (P.F. and E.R.V) and a NSERC PGS scholarship (S.K.A). M. Ioannidis was helpful in providing porous media images.



(a) (b)

Fig. 1: Binary porous media examples

2. ANNEALING APPROACHES

Simulated Annealing [3] has seen a vast range of applications in image synthesis and estimation. Its strength lies in flexibility in choice of prior and measurement models. However there is a drawback in computational complexity — particularly so for large domains and complex models. The overwhelming majority of SA implementations involve a single (flat) lattice, which is slowly annealed by brute force. In many applications, this is computationally impractical.

The reason for this computational cost is that, by their very nature, local MRF models are subject to a phenomenon of *critical slowing down*. The issue essentially is this: all structure in the model comes from local interactions, hence when longer scale (i.e. non-local) structure exists, changes in this structure can only be made after many local interactions. At phase transitions (the crossing of *critical* temperatures), non-local structure appears in the sample. Thus it is at exactly these points where significant changes to the energy become increasingly more expensive (more local interactions needed), *slowing down* convergence. This is a key issue to be attacked if one wishes to accelerate an annealing process.

To be sure, hierarchical approaches have been proposed. With few exceptions, these fall into two classes:

1. Hierarchical estimation assuming a dense first scale, and

2. Region-based sampling.

Region-based sampling (e.g. clustering) has been successfully applied to address the computational problem in local models, but it is specialized to Ising (or Ising-like) models at or near the critical temperature, where Ising exhibits the most interesting behaviour (phase change) and structures. Our models are quite different from the Ising class, and exhibit complex structure down to very low temperatures (i.e. for long temperature scales as we anneal) so standard region-based methods do not apply.

Many authors have discussed methods to use a hierarchical approach to accelerating image estimation. In particular a "label" pyramid is build above the image at the finest scale [4, 5]. However. given a dense image at the finest scale, the estimation problem is very well-conditioned, and converges well with even only the most rudimentary information from the coarser scales. Indeed, in most cases the coarse scales are not even annealed.

Because our scientific problems of interest may involve sparse data (or, indeed, no data in the case of pure sampling), the former approaches are not effective. We seek a hierarchical approach in which the large-scale structure is efficiently determined at coarse scales, leaving only more detailed refinements at finer scales.

Our proposal is to apply the central idea of renormalization theory to help reduce the computational cost of annealing. The key insight that we bring out of renormalization methods is this: the effective temperature for a given feature size is scale dependent. For some temperature at some intermediate scale, coarser scales are cold (meaning that the large, coarse features are 'frozen'), and finer scales are hot (meaning that the tiny features, not resolved at the intermediate scale, are rapidly changing). This implies that we need concentrate on only one intermediate scale at a time.

Some related work has been done, particularly in renormalization for classic image estimation problems [6]. However there are some interesting open questions, and our approach is distinct from existing research on hierarchical structures and accelerated annealing.

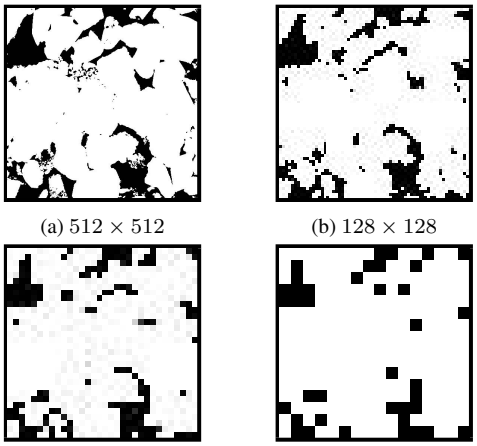
3. HIERARCHICAL SAMPLING

We are interested in simulated annealing, which is formalized as sampling from a Gibbs random field with density:

$$\pi_\beta(x) = \frac{e^{-\beta \mathcal{E}(x)}}{Z_\beta} \quad , \tag{1}$$

where $\beta = 1/T$ is the inverse temperature parameter. The partition function Z_β is not evaluated in simulated annealing. The process of doing this on a flat scale is usually based on [3]. We wish to approach this in a hierarchical way, by rescaling the model at appropriate temperatures.

The question then is: At any given level in this multiscale hierarchy, what image features are represented? In particular, we wish to work 'down' a hierarchy, from coarsest to finest resolution. How may we anneal in such a way that features are represented at the current level, and can be meaningfully projected to the the next finest level? Figure 2 illustrates this point.



(a) 512×512 (b) 128×128 (c) 64×64 (d) 32×32

Fig. 2: A porous media image viewed at several resolutions: How do local and non-local features scale?

Our approach to hierarchical annealing is as follows. Consider a hierarchy $\{X_n\}_{n=0}^{M}$ of coarse-grainings of the configuration space where each increase in level represents decimation by a factor of two (X_0 is finest resolution). At each higher level in the hierarchy, the energy function for that level is $\mathcal{E}^s$. We denote projection (coarse to fine) from level s to level $s-1$ as P_{s-1}. Annealing is performed as shown in Algorithm 1.

Algorithm 1 Hierarchical Annealing

```
k ⇐ 0
for s = S to 0 do
  while E^s(X_s) not converged do
    β ⇐ 1/T_k
    X_s ⇐ sample π_β        { draw a sample from π_β }
    k ⇐ k + 1
  end while
  X_{s-1} ⇐ P_{s-1}(X_s)    {map to next finer resolution}
end for
```

There are two sources of computational benefit in this approach. First, the size of the coarse domains is small, allowing rapid iterations of the sampler. Second, as suggested in the previous section, at an intermediate scale the algorithm needs to iterate only long enough to allow relatively local structure to converge, since the larger structures converged at coarser scales. Hence we work in a decimated configuration space (at less computational cost) until some appropriate time, and then project onto the next larger space and continue annealing.

Annealing with too low an initial temperature (not en-

ough energy) approaches a greedy algorithm, prone to finding local minima. On the other hand, too high an energy may destroy larger structures. Since this holds true at any level in the hierarchy, clearly there is a delicate balance in achieving computational gains while retaining good optimization performance.

An analysis of the interactions between the annealing (cooling) schedule, the scale-to-scale projection, and the convergence of the stochastic sampler is not straightforward. At present, we rely on heuristic rules to determine the cooling/projection schedule: at each level the process is annealed until it has converged (in the sense of energy). This heuristic is not optimal in any sense, although, the experiments which follow show it to be effective.

4. MODELS

The statistical sampling methods discussed here are characterized in terms of an energy function $\mathcal{E}$ in (1) . When considering a hierarchical approach, the key problem is how to define it at coarser scales. The canonical example of a local MRF model on a lattice of binary values is the *Ising model*. The usual simplified form (with no external field) can be written as:

$$\mathcal{E}_{\text{Ising}} = J \sum_{[i,j]} \sigma_i \sigma_j \quad . \tag{2}$$

where J is a coupling coefficient, σ_n is the spin of the n^{th} particle, and summation is taken over $[i, j]$, adjacent horizontal and vertical neighbours only.

The difficulty in considering hierarchical annealing with this model is to decide how J ought to scale. As previously described, the nature of this model leads to difficulties in the analysis. Because of the importance of the Ising model, a lot of effort has been spent on this question, and several approaches have been used with success, but they were not easily found.

In contrast, some non-local descriptions are inherently rescaleable. In the motivating application of porous media, two very important characterizations of structure are the *two-point correlation* function, and the distribution of *chord lengths*.

We are considering binary dense/pore structures [2] in an image. We can denote $I^s(x)$ for the index function for our binary image at scale s, (yielding 0 for pore, and 1 for density). If we let $< \cdot >$ denote a spatial average over the image, then the average density ϕ (or "one-point" correlation can be denoted as

$$S^s(r) = < I^s(x + r) > = \phi^s \quad . \tag{3}$$

Similarly, two-point correlation:

$$S^s(r_1, r_2) = < I^s(x + r_1) I^s(x + r_2) > \quad . \tag{4}$$

Notationally, let us take $\hat{\cdot}$ to denote trained/target values. Furthermore, denote the lattice size of the image at a particular scale s as $\mathcal{O}(s)$ (then for initial lattice of $N \times N$ $\mathcal{O}(s) = N/2^s$).

One-point energy function: With the above notation, $\hat{\phi}^s$ denotes target values for the image density at scale s and we can express an energy function for this scale based on one-point correlation as

$$\mathcal{E}_1^s = \| \hat{\phi}^s - \phi^s \| \quad . \tag{5}$$

Two-point energy function: Furthermore, if we restrict ourselves to the horizontal and vertical directions (to reduce computations), two-point correlation gives us

$$\mathcal{E}_2^s = \sum_{r=1}^{\mathcal{O}(s)/2} \| \hat{S}^s(0, r) - S^s(0, r) \| + \| \hat{S}^s(r, 0) - S^s(r, 0) \| \quad . \tag{6}$$

Chordlength energy function: Another interesting measurement is the distribution of chord-lengths in the image [2]. If we again restrict ourselves to the horizontal and vertical directions again, this is essentially the distribution of length of contiguous "runs" of density pixels in these directions. Denoting these probability mass functions as p_C^h and p_C^v for the horizontal and vertical directions, respectively, we may construct an energy function:

$$\mathcal{E}_3^s = \sum_{n=1}^{\mathcal{O}(s)} \| \hat{p}_C^h(n) - p_C^h(n) \| + \| \hat{p}_C^v(n) - p_C^v(n) \| \quad . \tag{7}$$

Here the sample pmf's are estimated by histograms from the image data.

Local histogram energy function: A similar difference-of-pmf's function arises from the approach of histogramming local neighbourhoods described in [7, 8]. This method has the properties of being very cheap to compute, and constraining local pixel configurations. Distributions are described for neighbourhoods surrounding both white and black pixels. So if we denote these distributions as p_H^w and p_H^b, and let the neighbourhood size be b, we can define an energy function:

$$\mathcal{E}_4^s = \sum_{n=1}^{2^b} \| \hat{p}_H^w(n) - p_H^w(n) \| + \| \hat{p}_H^b(n) - p_H^b(n) \| \quad . \tag{8}$$

There are a couple of things to note in the above: The distance used in the various energy functions is not specified — there are several possibilities in most cases. In the numerical experiments reported here, the l_2-norm is used, but the best choice is not clear.

Since we can look at separate scales, and also mixtures of these models, the general formulation will be

$$\mathcal{E}^s(x) = \sum_i c_i \mathcal{E}_i^s(x) \qquad c_i > 0 \; \forall i \quad . \tag{9}$$

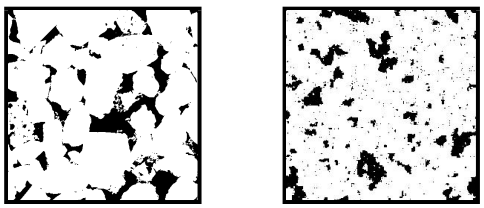
It is interesting to consider the scale-dependent behaviour of such models. For example, consider a proposed model of

$$\mathcal{E} = (\mathcal{E}_1 + \mathcal{E}_2) + \alpha\mathcal{E}_3 \ . \tag{10}$$

The mixing parameter α can be seen to balance fine behaviour against coarse (if we set up the $\mathcal{E}$'s correctly). It is intuitive that some energy functions may be more relevant at some scales than others.

5. RESULTS & CONCLUSIONS

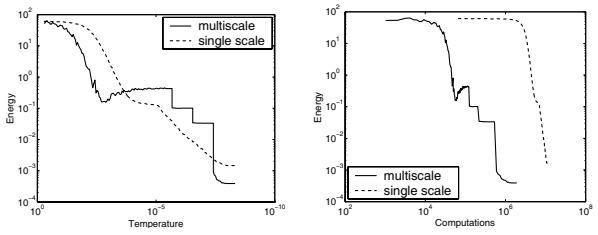
Experiments have been run with different data sets and energy functions. The following results are for images with final resolution of 256×256 pixels. The included example shows a real sandstone image 3(a) from one training set, and a sample 3(b) from the hierarchical method describe here using a mixture of two-point and chordlength energy functions (i.e. $\mathcal{E} = \mathcal{E}_2 + \alpha\mathcal{E}_3$)



(a) (b)

Fig. 3: real (a) and sampled (b) porous media images

For typical annealing runs on this data, a comparison of sampling performance for flat and hierarchical annealing is shown in Figures 4(a) and 4(b).



(a) (b)

Fig. 4: Convergence curves for flat and hierarchical sampling. Energy vs. (a) Temperature and (b) Computations

Figure 4(a) demonstrates good convergence of the hierarchical annealing sample. Figure 4(b) shows the large computational gain made by using this approach.

In these graphs, the two processes used the same cooling schedule (geometric, with $T_{n+1} = 0.9T_n$) and all energies are measured at the finest scale (by projection, in the hierarchical case). Note that this constrains the hierarchical annealing in a way that may not be optimal. Indeed, the convergence of the hierarchical process shows one areas that are not monotonically decreasing in energy — this seems to be an artifact of the cooling schedule's interaction with projection in the hierarchy.

Following these outstanding gains in convergence and computational cost, we are motivated to pursue this approach further. Additional discussion of the work in progress will appear at the website:
`http://ocho.uwaterloo.ca/~sk2alexa`

Ongoing directions for this work include further analysis of the hierarchical cooling schedule. We are also interested in analysis of random sampling by annealing to nonzero temperatures. At final temperatures $T > 0$, metastable structures will be present and sampling time potentially reduce greatly . However, in this case we are no longer sampling from the null space of a model, and the hierarchical annealing process is more subtle.

We have demonstrated concrete benefits to the hierarchical annealing approach, and also sketched some of the many interesting directions in which this work can be taken.

6. REFERENCES

[1] Z. Liang, M.A. Ioannidis, and I Chatzis, "Reconstruction of 3d porous media using simulated annealing," in *Computational Methods in Water Resources XIII*, Bentley et al., Ed., Balkema, Rotterdam, 2000.

[2] M. S. Talukdar, O. Torsaeter, and M. A. Ioannidis, "Stochastic recontruction of particulate media from two-dimensional images," *Journal of Colloid and Interface Science*, vol. 248, pp. 419–428, 2002.

[3] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Trans. on PAMI*, vol. 6, pp. 721–741, 1984.

[4] Z. Kato, M. Berthod, and J. Zerubia, "A hierarchical markov random field model . . . ," *GMIP*, vol. 58, no. 1, pp. 18–37, 1996.

[5] C. Bouman and M. Shapiro, "A multiscale random field model for bayesian image segmentation," *Image Processing*, vol. 3, no. 2, pp. 162–177, March 1994.

[6] B. Gidas, "A renormalization group approach to image processing problems," *IEEE Trans. on PAMI*, vol. 11, no. 2, pp. 164–180, 1989.

[7] S. K. Alexander, P. Fieguth, and E. R. Vrscay, "Hierarchical annealing for random image synthesis," in *EMMCVPR'03*. 2003, number 2683 in LNCS, Springer.

[8] S. K. Alexander, P. Fieguth, and E. R. Vrscay, "Image sampling by hierarchical annealing," in *ICIP'03*. IEEE, 2003.