CONVERGENCE OF GIBBS MEASURES ASSOCIATED WITH SIMULATED ANNEALING: THE CASE OF DISTANCE SQUARED

We start with an explanation of Simulated Annealing. In order to do this, we quote the paper “Convergence of Gibbs Measures Associated with Simulated Annealing” [CHK], by professors Dennis Cox of the Rice University Statistics department, Robert Hardt of the Rice University Mathematics department, and Petr Klouček of the University of Houston Texas Learning and Computation Center, which states that “Simulated annealing is a stochastic optimization algorithm that mimics the physical process of a thermodynamic system settling into a state of minimal energy while lowering the ‘temperature.’”

When a system is at a high temperature, the particles in the system are more spread out, and have a high amount of energy. In contrast, when the system is near absolute zero, the particles concentrate on the system and move very slowly. Annealing explains this process.

Simulated annealing is given by a Gibbs measure, which we now describe:

Let

\[ P_\lambda(B) = \frac{\int_B e^{-\lambda J(x)} dx}{\int_{\mathbb{R}^n} e^{-\lambda J(x)} dx}, \]

where \( J(x) \geq 0 \) is a continuous function such that the denominator of \( P_\lambda(B) \) is finite, and \( B \) is a Borel set. Therefore, \( P_\lambda(B) \) is a probability measure, called a Gibbs measure. We see that as \( \lambda \) goes to infinity, this measure concentrates more and more on the zero set of \( J(x) \). The thermodynamic interpretation of \( \lambda \) (to tie \( P_\lambda(B) \) in with simulated annealing) is that \( \lambda \) represents \( (kT)^{-1} \), where \( T \) is temperature, and \( k \) is Boltzmann’s constant, which is approximately \( 1.3806505 \times 10^{-23} \) joules/kelvin. Boltzmann’s constant gives the energy of a particle in the system given a temperature (see [W]). So we see that at absolute zero, a particle has zero energy.

We would like to know what happens to \( P_\lambda(B) \) when \( \lambda \to \infty \) (so when \( T \to 0 \)) in terms of weak convergence, which we define now:

**Definition 0.1.** A sequence of probability measures \( \{ P_n \} \) converges weakly to a probability measure \( P \), denoted \( P_\lambda \to P \), if \( \ldots \)
\[
\int \phi dP_n \to \int \phi dP
\]
for all bounded continuous real-valued functions \(\phi\).

With this definition in hand, we state a theorem from the paper “Convergence of Gibbs Measures Associated with Simulated Annealing” (the proof of which can be found in [CHK]), which states the following:

**Theorem 0.2.** Assume

1. \(J \in C^3(\mathbb{R}^n)\).
2. \(J \geq 0\).
3. \(J(x) \geq \|x\|^p\) for \(x\) sufficiently large, \(p > 0\).
4. \(M = \{x \in \mathbb{R}^n : J(x) = 0\}\) is nonempty and bounded.
5. The Hessian \(D^2J\) has constant rank \(k\) “near” \(M\) (on sets covering \(M\)).

Then \(M\) is a \(C^{2,\alpha}\) \(n-k\) dimensional manifold, and as \(\lambda \to \infty\), \(P_\lambda \to P\), where

\[
P(B) = \frac{\int_{B \cap M} \Lambda(y)^{-1/2} dH^{n-k}y}{\int_M \Lambda(y)^{-1/2} dH^{n-k}y},
\]

where \(H^{n-k}\) is \(n-k\) dimensional Hausdorff measure, and \(\Lambda(y)\) is the product of the non-zero eigenvalues of \(D^2J\).

In the above theorem, the set \(M = \{x : J(x) = 0\}\) is where our probability measure concentrates. It is our constraint for simulated annealing. But many times we have a constraint in mind, and we would like to choose a function \(J\) based on a given \(M\). With that in mind, we study in [S] the distance squared function; that is, \(J(x) = \text{dist}^2(x, M)\). This function is nice enough that the theorem above applies to it, for nice enough sets \(M\). However, now we only have constant eigenvalues of \(D^2J\), as the reader can check. Therefore, we can relax our restriction on \(M\) having to be a \(C^{2,\alpha}\) manifold and obtain some nice, and sometimes surprising results.

The following examples from [S] are presented without proof.

**Example 0.3.** For \(M\) = the boundary of a unit square in \(\mathbb{R}^2\), and with \(J(x) = \text{dist}^2(x, M)\), we get

\[
P(B) = \frac{\mathcal{H}^1 \bot M}{4} = \frac{\mathcal{H}^1 \bot M}{\mathcal{H}^1(M)},
\]

where \(\mathcal{H}^1 \bot M\) is Hausdorff measure restricted to the set \(M\). Therefore, \(P(B)\) is a probability distribution for the boundary of the square, and we do not get any point masses.

**Example 0.4.** Fractals:
For $M = \text{the Koch curve in } \mathbb{R}^2$ (see [M], pages 65-67), and with $J(x) = \text{dist}^2(x, M)$, we get

$$P(B) = \frac{\mathcal{H}^{\log 4/\log 3} M}{\mathcal{H}^{\log 4/\log 3}(M)}.$$ 

That is, $P(B)$ is a probability distribution, where $\log 4/\log 3$ is the Hausdorff dimension of the Koch curve. Therefore, $P(B)$ is the probability measure on $M$. 

Similarly, for $M = C_4 \times C_4$ in $\mathbb{R}^2$, that is, the one-fourth Cantor Set crossed with itself (see [Mo], pages 32-33), and with $J(x) = \text{dist}^2(x, M)$, we get

$$P(B) = \frac{\mathcal{H}^1 M}{\mathcal{H}^1(M)},$$

where 1 is the the Hausdorff dimension of $C_4 \times C_4$.

Using these two results, we generalize in [S] the above results to all fractals $F$ with Hutchinson’s open set condition (see [H]). That is,

$$P(B) = \frac{\mathcal{H}^{\text{dim}(F)} M}{\mathcal{H}^{\text{dim}(F)}(F)}.$$

**Example 0.5. Targets:**

Let $M = \text{concentric circles of radii } \frac{1}{n}$, for $n$ a natural number. Let $J_\delta(x) = \chi_{M_\delta}$. So $J_\delta(x)$ is the indicator function of the delta neighborhood around $M$. We can prove that $J_\delta(x)$ converges weakly to $e^{-\lambda \text{dist}^2(x, M)}$ as $\delta \to 0$ and $\lambda \to \infty$. Using $J_\delta(x)$, we prove in [S] that $P(x)$ is a point mass at the center of the circles. So the center has probability 1, while the rest of the circles has measure zero.

Now, if we let $M = \text{two collections of concentric circles located some distance away from each other, one collection containing circles of radii } \frac{1}{n}$ called $T_1$, and the other collection containing circles of radii $\frac{1}{2n}$ called $T_2$, for $n$ a natural number, and using $J_\delta(x)$, we find that $P(x)$ is a point mass measure, giving the center of $T_1$ probability $\frac{2}{3}$ and the center of $T_2$ probability $\frac{1}{3}$.

Finally, if we let $M = \text{two collections of concentric circles located some distance away from each other, one collection containing circles of radii } \frac{2}{n}$, called $T$, and the other collection containing circles of radii $1 + \frac{1}{n}$, called $C$ (for circle, since the circles converge to a center circle, rather than a point), for $n$ a natural number, and using $J_\delta(x)$, we find that $P(x)$ gives the center point of $T$ probability $\frac{1}{2}$, while the other half of the measure is distributed among the center circle of $C$. Thus we have obtained a probability measure split equally between a point mass and the distribution of a circle.

Using $J(x) = \text{dist}^2(x, M)$ or an equivalent function (as in the last example) I hope to find more interesting examples of probability measures $P(x)$ for more interesting zero sets $M$. In particular, I would like to prove a result for algebraic varieties in real Euclidean space.
References


