A quantum diffusion network

George Kesidis
CSE and EE Depts
Pennsylvania State University
kesidis@engr.psu.edu

Technical Report No. CSE 09-012, Aug. 10, 2009
August 12, 2009

Abstract

Wong’s diffusion network is a stochastic, zero-input Hopfield network [8] with a Gibbs stationary distribution over a bounded, connected continuum [13]. Previously, logarithmic thermal annealing was demonstrated for the diffusion network [10] [7] and digital versions of it were studied and applied to imaging [14]. Recently, “quantum” annealed Markov chains have garnered significant attention [12] [3] because of their improved performance over “pure” thermal annealing. In this note, a joint quantum and thermal version of Wong’s diffusion network is described and its convergence properties are studied. Different choices for “auxiliary” functions are discussed, including those of the kinetic type previously associated with quantum annealing.

1 Introduction

The optimization of a function $V(x)$, $x \in D$, when the dimension of the space $D$ is large and multiple local minima exist, is a computationally difficult problem. A class of stochastic algorithms, known as simulated annealing, has been developed for the case where $D$ is countable [7]. For optimization over a bounded continuum, a diffusion network was proposed in [13] and its thermal annealing properties were established in [10] after [7]. In this note, we study a quantum version of this system that, unlike thermal annealing, modify the objective function $V$ in a nonlinear, nonuniform way. Quantum annealing proposals in the past include those involving the Schrödinger operator with potential $V$ [11], and those that add an auxiliary function to $V$ that depends on $\nabla V$ (e.g., the Ising spin glass model with an external field [3]). We consider here the latter type. Generally, the intuition behind the use of an auxiliary function is to initially perform a greater breadth of search than under pure thermal annealing search.
2 Specification of a Quantum Diffusion Machine

Consider a time-inhomogeneous system described by
\[ \begin{align*}
du(t) &= -\nabla [V(x(t)) - \Gamma(t)\tilde{V}(x(t))] \, dt + \Sigma(T(t), x(t)) \, dW(t) \\
x(t) &= G(u(t))
\end{align*} \tag{1} \]

where the first equation is a stochastic differential equation of the Itô type,
\[ \nabla \text{ is gradient with respect to the } x \text{ variables,} \]
\[ u(t) \in \mathbb{R}^n \text{ where } u(t) = (u^1(t), \ldots, u^n(t)), \ n \geq 1, \]
\[ x(t) \in (-1, 1)^n, \ x(t) = (x^1(t), \ldots, x^n(t)), \]
\[ W(t) \text{ is } n\text{-dimensional Brownian motion,} \]
\[ G : \mathbb{R}^n \to (-1, 1)^n, \]
\[ V, \tilde{V} : [-1, 1]^n \to \mathbb{R}, \ V, \tilde{V} \in C^2, \]
\[ M := \sup_{x \in [-1, 1]^n} V(x) - \inf_{x \in [-1, 1]^n} V(x) < \infty, \]
\[ \tilde{M} := \sup_{x \in [-1, 1]^n} \tilde{V}(x) - \inf_{x \in [-1, 1]^n} \tilde{V}(x) < \infty, \]
\[ T \geq 0 \text{ the deterministic thermal/temperature process, and} \]
\[ \Gamma \geq 0 \text{ the deterministic quantum parameter process.} \]

\( G \) is such that \( x^k_t = g(u^k_t) \) where \( g \) is a sigmoid threshold function commonly found in neural networks:
\[ g(u^k_t) = \tanh(u^k_t/w) \text{ with } w > 0. \]

If \( \tilde{V} \equiv 0 \) and \( \Sigma \equiv 0 \) then the two relations in (1) describe a continuous-time Hopfield network with Lyapunov function \( V \) and no external inputs (easily realized as a “neural” network when \( V \) is quadratic). If \( \tilde{V} \equiv 0 \) and
\[ \Sigma(T(t), x_t) = \text{diag} \left( \sqrt{\frac{2T(t)}{f(x^1(t))}}, \ldots, \sqrt{\frac{2T(t)}{f(x^n(t))}} \right) \]
where
\[ f(y) = g'(g^{-1}(y)) = \frac{1}{w}(1 - y^2) \]
and \( T > 0 \) is constant, then the stationary distribution of the \( x \) process is Gibbs
\[ \mu(x) := \frac{1}{Z} \exp(-V(x)/T) \tag{2} \]
where \( Z \) is the partition (normalization) function. This is immediately seen by applying Itô’s rule to (1), after which the Fokker-Planck operator (3) governing the distribution \( p \) of the \( x \) process is seen to be:
\[ L_\Gamma(p) = \text{div}[A(T\nabla p + p\nabla(V - \Gamma\tilde{V}))] \tag{3} \]
where
\[ A(x) = \text{diag}(f(x^1), \ldots, f(x^n)). \]
That is, $L_0(\mu) \equiv 0$. Furthermore, if $T(t) = T(0)/\log_2(2 + t)$ (logarithmic thermal cooling), $T(0) > 2M$, and the global extrema of $V$ are assumed in the interior $(0,1)^n$, then time-inhomogeneous process $x_t$ converges in probability to the (ground state) set that globally minimizes the objective function $V$ [7].

If fixed $T, \Gamma > 0$, then the invariant distribution is clearly

$$\mu_\Gamma(x) := \frac{1}{Z_\Gamma} \exp(-V(x) - \Gamma \hat{V}(x))/T).$$

(4)

So, if $\Gamma = o(T)$, $\mu_\Gamma$ is like a Gibbs distribution in the sense that it tends to indicate the globally minimizing (ground) states of $V$ as $T \to \infty$.

3 Quantum convergence to the Gibbs invariant

In [12] (and as explained in the recent survey [3]), a quantum annealing process is considered. They show that a faster-than-logarithmic quantum cooling schedule, $\Gamma(t) \downarrow 0$ as $t \to \infty$, can be used to establish convergence to the Gibbs invariant for fixed $T > 0$, i.e., not to the ground states. We now prove the analogous result for the diffusion network, subject to a more rapid cooling schedule, by adapting the thermal convergence proof in [7, 10]. To this end, we show how the distribution $m_t$ of $x_t$ “tracks” the distribution $\mu_{\Gamma(t)}$ (note that this is obvious for all sufficiently large $t$ if $\Gamma$ reaches zero in finite time). As the proof is a more substantive variation of [7] than for pure thermal annealing of the diffusion network, we give it in greater detail here than we did in [10].

We begin by defining

$$z_t := \int_{(0,1)^n} \frac{m_t^2(x)}{\mu_{\Gamma(t)}(x)} \, dx,$$

where

$$\dot{m}_t = L_{\Gamma(t)} m_t.$$

Let $\gamma_{\Gamma}$ be the gap between 0 and the rest of the spectrum of $L_{\Gamma}$ [2]:

$$\gamma_{\Gamma} = \inf_{\phi \neq 0} \frac{2T \int (\nabla \phi)^T A(\nabla \phi) \mu_{\Gamma(t)} \, dx}{\int \int (\phi(x) - \phi(y))^2 \mu_{\Gamma(t)}(x) \mu_{\Gamma(t)}(y) \, dy \, dx}$$

(5)

subject to the constraint that $\phi$ is not constant, where integration is over $(0,1)^n$.

Equivalently,

$$\gamma(\Gamma) = \inf_{\phi \neq 0} T \int (\nabla \phi)^T A(\nabla \phi) \mu_{\Gamma(t)} \, dx$$

subject to $\int \phi^2 \mu_{\Gamma} \, dx = 1$ and $\int \phi \mu_{\Gamma} \, dx = 0$.

**Theorem 3.1** For any nonincreasing, differentiable quantum schedule $\Gamma$ with $\Gamma(\infty) = 0$ and any constant temperature $T > 0$:

$$\dot{z}_t = O \left( \left( 1 + \frac{\dot{M}}{2T} \cdot \frac{\dot{\Gamma}(t)}{\gamma(\Gamma(t))} \right)^{-1} \right).$$

3
Proof: Take

\[ \phi_t = m_t / \mu_{\Gamma(t)}. \]

By direct differentiation,

\[
\dot{z}_t = 2 \int \phi_t \dot{m}_t \, dx - \int \phi_t^2 \dot{\mu}_t \, dx
= 2 \int \phi_t L_t \dot{m}_t \, dx - T^{-1} \hat{\Gamma}(t) \int (\vec{V} - \langle \vec{V} \rangle) \phi_t^2 \dot{\mu}_t \, dx
\leq 2 \int \phi_t L_t (\phi_t \mu_t) \, dx - T^{-1} \hat{\Gamma}(t) \hat{M} z_t
\]

where the last step is integration by parts using \( A(0) = 0 = A(1) \). Thus, by the previous expression for \( \dot{\gamma}(\Gamma(t)) \) (noting \( \int (\phi_t - 1) \mu_t \, dx = 0 \)),

\[
\dot{z}_t \leq \gamma(\Gamma(t)) \int (\phi_t - 1)^2 \mu_t \, dx - T^{-1} \hat{\Gamma}(t) \hat{M} z_t,
\]

Integrating in time, we get an inequality of the form \( z_t \leq \alpha_t + \int_0^t \beta_s z_s \, ds \) where \( \alpha_t := z_0 + \int_0^t 2 \gamma(\Gamma(s)) \, ds \) and

\[ \beta_t := -2 \gamma(\Gamma(t)) - T^{-1} \hat{\Gamma}(t) \hat{M}. \]

So by applying Gronwall’s lemma and then multiplying by \( 1 = \exp(-\int_0^t \beta_r \, dr) / \exp(-\int_0^t \beta_r \, dr) \), we get

\[
z_t \leq \alpha_t + \int_0^t \alpha_s \beta_s \exp\left(\int_s^t \beta_r \, dr\right) \, ds
= \frac{\alpha_t \exp(-\int_0^t \beta_s \, ds) + \int_0^t \alpha_s \, d \exp(-\int_s^t \beta_r \, dr)}{\exp(-\int_0^t \beta_r \, dr)}
= \frac{z_0 + \int_0^t 2 \gamma(\Gamma(s)) \exp(-\int_s^t \beta_r \, dr) \, ds}{\exp(-\int_0^t \beta_r \, dr)}
\]

where the last step is integration by parts (resulting in term cancellation in the numerator) and the fact that \( \alpha_0 = z_0 \) and \( \dot{\alpha}_t = 2 \gamma_t \). Now note that as \( t \to \infty \),

\[ \gamma(\Gamma(t)) \to \gamma(\Gamma(\infty)) := \gamma(0) > 0 \] and \( \hat{\Gamma}(t) \to 0 \), and therefore \( \beta_t \to -2 \gamma(0) < 0 \).

Thus, the numerator and denominator of the previous display both diverge as \( t \to \infty \). Applying L'Hôpital’s rule gives that

\[
\dot{z}_t \leq \frac{2 \gamma(\Gamma(t))}{-\beta_t} \text{ as } t \to \infty.
\]
Lemma 3.1 \exists c > 0, which does not depend on T or \( \Gamma \), such that
\[
\gamma(\Gamma) \geq cT \exp(-2M^*(\Gamma)/T)
\]
where
\[
M^*(\Gamma) := \sup(V - \Gamma \tilde{V}) - \inf(V - \Gamma \tilde{V}). \tag{6}
\]

**Proof:** By (5), \( \gamma(\Gamma) \geq cT \inf \mu_{\Gamma}/(\sup \mu_{\Gamma})^2 \) where
\[
c := \inf_{\phi} \frac{2 \int (\nabla \phi)^T A(\nabla \phi) \, dx}{\int \int (\phi(x) - \phi(y))^2 \, dy \, dx}. \]
So, \( \gamma(\Gamma) \geq cT Z_{\Gamma} \exp(-\sup(V - \Gamma \tilde{V})/T)/\exp(-2\inf(V - \Gamma \tilde{V})/T) = cT \exp(-M^*(\Gamma)/T) \int \exp(-[(V - \Gamma \tilde{V}) - \inf(V - \Gamma \tilde{V})]/T) \, dx. \)

\[\square\]

Completing our adaptation of the arguments in [7, 10]:

**Corollary 3.1** For any nonincreasing, differentiable quantum schedule \( \Gamma \) with \( \Gamma(\infty) = 0 \) and any constant temperature \( T > 0 \), there is a constant \( K < \infty \) such that for any \( S \subset (0,1)^n \),
\[
P(x_t \in S) \leq K \left( \int_S \mu_{\Gamma(t)}(x) \, dx \right)^{1/2} \forall t \geq 0.
\]

**Proof:** Let
\[
B(\Gamma, \dot{\Gamma}, t) := \left( 1 + \frac{M}{2cT^2} \dot{\Gamma}(t) \exp(2M^*(\Gamma(t))/T) \right)^{-1}. \tag{7}
\]
By the previous lemma and theorem,
\[
\lim_{t \to \infty} z_t \leq \lim_{t \to \infty} B(\Gamma, \dot{\Gamma}, t) = 1. \tag{8}
\]
Thus, by the continuity of \( z_t \), there exists a positive constant \( K < \infty \) such that \( z_t \leq K^2 \) for all \( t \geq 0 \). So, by the Cauchy-Schwarz inequality,
\[
P(x_t \in S) = \int 1_S m_t \, dx = \int 1_S \phi_t \mu_{\Gamma(t)} \, dx \leq (\int \phi_t^2 \mu_{\Gamma(t)} \, dx)^{1/2} (\int 1_S^2 \mu_{\Gamma(t)} \, dx)^{1/2} \leq (\int z_t m_t \, dx)^{1/2} (\int \mu_{\Gamma(t)} \, dx)^{1/2}.
\]
Substituting $z_t \leq K^2$ completes the proof. □

Note that $K$ will depend on the parameter $z_0$.

4 Global optimization of joint annealing

To interpret this result, note that as $t \to \infty$, $\mu_t$ defined in (4) is tending to the Gibbs distribution $\mathbb{I}$ for fixed $T > 0$. Therefore, if $T > 0$ is small and $S$ does not include the ground states of $V$ (e.g., $S = \{x \in (0, 1)^n \mid V(x) \geq \theta + \inf V\}$ for some sufficiently large $\theta > 0$), then $\mathbb{P}(x_t \in S)$ will be small. To sharpen this statement, consider joint quantum and thermal annealing.

**Theorem 4.1** If $\Lambda(t) := \Gamma(t)/T(t) \to 0$, i.e., $\Gamma = o(T)$, and $D(t) := 1/T(t) = \log_2(2 + t)/T(0)$ with $T(0) > 2M$, then $\exists K^* < \infty$ such that

$$\mathbb{P}(x_t \in S) \leq K^* \left( \int_S \mu_t(x) \, dx \right)^{1/2} \quad \forall t \geq 0,$$

where $\mu_t$ is given by (2) with $T = T(t)$.

**Proof:** Argue as for (8) that

$$z_t = O \left( \left( 1 + \frac{\tilde{M}\tilde{\Lambda}(t) - M\tilde{D}(t)}{2\gamma_t} \right)^{-1} \right)$$

and

$$\gamma_t \geq \frac{c}{D(t)} \exp(-2D(t)M^*(\Gamma(t))),$$

and so conclude as in the previous corollary, where the condition $T(0) > 2M = 2M^*(0)$ figures in the resulting exponent of $(1 + t)$ after substituting for $D$. □

So, if $S$ does not contain any of the ground states of $V$, then $\lim_{t \to \infty} \mathbb{P}(x_t \in S) \to 0$.

5 Discussion: Choices for auxiliary function

5.1 Homotopy methods

In “homotopy” based search [4], the auxiliary function is taken to be $\tilde{V} := V - V_0$

where $\Gamma(0) \approx 1$ and $V_0$ is unimodal (only one local minimum which is, of course, its global minimum). Therefore, the ground states of $V - \Gamma(t)\tilde{V}$ are quickly found initially (i.e., when $t > 0$ is small so that $V - \Gamma(t)\tilde{V} \approx V_0$). Ideally, $V_0$ is the best such function approximating $V$ if suitable “global” information about $V$ is available to determine it a priori; in this case, the initial ground states (of $V - \Gamma(t)\tilde{V}$ for small $t > 0$) are close to those of objective function $V$. 

6
5.2 Contracting the objective function

Suppose that the auxiliary function is simply

$$\tilde{V} := V$$

and that $\Gamma(0) < 1$. In this case, the quantum diffusion network is performing a kind of thermal annealing from temperature $T/(1 - \Gamma(0))$ down to $T > 0$. So, this choice of auxiliary function has the effect of linearly contracting the objective function $V$, as in “pure” thermal annealing, thereby facilitating a greater breadth of search initially.

An example nonlinear contraction of the objective function $V$ is obtained by using the auxiliary function

$$\tilde{V}(x) := -\varepsilon^T \nabla^2 V(x) \varepsilon, \ x \in (-1, 1)^n,$$

where $\varepsilon$ is a fixed $n$-vector. Note that $\tilde{V}(x) > 0$, respectively $\tilde{V}(x) < 0$, when $x$ is a local maximum, respectively minimum, of $V$.

The use of the auxiliary (9) may not result in significant contraction of the objective function (i.e., from $V$ to $V - \Gamma \tilde{V}$) in situations where the peaks or valleys of $V$ are very deep. In a one dimensional ($n = 1$) setting, we can deal with this problem in the case where there is a local extremum ($V' = 0$) between successive points at which $V'' = 0$ (no saddle points in particular) by augmenting this auxiliary using “kinetic” components (i.e., involving $V'$) which are typically associated with “quantum” annealing, e.g.,

$$\tilde{V} := -(\varepsilon^2 + |V'|^2)V''.$$  

This example has a natural multidimensional form: $\tilde{V} := -\varepsilon^T \nabla^2 V \varepsilon - (\nabla V)^T \nabla^2 V \nabla V$.

In the case where it is advantageous to further contract the objective $V$ at the points where $V'' = 0$ ($V$ and $V - \tilde{V}$ for “quantum” auxiliary $\tilde{V}$ of (10) are equal at these points), one can similarly propose to augment the auxiliary function with $-(\varepsilon^2 + |V''|^2)V'', \ etc.$

References

[1] B. Appoloni, C. Carvalho, and D. deFalco. “Quantum stochastic optimization.” Stochastic Processes and their Applications, Vol. 33, No. 2: pp. 233-244, 1989.

[2] H. Brand, A. Schenzle, and G. Schroder. “Lower and upper bounds for the eigenvalues of the Fokker-Planck operator in detailed balance.” Physical Review A, Vol. 25, No. 4: pp. 2324-2338, April 1982.

[3] A. Das and B.K. Chakrabarti. “Quantum annealing and analog quantum computation.” Reviews in Modern Physics, in press.
[4] D.M. Dunlavy and Dianne P. O'Leary. “Homotopy Optimization Methods for Global Optimization.” Sandia National Laboratories Technical Report, SAND2005-7495, Dec. 2005

[5] S. Geman and C. Hwang. “Diffusions for global optimization.” SIAM J. Control and Optimization, Vol. 24, No. 5: pp. 1031-1043, Sept. 1986.

[6] B. Gidas. “Global optimization via the Langevin equation.” In Proc. of the 24th Conf. on Decision and Control, Ft. Lauderdale, FL, pp. 774-778, Dec 1985.

[7] R. Holley and D. Stroock. “Simulated Annealing via Sobolev inequalities.” Commun. in Math. Phys., No. 115: pp. 553-569, 1988.

[8] J.J. Hopfield. “Neural networks and physical systems with emergent computational abilities.” In Proc. Natl Acad. of Sci. USA, Vol. 79: pp. 2554-2558, April 1982.

[9] I. Karatzas and S. Shreve. Brownian motion and stochastic calculus. Second Edition. SpringerVerlag, New York, 1991.

[10] G. Kesidis. “Analog optimization with Wong’s stochastic neural network.” IEEE Trans. Neural Networks 6(1), pp. 258-260, Jan. 1995.

[11] G. Kesidis and E. Wong. Optimal Acceptance Probability for Simulated Annealing. Stochastics and Stochastics Reports, Vol. 29, pp. 221-226, 1990.

[12] S. Morita and H. Nishimori. “Convergence theorems for quantum annealing.” J. Phys. A: Math. Gen., vol. 39, pp. 13903-13920, 2006.

[13] E. Wong. “Stochastic Neural Networks.” Algorithmica, Vol. 6, pp. 466-478, 1991.

[14] G. Yin and P.A. Kelly. “Convergence Rates of Digital Diffusion Network Algorithms for Global Optimization with Applications to Image Estimation.” Journal of Global Optimization, 23(3-4), Aug. 2002.