

確率測度の生成と最適化問題

(株) 東芝 総合研究所 情報システム研究所
上林 達

スピン系の連続時間 Markov 過程をシミュレートすることによって、有限集合上に任意の定常確率測度を高速に生成する確率的アルゴリズムを提案する。それは、本質的に並列計算である。このアルゴリズムは数理計画法における大域的最適解を効率的に探索する方法の基礎を与える。確率測度の生成に基礎をおく最適解探索法の特徴と有効性を示すため、計算機実験を行なった。

An Application of Probability Distribution to Optimization Problems

Tohru Kambayashi

Information Systems Laboratory, R & D center
Toshiba

We offer an Monte Carlo algorithm that generates rapidly an arbitrary probability distribution on a finite set of numerous elements. The algorithm is a simulation of a Markov process with continuous time parameter on a spin model, and is essentially parallel computation. It will be applied to mathematical programming problems, especially to nonlinear or integer programming problems. In order to show the property and the efficiency of our method, we made some experiments.

1 はじめに

まず、次のような簡単な問題を考えてみよう:

Q. 1 $0 \leq x \leq 1$ において, $f(x) = x \sin(16\pi x)$ を最小化せよ. |

我々のアイデアは、最適解が高い確率で出現するように確率測度を構成するということである。例えば、“区間 $[0, 1]$ 上に値をとる確率変数でその分布が $e^{-\beta f(x)}$ に比例するものを構成する”ことができれば、我々は $f(x)$ に関する大域的最適解(ないしは、それに近い解)を得ることができるであろう。ただし、 β は正定数。確率変数の値を多数観測すれば、 $f(x)$ の最適解の近傍を捉えることができる。それを初期条件にして、降下法等を適用することもできよう。或いは、区間を縮小して再び確率分布を構成するということを繰り返して、最適解を“追いつめる”こともできる。このように都合の良い確率分布を持つ確率変数を計算機上でつくり出すことは常に可能であろうか。答えは“yes”である。その確率変数を構成するのにどの位の時間がかかるだろうか。答えは“一般に極めて高速に”である。 実際

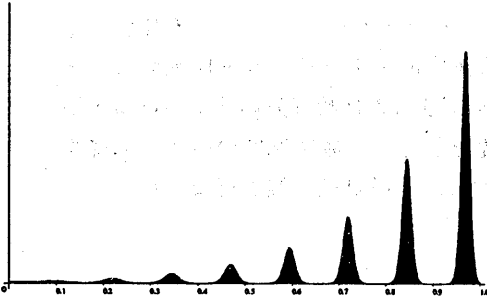


図 1: $0 \leq x \leq 1$ における $e^{-\beta f(x)}$ に比例する確率分布

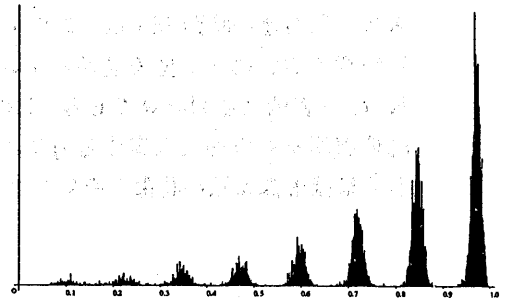


図 2: PMC アルゴリズムを用いて得られた分布。 $[0, 1]$ 区間を 2^{32} 等分することによって離散化している。縦軸は出現の累積度数。

には、多数の分点で f を離散化することになるのであるが、極めて大きな有限集合上に任意の確率測度を速やかに構成することを可能にするのが我々の確率アルゴリズムである。これを今後 PMC アルゴリズム (Probability measure construction algorithm) と呼ぶことにする。また PMC アルゴリズムに基礎をおく数値計画法を PMC 法と呼ぶ。PMC アルゴリズムは非常に一般的な性格を持っている為 PMC 法には様々なヴァリエーションが考えられる。例えば、 f の代わりに f と単調増加関数 H との合成 $H \circ f$ を用いて良い結果が期待できる場合もある。

上に挙げた関数 $f(x)$ に関しては、初期条件を random にとって降下法を適用することによって、大域的最適化が困難に行なえることは明らかである。しかし、例えば次のような問題の場合はどうであろうか。

Q. 2 $(x, y, z, w) \in [-1, 1]^4$ において

$$f(x, y, z, w) = \frac{1}{\sqrt{2}}(5x^2 + 5y^2 + 19z^2 + 19w^2 - 6xy - 2zw)^{1/2} \times \{1 + \frac{1}{4}(\sin^2(\pi ax) + \sin^2(\pi ay) + \sin^2(\pi az) + \sin^2(\pi aw))\}$$

を最適化せよ。ただし、 $a = 500$. |

この関数は原点において最小値 0 をとる。このような問題に対しても、PMC 法によれば短い時間で満足すべき解を見出すことができる。(Section 3参照)

PMC 法は、2 値のスピン系で定義された 2 次関数の最適化に関して S. Kirkpatrick らが提案した simulated annealing という Monte Carlo 法に示唆を受けている。(文献 [5]) Simulated annealing は、液相から単結晶を成長させる為の徐冷法にその着想を得ている。Simulated annealing のアイデアは面白い、大域的な最適解を捜し出すこともできるが、それにはとてつもない時間がかかる為 annealing の過程を正直に実行するのは現実的ではない。スピんに逐次処理を行なうことも、計算時間の点から効率的ではない。Simulated annealing 法と PMC 法の比較という見地では、後者は主として次の 2 点で前者と異なっている:

- 最適解そのものに到達するためではなく、最適解にピークを有する確率測度 を生成する様にスピン系の確率過程が構成される。
- 系の遷移に 連続な時刻 を導入することで、理論的に妥当な並列化を可能にした。

Section 2.1 において、PMC アルゴリズムの詳細を述べる。ついで、section 2.2 においてアルゴリズムの妥当性に数学的検討を加える。PMC 法に関する計算機実験とその結果は、section 3 に見ることができる。

2 アルゴリズムとその解析

この section では、PMC アルゴリズムとその数学的背景を述べる。

N を自然数とする。2 値 (簡単な為、 ± 1 とする) をとる N 個のスピンからなる系 Ξ と、 Ξ 上の実数値関数 V とが与えられているものとする:

$$\Xi = \prod_{j=1}^N \{-1, +1\}, \quad \Xi \ni \xi: \{1, 2, \dots, N\} \rightarrow \{-1, +1\}$$

$$V: \Xi \rightarrow \mathbf{R}$$

PMC アルゴリズムの目標は、 Ξ -値確率変数 X で、その確率分布が $\mu_\beta \stackrel{\text{def}}{=} (1/Z_\beta)e^{-\beta V} \mu_0$ に等しいものを、構成することである。ここに、 β は正定数、 μ_0 は Ξ 上の一様な確率測度。また

$$Z_\beta = \int_{\Xi} e^{-\beta V} d\mu_0 = \frac{1}{2^N} \sum_{\xi \in \Xi} e^{-\beta V(\xi)}$$

は正規化定数である。なお、今後 V をスピン上のポテンシャルと呼ぶことがある。

我々は、連続な時変数 t を parameter に持つ Ξ -値 確率変数 $\xi(t)$ を、スピン系の離散時間の Markov 過程によって近似的に構成する。それが $t \rightarrow \infty$ の極限で X に法則収束¹ すること、及び master equation において、 X の確率分布が指数安定であることを証明する。この証明においては、Holley & Stroock (文献 [4]) が simulated annealing の解析で用いた確率論の概念 — Dirichlet form — を利用する。さらに、我々は Dirichlet form に検討を加え、収束の速度を V と β によって定まる式で評価する。

2.1 アルゴリズム

Spin 系の状態が ξ であるとき、 μ 番目の spin の値が $\xi(\mu)$ から $-\xi(\mu)$ に遷移する単位時間当りの確率を $w_\mu(\xi; \beta)$ とする。ここに、 β は正定数。従って、spin μ が、微小時間 Δt の間に $\xi(\mu)$ から $-\xi(\mu)$ に遷移する確率は、 $\Delta t w_\mu(\xi; \beta)$ となる。各 spin は同時に遷移を行なうものとする。このような確率遷移は、Monte Carlo 法によって実現することが出来る。

$w_\mu(\xi; \beta)$ の定め方を述べよう。 $\xi \in \Xi$ に対し、 $\tau_\mu \xi \in \Xi$ を次のように定める:

$$(\tau_\mu \xi)(\nu) = \begin{cases} -\xi(\mu) & (\nu = \mu), \\ \xi(\nu) & (\nu \neq \mu). \end{cases}$$

系の発展は、次の遷移確率を持つ Markov 過程となる:

$$(1) \quad P_{t, t+\Delta t}(\xi, \eta) = \begin{cases} \prod_{\mu=1}^N \{1 - \Delta t w_\mu(\xi; \beta)\} & (\eta = \xi), \\ \Delta t w_j(\xi; \beta) \prod_{\mu \neq j} \{1 - \Delta t w_\mu(\xi; \beta)\} & (\eta = \tau_j \xi), \\ o(\Delta t) & (\eta \text{ が上記以外}). \end{cases}$$

但し、 $P_{s, t}(\xi, \eta)$ は時刻 s に系が状態 ξ にあったとき、時刻 $t (\geq s)$ に系の状態が η となる条件つき確率である。連続化 ($\Delta t \rightarrow 0$) することにより、我々は master equation

$$(2) \quad \frac{\partial}{\partial t} \rho_t(\eta) = - \sum_{j=1}^N \rho_t(\eta) w_j(\eta; \beta) + \sum_{j=1}^N \rho_t(\tau_j \eta) w_j(\tau_j \eta; \beta)$$

を得る。但し、 ρ_t は時刻 t における確率密度。

さて、記号を次のように定義する: $Z_\beta = 2^N Z_\beta$, $\rho_\beta = (1/Z_\beta)e^{-\beta V}$. 我々は、 $\rho_t = \rho_\beta$ に対して (2) の右辺が 0 になるように $w_j(\xi; \beta)$ を定める。即ち、 μ_β が定常分布になる様にする。 $w_j(\tau_j \xi; \beta) / w_j(\xi; \beta) = \rho_\beta(\xi) / \rho_\beta(\tau_j \xi)$ であればその条件は満たされるから、例えば

$$(3) \quad w_j(\xi; \beta) = e^{-\beta(V(\tau_j \xi) - V(\xi))^+} = e^{\beta(V(\xi) - V(\tau_j \xi))^-},$$

あるいは

$$w_j(\xi; \beta) = \{1 + \exp\{\beta(v(\tau_j \xi) - v(\xi))\}\}^{-1}$$

ととることができる。但し、 $x^+ = x(x \geq 0)$, $0(x < 0)$. $x^- = -(-x^+)$.

PMC アルゴリズムは、例えば SIMD-type の並列計算機上で高速に稼働するであろう。その際プロセッサ群に tree-structure を導入すると良いであろう。スピンの個数が p^n であるとする。spin j にはプロセッサ U_j とメモリ M_j が割り当てられており、spin の値はメモリに格納される。(ここでは、 $j = 0, \dots, p^n - 1$ であるとする。)

¹ 確率分布の distribution の意味での収束

$$G_m = \{U_j | j \bmod p^m = 0, j \bmod p^{m+1} \neq 0\} \quad (0 \leq m < n)$$

とおく。ただし、mod は剰余を表す。まず $m = 0$ として、 G_m に属するプロセッサ U_j はプロセッサ $U_{j-(j \bmod p^{m+1})}$ に spin の値を送る。次に m を 1 増やして、同じことを行う。 $m = n$ になれば、プロセッサ U_0 には計算に必要なデータが集まっているわけだから、それを、全てのプロセッサに broadcast する。 p の値は、通信の時間的コストと各プロセッサのデータ処理のそれとの関係から決定されよう。この方法によれば、データ転送の手間を減らすことができ、各プロセッサの同期の問題も解決される。

2.2 数学的解析

Master equation (2) は線型微分方程式であり、係数は定数である。ここでは、確率論を用いて (2) の解の ρ_β への収束性 — 及び収束の order — を考察する。今後 w_j の形として (3) を仮定する。また $V \neq \text{const.}$ とする。従って $\min_\zeta V(\zeta) = 0$, $\max_\zeta V(\zeta) = 1$ と仮定して一般性を失わない。実際、遷移確率は V の大きさの差に依存するし、 V の amplitude は β の大きさで置き換えて良い。

初期分布を m_0 とする: $m_0 = 2^N \rho_0 \mu_0$ ($\rho_0 \geq 0$, $\int_{\Xi} dm_0 = 1$). 時刻 t (≥ 0) における分布を $m_0 P_{0,t}$ と書く: $m_0 P_{0,t} = 2^N \rho_t \mu_0$. ただし

$$\rho_t(\xi) = \int_{\Xi} P_{0,t}(\eta, \xi) dm_0(\eta) = \sum_{\eta \in \Xi} \rho_0(\eta) P_{0,t}(\eta, \xi).$$

Radon-Nikodym derivative $d(m_0 P_{0,t})/d\mu_\beta$ を f_t とおく: $f_t = (Z_\beta/e^{-\beta V})\rho_t$. また、三上の実数値関数の全体に内積

$$(\phi, \psi) = \int_{\Xi} \phi \psi d\mu_\beta = \sum_{\xi \in \Xi} \phi(\xi) \psi(\xi) \cdot \frac{1}{Z_\beta} e^{-\beta V(\xi)}$$

をいれた空間を L_β^2 で表すことにする。我々は $\|f_t - 1\|_\beta^2 = (f_t - 1, f_t - 1)$ の $t \rightarrow \infty$ での振舞いを調べよう。実際、 $\|f_t - 1\|_\beta \rightarrow 0$ ($t \rightarrow \infty$) は $\rho_t \rightarrow \rho_\beta$ ($t \rightarrow \infty$) を意味するからである。更に、 $\|f_t - 1\|_\beta$ を $-t$ の exponential で上から評価する。

まず、三上の実数値関数 ϕ に対して三上の実数値関数 $\mathcal{L}_\beta \phi$ を次で定める:

$$(\mathcal{L}_\beta \phi)(\xi) = \sum_{j=1}^N w_j(\xi; \beta) (\phi(\tau_j \xi) - \phi(\xi)).$$

又

$$W_j(\xi; \beta) = w_j(\xi; \beta) \cdot \frac{1}{Z_\beta} e^{-\beta V(\xi)}$$

とおく。 w_j の形が (3) であるとき、 $W_j(\xi; \beta) = (1/Z_\beta) \exp\{-\beta(V(\xi) \vee V(\tau_j \xi))\}$ となる。ただし、 $x \vee y = (x - y)^+ + y$. $W_j(\tau_j \xi; \beta) = W_j(\xi; \beta)$ が成り立つことに注意する。

さて、三上の任意の実数値関数 ϕ, ψ に対して

$$E_\beta(\phi, \psi) = \frac{1}{2} \sum_{\xi \in \Xi} \sum_{j=1}^N (\phi(\tau_j \xi) - \phi(\xi)) (\psi(\tau_j \xi) - \psi(\xi)) W_j(\xi; \beta)$$

とおく。

Lemma 2.1 $\mathcal{L}_\beta : L_\beta^2 \rightarrow L_\beta^2$ は、自己共役作用素である:

$$\int_{\Xi} \psi(\mathcal{L}_\beta \phi) d\mu_\beta = \int_{\Xi} \phi(\mathcal{L}_\beta \psi) d\mu_\beta = -E_\beta(\phi, \psi).$$

(Proof)

$$\sum_{\xi} \sum_{j=1}^N \psi(\tau_j \xi) (\phi(\tau_j \xi) - \phi(\xi)) W_j(\xi; \beta) = - \sum_{\xi} \sum_{j=1}^N \psi(\xi) (\phi(\tau_j \xi) - \phi(\xi)) W_j(\tau_j \xi; \beta)$$

が成立する。後は、定義に従った簡単な計算により、主張が証せられる。□

$$(P_{s,t} \phi)(\xi) = \int_{\Xi} P_{s,t}(\xi, d\eta) \phi(\eta) = \sum_{\eta} P_{s,t}(\xi, \eta) \phi(\eta)$$

とおこう。(1)を用いることにより,

$$\frac{\partial}{\partial t}(P_{s,t}\phi)(\xi) = (P_{s,t}(\mathcal{L}_\beta\phi))(\xi)$$

を得る.

Lemma 2.2

$$\frac{\partial}{\partial t}f_t = \mathcal{L}_\beta f_t$$

(Proof) 任意の実数値関数 ϕ に対して,

$$\int_{\Xi} \phi f_t d\mu_\beta = \int_{\Xi} \phi d(m_0 P_{0,t}) = \int_{\Xi} (P_{0,t}\phi) dm_0$$

が成立する. 従って,

$$\frac{d}{dt} \int_{\Xi} \phi f_t d\mu_\beta = \int_{\Xi} (P_{0,t}(\mathcal{L}_\beta\phi)) dm_0 = \int_{\Xi} (\mathcal{L}_\beta\phi) f_t d\mu_\beta = \int_{\Xi} \phi (\mathcal{L}_\beta f_t) d\mu_\beta.$$

ϕ は任意であったから, 主張が従う. \square

$$c_\beta \stackrel{\text{def}}{=} \inf \{ E_\beta(\phi, \phi) \mid \phi \in L_\beta^2, \|\phi\|_\beta = 1, \int_{\Xi} \phi d\mu_\beta = 0 \} \text{ とする.}$$

$$c_\beta = \inf \left\{ \frac{E_\beta(\phi, \phi)}{\text{var}_\beta(\phi)} \mid \phi \in L_\beta^2 \right\}$$

が成り立つことに注意する. ただし, $\text{var}_\beta(\phi) = \int \phi^2 d\mu_\beta - (\int \phi d\mu_\beta)^2$.

Theorem 2.1 (Holley & Stroock [4])

任意の $t \geq 0$ に対して $\|f_t - 1\|_\beta \leq e^{-\frac{1}{2}c_\beta t} \|f_0 - 1\|_\beta$ が成立する.

(Proof)

$$\frac{d}{dt} \|f_t - 1\|_\beta^2 = -E_\beta(f_t - 1, f_t - 1) \leq -c_\beta \|f_t - 1\|_\beta^2.$$

一方, $\exists T \geq 0, \|f_T - 1\|_\beta = 0 \Rightarrow \forall t \geq T, \|f_t - 1\|_\beta = 0$ が成立する. よって主張は明らか. \square

収束の速度は, 定数 c_β に依存する. 従って, PMC アルゴリズムの実際的な有効性を考える上で, c_β のできるだけ sharp な評価を得ることが重要である.

$\xi, \eta \in \Xi (\xi \neq \eta)$ に対して, Ξ の点列 $\{\sigma_k\}_{k=0}^m$ で, 次の条件を満たすものを ξ, η を結ぶ path と呼び, m をその長さという:

$$\sigma_0 = \xi, \sigma_m = \eta \text{ and } \forall k \in \{1, \dots, m\}, \exists j \in \{1, \dots, N\}, \sigma_k = \tau_j \sigma_{k-1}.$$

さらに, ξ, η を結ぶ最短の path の長さを $[\xi : \eta]$ と書くことにする. $[\xi : \eta]$ は ξ と η の間の Hamming distance に等しい. ξ, η を結ぶ最短の path 全体の集合を $\sigma(\xi, \eta)$ で表す: $\sigma(\xi, \eta) = \{\sigma^{(p)}(\xi, \eta) \mid p = 1, 2, \dots, [\xi : \eta]!\}$. 各 p に対し, $\sigma^{(p)}(\xi, \eta)$ は, ξ, η を結ぶ最短の path の 1 つである: $\sigma^{(p)}(\xi, \eta) = \{\sigma_k^{(p)}(\xi, \eta)\}_{k=0}^{[\xi : \eta]}$. $\sigma_k^{(p)}(\xi, \eta)$ は, path $\sigma^{(p)}(\xi, \eta)$ の k 番目の点である.

さて, $\lambda(\xi, \eta) \stackrel{\text{def}}{=} \max_p \max_k (V(\sigma_k^{(p)}(\xi, \eta)) \vee V(\sigma_{k-1}^{(p)}(\xi, \eta)) - V(\xi) - V(\eta))$ とおき, さらに $\Lambda = \max_{\xi \neq \eta} \lambda(\xi, \eta)$ とおく. $\Lambda \leq 1$ は明らか. まず,

$$E_\beta(\phi, \phi) = \frac{1}{2Z_\beta} \sum_{[\xi : \eta]=1} (\phi(\eta) - \phi(\xi))^2 \exp\{-\beta(V(\xi) \vee V(\eta))\}$$

と書ける. また

$$\begin{aligned} (4) \quad 2 \text{var}_\beta(\phi) &= \sum_{\xi, \eta} (\phi(\eta) - \phi(\xi))^2 \cdot \frac{1}{Z_\beta} e^{-\beta V(\xi)} \cdot \frac{1}{Z_\beta} e^{-\beta V(\eta)} \\ &= \sum_{[\xi : \eta]=1} (\phi(\eta) - \phi(\xi))^2 \cdot \frac{1}{Z_\beta} e^{-\beta V(\xi)} \cdot \frac{1}{Z_\beta} e^{-\beta V(\eta)} \\ &\quad + \sum_{r=2}^N \sum_{[\xi : \eta]=r} (\phi(\eta) - \phi(\xi))^2 \cdot \frac{1}{Z_\beta} e^{-\beta V(\xi)} \cdot \frac{1}{Z_\beta} e^{-\beta V(\eta)}. \end{aligned}$$

ここで,

$$(5) \sum_{[\xi:\eta]=1} (\phi(\eta) - \phi(\xi))^2 \cdot \frac{1}{Z_\beta} e^{-\beta V(\xi)} \cdot \frac{1}{Z_\beta} e^{-\beta V(\eta)} \leq \frac{2e^{\beta\Lambda}}{Z_\beta} E_\beta(\phi, \phi)$$

は明らか。一方 $[\xi:\eta] = r$ ($r \geq 2$) とすると,

$$(\phi(\eta) - \phi(\xi))^2 \leq \frac{1}{(r-1)!} \sum_{p=1}^{r-1} \sum_{k=1}^r \{\phi(\sigma_k^{(p)}(\xi, \eta)) - \phi(\sigma_{k-1}^{(p)}(\xi, \eta))\}^2$$

が成り立つから,

$$(6) \sum_{[\xi:\eta]=r} (\phi(\eta) - \phi(\xi))^2 e^{-\beta V(\xi)} e^{-\beta V(\eta)} \leq \frac{e^{\beta\Lambda}}{(r-1)!} \sum_{[\xi:\eta]=r} \sum_{p=1}^{r-1} \sum_{k=1}^r \{\phi(\sigma_k^{(p)}(\xi, \eta)) - \phi(\sigma_{k-1}^{(p)}(\xi, \eta))\}^2 \exp\{-\beta\{V(\sigma_k^{(p)}(\xi, \eta)) \vee V(\sigma_{k-1}^{(p)}(\xi, \eta))\}\}$$

各 $\zeta \in \Xi$ と各 $j \in \{1, 2, \dots, N\}$ に対して,

$$(\phi(\tau_j \zeta) - \phi(\zeta))^2 \exp\{-\beta(V(\tau_j \zeta) \vee V(\zeta))\}$$

なる項は, 式 (6) の右辺の和の中に丁度 $r(N-1)(N-2)\dots(N-r+1)$ 回現れるから,

$$\sum_{[\xi:\eta]=r} (\phi(\eta) - \phi(\xi))^2 e^{-\beta V(\xi)} e^{-\beta V(\eta)} \leq e^{\beta\Lambda} r \binom{N-1}{r-1} \sum_{[\xi:\eta]=1} (\phi(\eta) - \phi(\xi))^2 \exp\{-\beta(V(\xi) \vee V(\eta))\}$$

従って (4), (5) と併せて,

$$\text{var}_\beta(\phi) \leq \frac{e^{\beta\Lambda}}{Z_\beta} \sum_{r=1}^N r \binom{N-1}{r-1} E_\beta(\phi, \phi)$$

を得る。

Theorem 2.2

$$c_\beta \geq e^{-\beta\Lambda} \frac{Z_\beta}{(N+1)2^{N-2}} \geq \frac{4}{N+1} \exp\{-\beta(\Lambda + \frac{1}{2N} \sum_{\xi} V(\xi))\}$$

なる評価が成立する。

(Proof) 第一の不等式は上の議論と,

$$\sum_{r=0}^{N-1} (r+1) \binom{N-1}{r} = (N+1)2^{N-2}$$

であることより従う。また \log が上に凸であることを用いれば,

$$\log Z_\beta \geq N \log 2 - \frac{\beta}{2N} \sum_{\xi} V(\xi)$$

を得る。従って第二の不等式も明らか。□

Theorem 2.2 は, $\Lambda + (1/2N) \sum_{\xi} V(\xi)$ が小であれば取束の速度を大きく見積もることができることを示している。

3 計算機実験

最適化問題 Q.2 を PMC アルゴリズムを用いて解いてみよう。A をアフィン変換

$$A: [0, 1]^4 \rightarrow [-1, 1]^4, \\ (x, y, z, w) \rightarrow (2x - 1, 2y - 1, 2z - 1, 2w - 1)$$

とする。\$[0, 1]^4\$ 上で関数 \$f \circ A\$ を最適化すれば良い。\$4N\$ 個の \$\{0, 1\}\$-スピン系の状態 \$\xi\$ に 2 進小数で表した座標をもつ点

$$b_N(\xi) \stackrel{\text{def.}}{=} \begin{aligned} & (0. \xi(0) \xi(1) \cdots \xi(N-1), \\ & 0. \xi(N) \xi(N+1) \cdots \xi(2N-1), \\ & 0. \xi(2N) \xi(2N+1) \cdots \xi(3N-1), \\ & 0. \xi(3N) \xi(3N+1) \cdots \xi(4N-1) \end{aligned} \in [0, 1]^4$$

を対応させ、\$U = f \circ A \circ b_N\$ と定める。

本実験の目的は PMC 法の特徴と有効性を示すことにある。そこで、我々は次の様な方針で最適解の探索を行なうことにした：

関数の大域的な傾向を掴むことと最適解を追い詰めて精密化を計ることとは互いに相反する要請であるが、大域的最適化を行なおうとする際は二者を調和させなければならない。従って、最適解探索は step by step で行なうことにした。各 step では目標とする関数値を設定し、目的関数の変形をポテンシャルにとって分布を生成する。変形とは、具体的には関数の上下を切断する。

既に見たように、スピン上のポテンシャル \$V\$ の平均値 \$(1/2^N A(V)) \sum_{\xi} V(\xi)\$ が小であれば目標となる確率測度への速い収束が期待できる。ただし、\$A(V)\$ は \$V\$ の amplitude。そこで \$V = H \circ U\$ を目的関数にとることにする。ただし、

$$H(y) = \begin{cases} c_1 & (\text{if } y > c_1), \\ y & (\text{if } c_2 \leq y \leq c_1), \\ m & (\text{if } y < c_2). \end{cases}$$

ここに \$m < c_2 \leq c_1\$。\$m\$ は“最小値の推定”と考えて良い。次々と出現する状態 \$\xi\$ を観測し、\$U(\xi)\$ が“目標値” \$c_2\$ より小になれば、\$c_1, c_2\$ 及び \$\beta\$ を更新して、新たなポテンシャルに基づいて分布の生成を行なう。\$c_1, c_2\$ の更新は次式に従って行なう：

$$\begin{aligned} c_1 &= C_1(U(\xi) - m) + m \\ c_2 &= C_2(U(\xi) - m) + m \end{aligned}$$

次に \$\beta\$ の更新を行なう：\$\beta = R/(c_1 - m)\$。ただし、\$C_1, C_2\$ 及び \$R\$ は、予め与える定数である。

実験結果を図 3 に示す。横軸は系の確率遷移の回数であり、その回数までの遷移で得られた最小値の常用対数をプロットしている。遷移を重ねるにつれて得られる値が指数的に減少していく様子を見ることが出来る。なお、単精度で解を得る為に \$N\$ は 32 にとる。また、\$1/\Delta t = rf \times \text{dim} \times N\$ と定めている。ここに \$\text{dim}\$ は問題の次元で、この場合は 4 である。また \$rf = 100\$ にとった。これは、(1) 式から (2) 式を導く際、\$1/10000\$ 以下の order を無視することに相当する。定数 \$C_1, C_2\$ 及び \$R\$ は次のように定めている：\$C_1 = 7, C_2 = 0.95, R = 70\$。これらの値は、経験的に決定した。また、\$m = 0\$ とおいた。ここで \$U\$ の最小値が 0 であるという知識を使っている。\$m\$ は \$\beta\$ や関数の変形を決定する為に便宜的に導入されたものであり、“最小値の推定”は PMC 法において本質的ではないが、上述の heuristic な戦略ではこの値のとり方次第で効果が左右されるであろう。PMC 法において確率測度の生成を如何なる戦略で利用するのが良いかという問題は、今後実証的に研究して行かなければならない。

4 おわりに

PMC 法の有効性が関数に依存することは、Section 2.2 に示した通りである — 1 点で 0、その点を除く \$2^{256} - 1\$ 個の点で 1 をとるような関数の最小値探索は、このアルゴリズムによっても現実的な目標とはなり得ない。しかし、我々のアルゴリズムが高い効果を発揮するような問題のクラスは極めて広いと考えられる。今後、整数計画法も含む様々な問題で有効性と信頼性を検証して行きたい。

PMC アルゴリズムを情報論的エントロピーの側面から考えてみよう。\$2^N\$ 個の実数の集合において各要素が“最小値である確率”を考える。その集合に関して何の情報もない時その系に関するエントロピーは \$N \log 2\$ であり、何らかの方法で最小値を確定した後エントロピーは 0 になる。Section 2.2 で定義した \$\Xi\$ と \$\Xi\$ 上の関数 \$V\$ に対し、要素を 1 つずつ調べる方法によって最小値を探索するとしよう。\$k\$ 個の要素を調べた段階で系のエントロピーは \$\log(2^N + 1 - k)\$ である。一方、PMC アルゴリズムはエントロピーが

$$H_{\beta} = \log Z_{\beta} - \beta \frac{d}{d\beta} \log Z_{\beta}$$

に近い確率を実現する。例えば、区間 \$[0, 1]\$ において関数 \$f(x) = x\$ の最小値を探索する問題に、Section 2 に述べた方法で、PMC 法を適用することを考える。この場合 \$H_{\beta} \approx N \log 2 + 1 - \log \beta\$ と計算されるから、等しいエントロピーを与えるという条件から

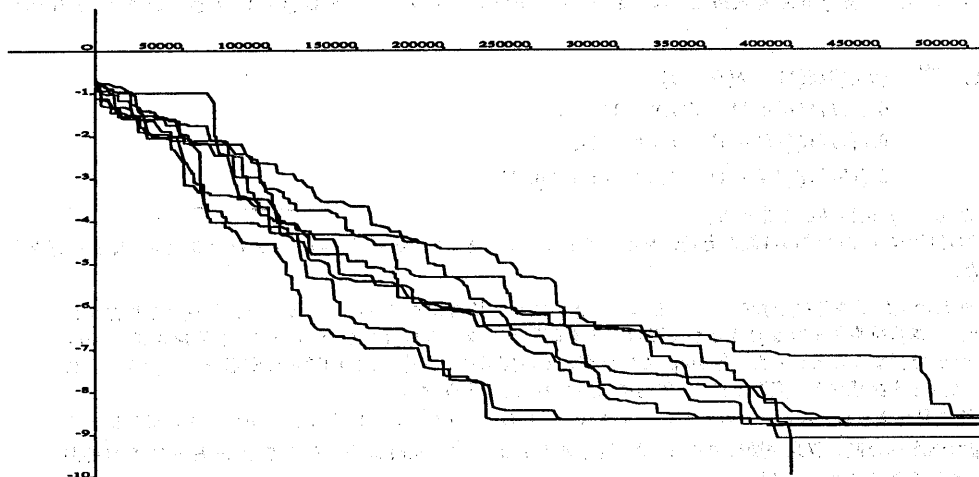


図 3: 実験結果

$$k = 1 + 2^N \left(1 - \frac{e}{\beta}\right)$$

なる関係式を得る。この式は、 N が大であるとき、 β の値が重大な意味を持つことを示している。即ち、theorem 2.1 及び theorem 2.2 を考慮すれば、 β の値は小である方が測度の生成は高速に行なわれるが、上式は β の値が小であっても生成された測度の持つ“効果”が大きいことを示している。

PMC アルゴリズムは、要するに幾つかの関数値を拾い出してはそれらの差に基づいて確率的な filter を作るということを繰り返すだけの単純なものである。この事は、エントロピーの減少と計算量との関係を知りたいという素朴な疑問に我々を導く。即ち

“どの位の手間を掛ければどの程度エントロピーを減少させることができるのだろうか。効率の上限はどこにあるのだろうか。”

参考文献

- [1] 伊藤 清: 確率論 III. 岩波講座 基礎数学, 1978
- [2] 今野 浩, 山下 浩: 非線形計画法. 日科技連出版社, 1978
- [3] R.J.Glauber: Time-Dependent Statistics of the Ising Model. Journal of Mathematical Physics, Vol. 4, No. 2, 1963, 294-307
- [4] R.Holley & D.Stroock: Simulated Annealing via Sobolev Inequalities. Communications in Mathematical Physics 115, 1988, 553-569
- [5] S.Kirkpatrick, C.D.Gelatt, Jr. & M.P.Vecchi: Optimization by simulated annealing. Science 220, 1983, 671-680
- [6] J.La Salle & S.Lefschetz: Stability by Liapunov's Direct Method with Application. Academic Press, 1961
- [7] P.J.M.van Laarhoven & E.H.L.Aarts: Simulated Annealing: Theory and Applications. Reidel, 1987
- [8] 斉藤 利弥: 常微分方程式 I. 岩波講座 基礎数学, 1976