Parameterized Hierarchical Annealing for Scientific Models

Simon K. Alexander[1]*, Paul Fieguth[2], and Edward R. Vrscay[1]

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

Abstract. The accurate synthesis of binary porous media is a difficult problem. Initial applications of simulated annealing in this context with small data sets and simple energy functions have met with limited success. Simulated annealing has been applied to a wide variety of problems in image processing. Particularly in scientific applications such as discussed here, the computational complexity of this approach may constrain its effectiveness; complex, non-local models on large 2D and 3D domains may be desired, but do not lend themselves to traditional simulated annealing due to computational cost. These considerations naturally lead to a wish for hierarchical/multiscale methods. However, existing methods are few and limited. In this paper a method of hierarchical simulated annealing is discussed, and a simple parameterization proposed to address the problem of moving through the hierarchy. This approach shows significant gains in convergence and computational complexity when compared to the simulated annealing algorithm.

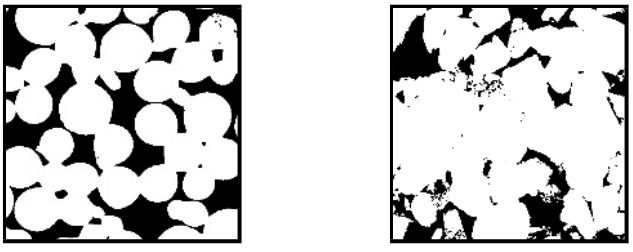
1 Introduction

We are interested in the problem of computational practicality of simulated annealing in large phase spaces. As is often the case, a constrained problem domain allows concentration on particular issues of interest. Hence, we choose as a motivational application the synthesis of binary porous media images. Figure 1 gives two examples, binary images representing density (white) and pore structures in a physical media (hence the name). Images such as these are important in the study of porous media [11]. The above examples, however, are physically imaged; researchers in the area are interested in ways to accurately synthesize such data. Resulting data sets can be used to perform many useful calculations [10].

The simulated annealing algorithm has been used to perform this sort of synthesis with some success [11]. Ultimately the computational complexity has limited the viability of this approach. Simulated annealing [6] has been used successfully in many

* This research was supported in part by the Natural Sciences and Engineering Council of Canada (NSERC), in the form of grants (P.F. and E.R.V.) and a Postgraduate Scholarship (S.K.A.). We also wish to thank M. Ioannidis, Dept. of Chemical Engineering University of Waterloo for providing sample porous media images and interesting discussions.

A. Campilho, M. Kamel (Eds.): ICIAR 2004, LNCS 3211, pp. 236–243, 2004.



(a) (b)

Fig. 1. Examples of binary porous media a) sintered glass spheres, b) sandstone

imaging applications, particularly in estimation problems. However, the computational cost of slow annealing schedules has limited its applicability, especially where large phase spaces are involved. In this work, we are interested in extending the discussion of *hierarchical* methods, primarily as a way to reduce the computational complexity.

Simulated annealing is controlled by a *cooling schedule* — a decreasing (although perhaps not strictly so) temperature parameter which affects the likelihood of non-energetically-favourable events. Theoretical results only exist for impractically slow cooling, or special cases [5]. Our work is motivated by a wish to leverage multiscale characteristics of a model to reduce the amount of computation needed.

The general approach of hierarchical annealing [3, 1, 2] is inspired in part, by renormalization group approaches in Markov random fields (e.g., [7]). A key realization is that while the scaling behavior of local-interaction models may be very difficult to analyze (hence renormalization difficult), other models may be proposed based on non-local quantities which are inherently renormalizable.

After briefly describing annealing approaches and how this example problem fits in, hierarchical annealing will be outlined. In this paper, we concentrate on the issue of parameterizing the hierarchical approach, and discuss the application to a particular model with comparison to 'flat' annealing methods. Empirical results are presented, along with a discussion of generalizations and future work.

2 Simulated Annealing and Hierarchical Approaches

Since its introduction to image processing by Geman & Geman [6], simulated annealing has been used for a large range of applications to image estimation and synthesis. The flexibility of this model is its main strength. If a problem can be stated in terms of minimum energy states of a Gibbs type density (1), or equivalently a Markov random field, simulated annealing will give a correct solution *given a long enough cooling schedule*. The usual way of describing such a density is

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

where $\beta = 1/T$ is the inverse temperature parameter. The partition function Z_β is not evaluated in simulated annealing.

However, the primary drawback is that while in theory this approach will work, this result relies on an intractably slow *logarithmic cooling schedule* in temperature T : $T_{n+1} \sim 1/\log T_n$.

Even computationally tractable cooling schedules (which are not proven to converge, in general) are expensive — especially for complex models and large domains. The majority of annealing implementations use a single ('flat') lattice at a particular resolution, annealed by brute force. For many interesting applications, this approach is simply too computationally expensive to be practical.

The fundamental reason for the slow convergence of such models is that local MRF models are inherently subject to a phenomena known as *critical slowing down*. Essentially what happens is the following: since all structure in the model is created by local interactions, all longer scale (i.e., non-local) structure *must* be created by many local interactions. Recall that we are obeying a *cooling schedule* which reduces the 'temperature' of the simulation over time. At *phase-transition* points (i.e. *critical temperatures*), non-local structure will appear. Once this temperature is reached, significant changes in the energy become more expensive, since the non-local structure must change through the cumulative effect of many local interactions. In order to address the cost of sampling such a model, attacking this critical slowing down is vital.

The method we are describing here, *hierarchical annealing* is an approach to do just this. Hierarchical approaches have been proposed in the literature. Primarily, they are of two types:

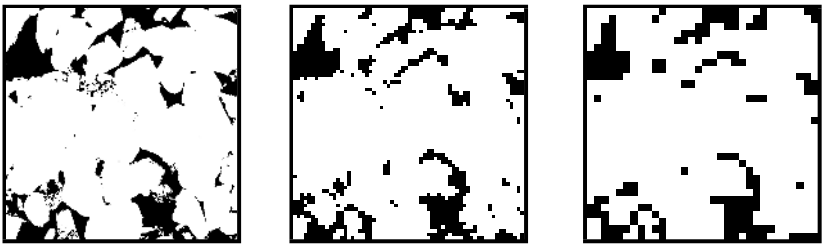
1. Hierarchical estimation (assuming a dense first scale), and
2. Region-based sampling methods.

Methods for region-based sampling have been applied successfully to the computation of local models in very particular cases – for example clustering methods for Ising or Ising-like models at or near the critical temperature. (Here Ising exhibits the interesting phase-change behaviour; studying this and the structures created near this temperature has driven much of the interest in these sorts of methods). Our models are in general quite different than this class, especially in that they exhibit complex structure at very low temperatures or, equivalently, for long temperature scales as we anneal.

There are several authors who have discussed hierarchical approaches to the related problem of accelerating image estimation. In particular, a 'label' pyramid may be built above the (full resolution, finest scale) image [9, 4, 8]. This is, however, quite different than our situation. Given a dense image at finest scale, this estimation problem tends to be well-conditioned, and converges well with very little information from coarser scales. In fact, often the coarse scale state is not annealed at all.

2.1 Hierarchical Annealing

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

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

Our approach to hierarchical annealing [1] is outlined 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 scale s from coarsest to finest do
  while E^s(X_s) not converged do
    β ⇐ 1/T_k
    X_s ⇐ apply Gibbs sampler to X_s
    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. So we work in a decimated configuration space (at less computational cost) until some appropriate condition is reached, and then project onto the next larger space and continue annealing.

Annealing with too low an initial temperature (not enough energy) approaches a greedy algorithm, that is, it will be prone to getting stuck in local minima. On the other hand, too high an energy may destroy larger structures (all such structure may be taken apart if you anneal for long enough). 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 (i.e., low temperatures).

Analysis of the interactions between the annealing (cooling) schedule, the scale-to-scale projection, and the convergence of the stochastic sampler is not straightforward. In this work, we present a discussion of a simple parameterization of the process, which is in no way presumed to be optimal. Empirical results will show this approach to be beneficial when applied to our example. Further work will refine these methods; from the preceding it should be clear that there are several related avenues for improvement.

2.2 Parameterization for H-A

Due to the aforementioned difficulties, choosing a 'best' way to parameterize the hierarchical annealing is not straightforward. In this work, we have explored a simple parameterization and present the benefits when compared to 'flat' annealing.

Parameterization is made more difficult by the fact that it is difficult to distinguish the effect of varying the cooling schedule and the number of iterations of the system. To simplify things, in the following we have taken a fixed, *geometric* cooling schedule:

$$T_{n+1} = \alpha^n T_0, \qquad 0 < \alpha < 1 \ . \tag{2}$$

In this approach, the process is broken into three steps. There is an initialization step, which is flat annealing, at the coarsest scale, from an initial temperature t_{init} to some t_0 according to a particular cooling schedule. At this point, the second step, begins: The hierarchical stepping of the system is begun. This process is controlled by two parameters θ_0, the number of iterations per step, and θ_1, the 'iteration offset' upon projection. For each scale, until the final, high-resolution scale is reached, the system is iterated under the given cooling schedule for θ_0 steps. At this point, the system is projected to the next highest scale, but the temperature is adjusted by θ_1 *steps* (i.e. as if n has changed in 2). The process is repeated until the final scale is reached. Finally, the system may be iterated at the finest scale until it has converged. Algorithm 2 describes this process:

Algorithm 2 Parametrized Hierarchical Annealing

```
{Initialize system at coarsest scale}
{Anneal until temperature t_0 is reached}
for scale s from coarsest to finest do
  for θ_0 iterations do
    β ⇐ 1/t
    X_s ⇐ apply Gibbs sampler to X_s
    t ⇐ α^n T_0
    n ⇐ n + 1
  end for
  X_{s-1} ⇐ P_{s-1}(X_s)        {map to next scale}
  n ⇐ n + θ_1
end for
{Final iterations done at highest resolution scale}
```

2.3 Models/Energy Functions

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. We are considering binary dense/pore structures [12] 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

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

Similarly, two-point correlation:

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

Quantities of particular interest in the study of porous media are the above mentioned correlation functions, and the distribution of chordlengths [12]. 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 $N \times N$ $\mathcal{O}(s) = N/2^s$). When considering chordlength distributions, if we 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 at scale s:

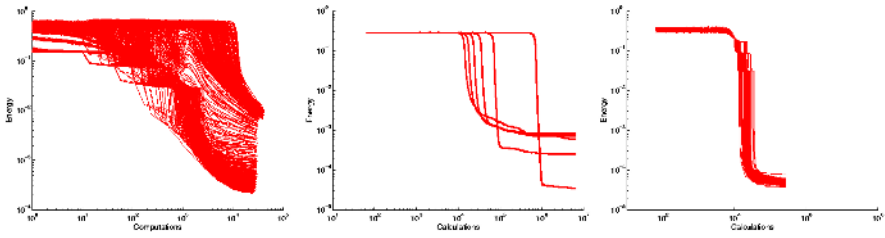
$$\mathcal{E}^s = \sum_{n=1}^{\mathcal{O}(s)} \|\hat{p}_C^h(n) - p_C^h(n)\| + \|\hat{p}_C^v(n) - p_C^v(n)\| \ . \tag{5}$$

Here the sample pmf's are estimated by histograms from the image data. Taken alone, this puts no constraint on the *amount* of pore or density in the image. Hence a constraint based on (3) must be added; either explicitly in the energy function, or implicitly by using a sampling algorithm that conserves density by exchanging pixels rather than flipping single sites. Other energy functions are of interest for this application [3], in particular two-point correlation (4) [12]. However, in the results following this section, we concentrate on chordlength distribution to construct energy functions like (5).

3 Results and Conclusions

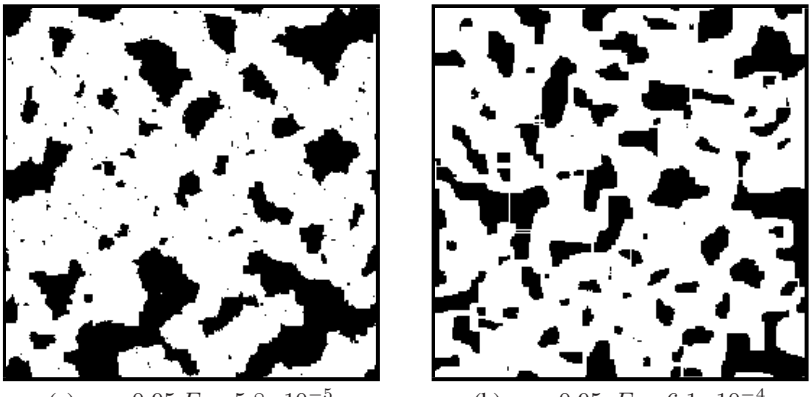
In order to demonstrate the efficacy of the ideas introduced in this work, we present some empirical results for a particular model, comparing the behaviour of 'flat' and (parameterized) 'hierarchical' annealing methods. The energy function used here is based on (5) with l_2 distance between histograms for both horizontal and vertical chordlength distributions (i.e., mean squared error). The chordlength distribution is important in the study of porous media, and has been used in annealing methods in the porous media literature [12, 11].

This model demonstrates how critical slowing down affects the flat annealing process. Figure 3 shows three plots: 3(a) gives energy vs. calculations results for a large number



(a) (b) (c)

Fig. 3. Flat and hierarchical sampling. Energy vs. Computations for (a) many hierarchical parameters, (b) flat annealing for several values of α, and (c) several 'good' hierarchical parameters for comparison to flat annealing results. Note similar low-temperatures are reached in (c) with much less computation than (b).



(a) $\alpha = 0.95$,$E = 5.8 \cdot 10^{-5}$ (b) $\alpha = 0.95$, $E = 6.1 \cdot 10^{-4}$

Fig. 4. Synthetic images from hierarchical (a), and flat (b) annealing. Energy E, and coefficient of geometric cooling (2), α, is given. Note the clear morphological differences between the right (flat) and left (hierarchical) images. Blockiness and line segments on the right are due to annealing too quickly.

of parameterizations of a chordlength model (without periodic boundaries) demonstrating a range of results; unsurprisingly a poor choice of parameterization leads to poor convergence. Figures 3(b) and 3(c) show for a similar model (with periodic boundaries) collection of good parameterizations (low final energy) to a number of flat annealing sessions with different cooling schedule parameters α,respectively.

Note that all energies are calculated at the finest scale, and results are shown on a log-log plot. Hierarchical annealing was done with the cooling parameter $\alpha = 0.95$. For flat annealing, α ranged from 0.9 to 0.999. Computations are normalized to the cost of a full 'sweep' of the Gibbs sampler on a 32×32 image. The longest flat annealing run takes approximately 2 days on a 3 Ghz. Pentium 4 class machine.

These plots show that hierarchical annealing may reach much lower temperatures for the same computational cost. Conversely, we may reach the same temperatures with far fewer computations. It should be noted that none of the above results are particularly good, morphologically speaking, when compared to the class of images they are trying

to reconstruct. Training images were taken from a set of glass sphere images (see Fig 1-a). However, Figure 4 shows that the lower energy images have significantly different characteristics from those resulting from 'flat' annealing. The problem of critical slowing down is illustrated in the right images (flat annealing). In particular, note that breaking a 'bad' long chord into two smaller chords need not be energetically favourable, hence such chord can be difficult to get rid of. Flat annealing images (right hand side) exhibit thin vertical and horizontal chords which are not seen in the (lower energy) hierarchical results.

In conclusion, we have discussed the need for hierarchical or multiscale methods in annealing, and related our proposed method to the literature. We have demonstrated the efficacy of a simple parameterized approach to hierarchical annealing when compared to standard, 'flat' annealing. Our method is shown to exhibit significant improvements to a practical application. It is, however, not in any way considered optimal and further analysis of the process is needed.

References

1. S. K. Alexander, P. Fieguth, and E. R. Vrscay, *Hierarchical annealing for random image synthesis*, EMMCVPR'03, LNCS, no. 2683, Springer, 2003.
2. _____, *Image sampling by hierarchical annealing*, ICIP'03, IEEE, 2003.
3. _____, *Hierarchical annealing for scientific models*, ICASSP'04, IEEE, 2004.
4. C. Bouman and M. Shapiro, *A multiscale random field model for Bayesian image segmentation*, IEEE Image Processing **3** (1994), no. 2, 162–177.
5. Pierre Brémaud, *Markov chains: Gibbs fields, monte carlo simulation, and queues*, Springer, 1998.
6. S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence **6** (1984), 721–741.
7. Basilis Gidas, *A renormalization group approach to image processing problems*, IEEE Transactions on Pattern Analysis and Machine Intelligence **11** (1989), no. 2, 164–180.
8. T. Hofmann, J. Puzicha, and J.M. Buhmann, *Unsupervised texture segmentation in a deterministic annealing framework*, Pattern Analysis and Machine Intelligence, IEEE Transactions on **20** (1998), no. 8, 803–818.
9. Zoltan Kato, Marc Berthod, and Josiane Zerubia, *A hierarchical Markov random field model and multitemperature annealing for parallel image classification*, Graphical Models and Image Processing **58** (1996), no. 1, 18–37.
10. Z. Liang, M.A. Ioannidis, and I Chatzis, *Geometric and topological analysis of three-dimensional porous media: Pore space partitioning based on morphological skeletonization*, Journal of Colloid and Interface Science **221** (2000), 13–24.
11. _____, *Reconstruction of 3d porous media using simulated annealing*, Computational Methods in Water Resources XIII (Balkema, Rotterdam) (Bentley et al., ed.), 2000.
12. M. S. Talukdar, O. Torsaeter, and M. A. Ioannidis, *Stochastic recontruction of particulate media from two-dimensional images*, Journal of Colloid and Interface Science **248** (2002), 419–428.