

連続最適化のための進化戦略の数理

秋本洋平 (筑波大学 / 理研 AIP / 大阪大学)

2025 年 8 月 28-30 日 統計数理研究所

連続最適化および関連分野に関する夏期学校 2025

内容 : (1+1)-ES の一次収束の証明と収束レートの導出

1. 進化戦略とは
2. 確率的アルゴリズムの一次収束
3. 確率過程の一次収束に役立つ数学的ツール
4. Spherical 関数のもとでの (1+1)-ES の一次収束
5. 一般の関数クラスにおける (1+1)-ES の一次収束

目標

- ・ 確率的かつ適応的なアルゴリズムの解析テクニックの理解と実践
- ・ 進化戦略の理論保証の紹介

進化戦略とは

進化戦略: Evolution Strategies (ES) (*Evolutionsstrategie*, in German)

- ・ Ingo Rechenberg と Hans-Paul Schwefel らが 1960 年代に提案
- ・ 当時ベルリン工科大で航空機の翼型や噴射ノズルの形状最適化に取り組む
- ・ 空気抵抗の評価に流体シミュレーションが必要, 目的関数の勾配計算が困難
- ・ 自然界の進化プロセス (突然変異と自然淘汰) に着想
- ・ Ingo Rechenberg の博士論文として, ES を体系的にまとめ上げ (Rechenberg, 1973)
- ・ 遺伝的アルゴリズム (GAs, Holland (1975)) とほぼ同時期に独立して提案され, 進化計算の分野の基礎を形成
- ・ Rechenberg は他にも自然界から着想を得て工学に応用する bionics 分野で活躍

進化戦略のフレームワーク

$f: \mathbb{R}^d \rightarrow \mathbb{R}$ to be minimized

- 1: **while** not happy **do**
- 2: sample $\{x_i\}_{i=1}^\lambda$ i.i.d. from $\mathcal{P}_\theta$
- 3: compute $f(x_i)$ for $i = 1, \dots, \lambda$ and their rankings $r_1, \dots, r_\lambda$
- 4: update θ and internal states using $\{(x_i, f(x_i))\}_{i=1}^\lambda$
- 5: **end while**

- ・（基本的には）連続最適化アルゴリズム
- ・ 確率分布（多変量正規分布）から解を生成する確率的アルゴリズム
- ・ 目的関数の勾配を用いない zero-th order 最適化法
- ・ 目的関数の大小関係のみを用いる comparison-based 最適化法
- ・ 確率分布のパラメータ（平均ベクトル, 共分散行列など）を更新していく適応的アルゴリズム

(1+1)-ES with 1/5 Success Rule (Rechenberg, 1973)

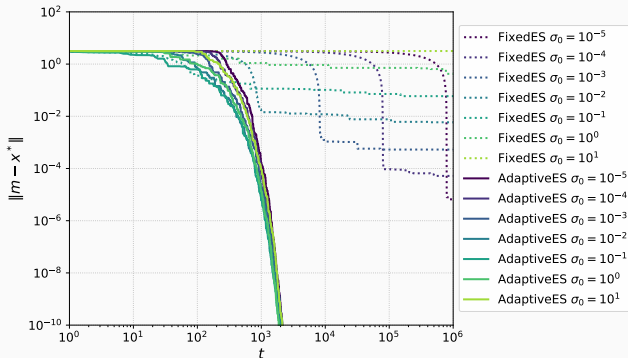
```
1: input  $m_0 \in \mathbb{R}^d, \sigma_0 > 0, f: \mathbb{R}^d \rightarrow \mathbb{R}$ 
2: parameter  $\alpha > 1$ 
3: for  $t = 1, 2, \dots$ , until stopping criterion is met do
4:   sample  $x_t \sim m_t + \sigma_t z_t$ , where  $z_t \sim \mathcal{N}(0, I)$ 
5:   if  $f(x_t) \leq f(m_t)$  then
6:      $m_{t+1} \leftarrow x_t$                                 ▷ move to the better solution
7:      $\sigma_{t+1} \leftarrow \sigma_t \alpha$                   ▷ increase the step size
8:   else
9:      $m_{t+1} \leftarrow m_t$                                 ▷ stay where we are
10:     $\sigma_{t+1} \leftarrow \sigma_t \alpha^{-\frac{1}{4}}$           ▷ decrease the step size
11:   end if
12: end for
```

アイデア : ステップサイズ σ の適応

成功 ($f(x_t) \leq f(m_t)$ となる) 確率が 1/5 になるように σ を適応

ステップサイズ適応は必要不可欠

$f(x) = \|x - x^*\|$ での結果 ($d = 10, m_0 = (1, \dots, 1)$)



σ 固定の場合の振る舞い

- $\|m - x^*\| \propto \sigma$ 程度で停滞
- 停滞するまでのイテレーションが $\propto 1/\sigma$

1/5 Success Rule のロジック

$f(x) = \|x\|$ を考える.

- ・ 仮に m を更新しないならば, 成功すると α 倍され, 失敗すると $\alpha^{-\frac{1}{4}}$ 倍されるので, σ は成功確率が 1/5 となる値で安定
- ・ ステップサイズが小さすぎる状況:
解生成される範囲において目的関数が線形関数で近似される
 $\Rightarrow$
 - ・ 成功確率が 1/2 に近い
 - ・ σ を大きくすることで, 成功確率が下がるが, 一度の更新で大きく改善できる
- ・ ステップサイズが多きすぎる状況:
目的関数が改善する領域が解生成される範囲に対して著しく狭い
 $\Rightarrow$
 - ・ 成功確率が 0 に近い
 - ・ σ を小さくすることで, 成功確率が上がり, 改善できる

Definition (Progress Rate)

$$\varphi(m, \sigma) = -\mathbb{E} \left[\log \frac{\|m_{t+1} - x^*\|}{\|m_t - x^*\|} \mid m_t = m, \sigma_t = \sigma \right]$$

$\varphi(m, \sigma)$ は高いほど望ましい：

$$\mathbb{E} [\log \|m_{t+1} - x^*\| \mid m_t = m, \sigma_t = \sigma] = \log \|m_t - x^*\| - \varphi(m, \sigma)$$

Definition (Success Probability)

$$p_0^{\text{succ}}(m, \sigma) = \Pr_{z \sim \mathcal{N}(0, I)} [f(m + \sigma z) \leq f(m)]$$

Sphere 関数のもとでの Progress Rate

$f(x) = \|x\|$ の場合 : Progress Rate は $\bar{\sigma} = \frac{d\sigma}{\|m\|}$ の関数

$$\begin{aligned} d\varphi(m, \sigma) &= -d\mathbb{E} \left[\log \frac{\|m_{t+1}\|}{\|m_t\|} \mid m_t = m, \sigma_t = \sigma \right] \\ &= -d\mathbb{E} \left[\log \frac{\|m + \sigma z \mathbb{I}\{\|m + \sigma z\| \leq \|m\|\}\|}{\|m\|} \right] \\ &= -d\mathbb{E} [\log \|e_m + (\bar{\sigma}/d)z \mathbb{I}\{\|m + \sigma z\| \leq \|m\|\}\|] \\ &= -d\mathbb{E} [\log \|e_m + (\bar{\sigma}/d)z \mathbb{I}\{\|e_m + (\bar{\sigma}/d)z\| \leq 1\}\|] \\ &= -d\mathbb{E} [\log \|e_1 + (\bar{\sigma}/d)\tilde{z} \mathbb{I}\{\|e_1 + (\bar{\sigma}/d)\tilde{z}\| \leq 1\}\|] \\ &= -d\mathbb{E} [\log \|e_1 + (\bar{\sigma}/d)z \mathbb{I}\{\|e_1 + (\bar{\sigma}/d)z\| \leq 1\}\|] \\ &= \varphi_d(\bar{\sigma}) \end{aligned}$$

$e_m = \frac{m}{\|m\|}$, $e_1 = (1, 0, \dots, 0)$, R : 直交行列 $e_1 = Re_m$, $\tilde{z} = Rz$.

注 : $z \sim \mathcal{N}(0, I)$ のとき, $\tilde{z} \sim \mathcal{N}(0, RR^\top = I)$ より, 両者は同じ多変量標準正規分布に従う

Sphere 関数のもとでの Success Probability

$f(x) = \|x\|$ の場合 : Success Probability は $\bar{\sigma} = \frac{d\sigma}{\|m\|}$ の関数

$$\begin{aligned} p_0^{\text{succ}}(m, \sigma) &= \Pr_{z \sim \mathcal{N}(0, I)} [f(m + \sigma z) \leq f(m)] \\ &= \Pr_{z \sim \mathcal{N}(0, I)} [\|m + \sigma z\| \leq \|m\|] \\ &= \Pr_{z \sim \mathcal{N}(0, I)} [\|e_m + (\bar{\sigma}/d)z\| \leq 1] \\ &= \Pr_{z \sim \mathcal{N}(0, I)} [\|e_1 + (\bar{\sigma}/d)\tilde{z}\| \leq 1] \\ &= \Pr_{z \sim \mathcal{N}(0, I)} [\|e_1 + (\bar{\sigma}/d)z\| \leq 1] = p_{0,d}^{\text{succ}}(\bar{\sigma}) \end{aligned}$$

Lemma (Lemma 3.1 in Akimoto, Auger, and Glasmachers, 2018)

1. The image of $p_{0,d}^{\text{succ}}$ is $(0, 1/2)$.
2. $p_{0,d}^{\text{succ}}$ is strictly decreasing, thus bijective.

Success Probability と Progress Rate の漸近的解析 (1/2)

$d \rightarrow \infty$ を考えることで解釈できる形に帰着させる

$$\begin{aligned} p_{0,d}^{\text{succ}}(\bar{\sigma}) &= \Pr_{z \sim \mathcal{N}(0,I)} [\|e_1 + (\bar{\sigma}/d)z\| \leq 1] \\ &= \Pr_{z \sim \mathcal{N}(0,I)} [\|e_1 + (\bar{\sigma}/d)z\|^2 \leq 1] \\ &= \Pr_{z \sim \mathcal{N}(0,I)} [d(\|e_1 + (\bar{\sigma}/d)z\|^2 - 1) \leq 0] \end{aligned}$$

ここで,

$$\begin{aligned} d \left(\left\| e_1 + \frac{\bar{\sigma}}{d} z \right\|^2 - 1 \right) &= d \left(\left(1 + \frac{\bar{\sigma}}{d} z_1 \right)^2 + \left(\frac{\bar{\sigma}}{d} \right)^2 \sum_{i=2}^d z_i^2 - 1 \right) \\ &= 2\bar{\sigma}z_1 + \frac{\bar{\sigma}^2}{d} \sum_{i=1}^d z_i^2 \end{aligned}$$

(補足) 大数の強法則

Theorem (Strong Law of Large Numbers for IID)

Let $X_1, X_2, \dots$ is an infinite sequence of independent and identically distributed (i.i.d.) random variables with expected value $\mathbb{E}[X_i] = \mu$.

Then, $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i \stackrel{w.p. 1}{=} \mu$, i.e.,

$$\Pr \left[\lim_n \frac{1}{n} \sum_{i=1}^n X_i = \mu \right] = 1.$$

$z_1^2, \dots, z_d^2$ は i.i.d. で $\mathbb{E}[z_i^2] = 1$ より,

$$d \left(\left\| e_1 + \frac{\bar{\sigma}}{d} z \right\|^2 - 1 \right) = 2\bar{\sigma}z_1 + \frac{\bar{\sigma}^2}{d} \sum_{i=1}^d z_i^2 \xrightarrow[w.p. 1]{d \rightarrow \infty} 2\bar{\sigma}z_1 + \bar{\sigma}^2$$

Success Probability と Progress Rate の漸近的解析 (2/2)

$d \rightarrow \infty$ を考えると Success Probability を標準正規分布の累積分布関数 Φ を用いて書き下せる :

$$\begin{aligned}\lim_{d \rightarrow \infty} p_{0,d}^{\text{succ}}(\bar{\sigma}) &= \lim_{d \rightarrow \infty} \Pr_{z \sim \mathcal{N}(0, I)} [d(\|e_1 + (\bar{\sigma}/d)z\|^2 - 1) \leq 0] \\ &= \Pr_{z_1 \sim \mathcal{N}(0,1)} [2\bar{\sigma}z_1 + \bar{\sigma}^2 \leq 0] \\ &= \Pr_{z_1 \sim \mathcal{N}(0,1)} \left[z_1 \leq -\frac{\bar{\sigma}}{2} \right] \\ &= \Phi \left(-\frac{\bar{\sigma}}{2} \right)\end{aligned}$$

同様にして Progress Rate も簡単化できる :

$$\lim_{d \rightarrow \infty} \varphi_d(\bar{\sigma}) = -2\bar{\sigma} \mathbb{E} \left[\left(z_1 + \frac{\bar{\sigma}}{2} \right) \mathbb{I}\{z_1 \leq -\bar{\sigma}/2\} \right]$$

d が十分に大きければ, これらの量を極限で近似できる

成功確率 1/5 は最適 $\bar{\sigma}$ のもとでの成功確率に近い値

Sphere 関数のもとでの Progress Rate が最大になる $\bar{\sigma}$

$$\bar{\sigma}_{\infty}^* = \operatorname{argmin}_{\bar{\sigma}} \varphi_{\infty}(\bar{\sigma})$$

最適 $\bar{\sigma}_{\infty}^*$ のもとでの成功確率

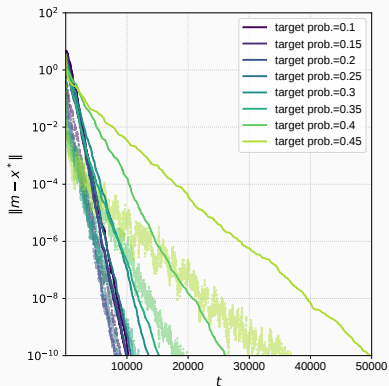
$$p_{0,\infty}^{\text{succ}}(\bar{\sigma}_{\infty}^*) = \Phi\left(-\frac{\bar{\sigma}_{\infty}^*}{2}\right) \approx 0.27$$

Corridor モデル（線形の目的関数と $2(d-1)$ 個の不等式制約を持つ問題）で同様の解析をすると

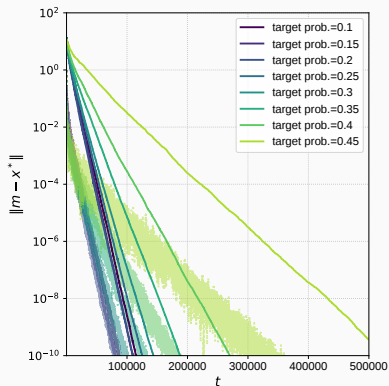
$$\lim_{d \rightarrow \infty} p_{0,d}^{\text{succ}}(\bar{\sigma}_d^*) \approx 0.184$$

これらの観察と経験則から 1/5 を成功確率のターゲットとしている。

凸二次関数のもとでの実験的検証



$d = 10$



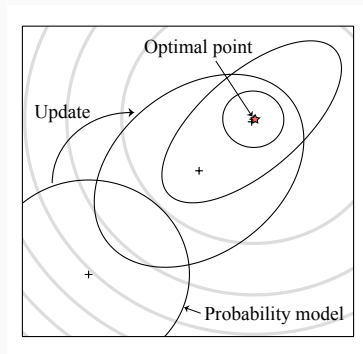
$d = 100$

目的関数：凸二次関数 $f(x) = \frac{1}{2} \sum_{i=1}^d 100^{\frac{i-1}{d-1}} x_i^2$

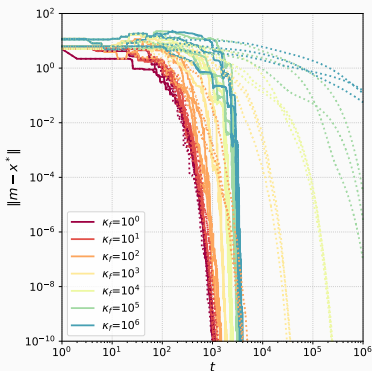
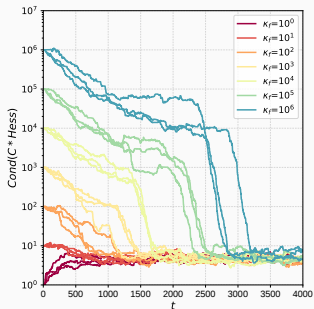
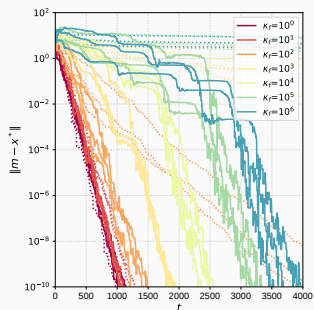
実線： $\|m - x^*\|$ ，破線： σ

(1+1)-CMA-ES : 共分散行列適応を行う (1+1)-ES

- 1: **input** $m_0 \in \mathbb{R}^d, \sigma_0 > 0, \Sigma_0 = I,$
 $f: \mathbb{R}^d \rightarrow \mathbb{R}$
- 2: **parameter** $\alpha > 1$
- 3: **for** $t = 1, 2, \dots$, **do**
- 4: sample $x_t \sim m_t + \sigma_t \mathcal{N}(0, \Sigma_t)$
- 5: **if** $f(x_t) \leq f(m_t)$ **then**
- 6: $m_{t+1} \leftarrow x_t$
- 7: $\sigma_{t+1} \leftarrow \sigma_t \alpha$
- 8: **update** Σ_{t+1}
- 9: **else**
- 10: $m_{t+1} \leftarrow m_t$
- 11: $\sigma_{t+1} \leftarrow \sigma_t \alpha^{-\frac{1}{4}}$
- 12: **update** Σ_{t+1}
- 13: **end if**
- 14: **end for**



悪スケールな凸二次関数における比較

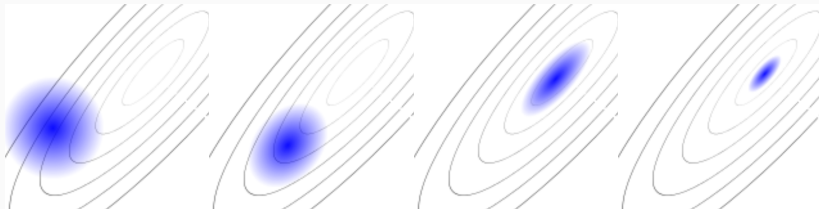


$f(x) = \frac{1}{2}x^T H x$
 $H = R^T D R$ (R : 直交行列,
 $D = \text{diag}(1, \dots, 10^{\kappa_f \frac{i-1}{d-1}}, \dots, 10^{\kappa_f})$)
 実線 : (1+1)-CMA-ES, 破線 : (1+1)-ES

“The” CMA-ES

Hansen and Ostermeier (2001), Hansen, Müller, and Koumoutsakos (2003), Hansen and Kern (2004), Jastrebski and Arnold (2006), and Akimoto and Hansen (2020)

1. Generate candidate solutions $\{x_i\}_{i=1,\dots,\lambda}$ from $\mathcal{N}(m_t, \sigma_t^2 C_t)$
2. Evaluate $f(x_i)$ and sort them, $f(x_{1:\lambda}) < \dots < f(x_{\lambda:\lambda})$.
3. Update (m, σ, C) using the sorted solutions $(x_{i:\lambda})_{i=1,\dots,\lambda}$.



(m, C) update based on the **Natural Gradient** of $J(\theta) = \mathbb{E}_{x \sim p_\theta}[(W \circ f)(x)]$, i.e., $\theta \leftarrow \theta + \eta \mathcal{I}_\theta^{-1} \nabla \mathbb{E}[(W \circ f)(x)]$ (Ollivier et al., 2017)

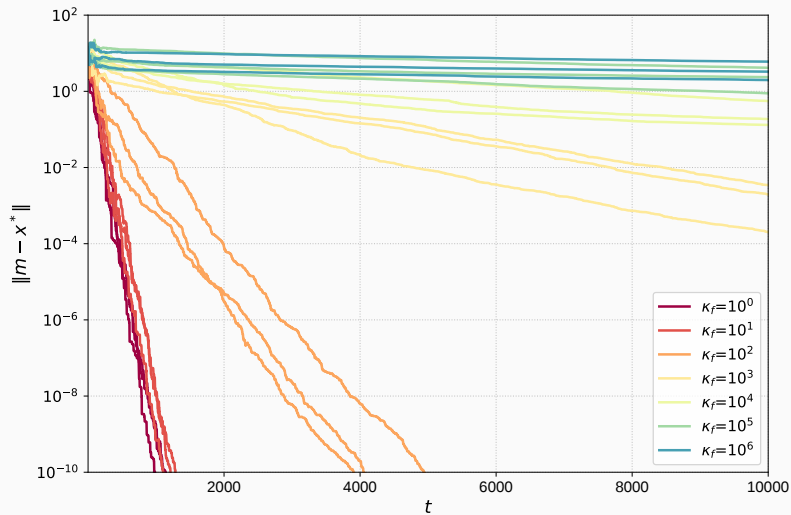
確率的アルゴリズムの一次収束： EFHT と確率 1 収束レート

(1+1)-ES のアルゴリズム (本日解析するアルゴリズム)

```
1: input  $m_0 \in \mathbb{R}^d, \sigma_0 > 0, f: \mathbb{R}^d \rightarrow \mathbb{R}$ 
2: parameter  $\alpha_{\uparrow} > 1 > \alpha_{\downarrow} > 0$ 
3: for  $t = 1, 2, \dots$ , until stopping criterion is met do
4:   sample  $x_t \sim m_t + \sigma_t \mathcal{N}(0, I)$ 
5:   if  $f(x_t) \leq f(m_t)$  then
6:      $m_{t+1} \leftarrow x_t$                                 ▷ move to the better solution
7:      $\sigma_{t+1} \leftarrow \sigma_t \alpha_{\uparrow}$           ▷ increase the step size
8:   else
9:      $m_{t+1} \leftarrow m_t$                                 ▷ stay where we are
10:     $\sigma_{t+1} \leftarrow \sigma_t \alpha_{\downarrow}$         ▷ decrease the step size
11:   end if
12: end for
```

成功確率のターゲットは $\frac{\log(1/\alpha_{\downarrow})}{\log(\alpha_{\uparrow}/\alpha_{\downarrow})}$

(1+1)-ES の一次収束



確率過程の一次収束に関する定義 1 : First Hitting Time

Definition (First Hitting Time, FHT)

Let $\{X_t : t \geq 0\}$ be a sequence of real-valued random variables with initial condition $X_0 = \beta_0 \in \mathbb{R}$. For $\beta < \beta_0$, the first hitting time T_β^X of X_t to the set $(-\infty, \beta]$ is defined as $T_\beta^X = \inf\{t : X_t \leq \beta\}$.

T_β^X は 0 以上の値を取る確率変数

X_t が確率変数なので, $X_t \leq \beta$ となるタイミングも確率的.

(1+1)-ES の First Hitting Time

(1+1)-ES の最適解の ϵ 近傍への First Hitting Time

$X_t = \log \|m_t - x^*\|$ (f の最適解 x^* からの距離の対数) とする. $\beta = \log \epsilon$ とすれば, $T_\beta^x = \inf \{t : X_t \leq \beta\}$ は (1+1)-ES の解が最適解の ϵ 近傍に初めて到達するまでのステップ数.

First Hitting Time について示したいこと

ある正の定数 C_T と C_R が存在し, 任意の $\epsilon > 0$ について Expected FHT (EFHT) が

$$\mathbb{E}[T_{\log(\epsilon)}^x] \leq C_T + C_R^{-1} \cdot \log \left(\frac{\|m_0 - x^*\|}{\epsilon} \right)$$

を満たす.

- ϵ 近傍に到達するまでに, $O(\log(1/\epsilon))$
(決定的アルゴリズムの一次収束と対応)
- $-C_R$ は収束曲線の傾きに対応

確率過程の一次収束に関する定義 2 : 収束レート

Definition (Convergence rate)

Let $\{X_t > 0\}_{t \geq 0}$ be the sequence of real-valued random variables. If there exists a constant $CR > 0$ satisfying

$$\Pr \left[\limsup_{t \rightarrow \infty} \frac{1}{t} \log \left(\frac{X_t}{X_0} \right) = -CR \right] = 1$$

then $\exp(-CR)$ is called the convergence rate.

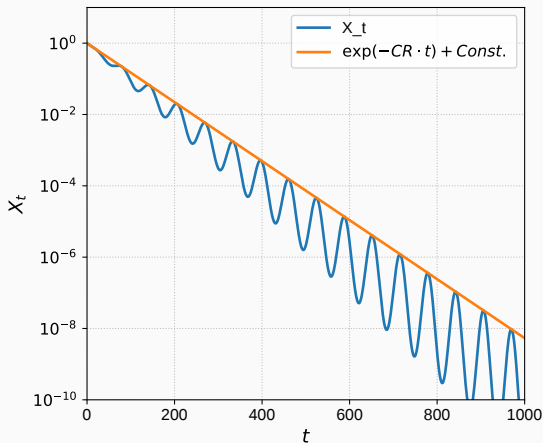
Cf. 決定的な系列についての収束レート

$$\lim_{t \rightarrow \infty} \frac{\|X_{t+1} - X^*\|}{\|X_t - X^*\|}$$

—CR は収束曲線の傾き

—CR は収束曲線を対数スケールで描いた際の傾きの極限

$\limsup$ をとっているので、上側の包絡線の傾き。



(1+1)-ES の収束レート

Convergence rate

$X_t = \|m_t - x^*\|$ とする。このとき, $CR > 0$ が存在し,

$$\Pr \left[\limsup_{t \rightarrow \infty} \frac{1}{t} \log \left(\frac{\|m_t - x^*\|}{\|m_0 - x^*\|} \right) = -CR \right] = 1$$

となる, すなわち, 解 m_t が確率 1 で最適解 x^* に向かって一次収束すること, および CR の値.

Note:

$$\frac{1}{t} \log \left(\frac{\|m_t - x^*\|}{\|m_0 - x^*\|} \right) = \frac{1}{t} \sum_{i=0}^{t-1} \log \left(\frac{\|m_{i+1} - x^*\|}{\|m_i - x^*\|} \right)$$

なので, 傾きの平均を考えているとも解釈される.

確率過程の一次収束に役立つ数学的 ツールの紹介

EFHT のバウンド技術 : ドリフト解析

確率過程の 1 ステップでの期待改善量 $\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] - X_t$ を考えることで, EFHT にバウンドを与える技術 (Hajek, 1982)

様々なバリエーションが存在 (証明したい結論や利便性に応じて選択)

- Additive Drift
- Multiplicative Drift
- Negative Drift

有界な状態空間のもとでの加法的ドリフト解析

Theorem (Additive Drift (Lengler and Steger, 2016))

Let $\{X_t : t \geq 0\}$ be a sequence of integrable ($\mathbb{E}[|X_t|] < \infty$) random variables adapted to a filtration $\{\mathcal{F}_t : t \geq 0\}$ with state space $[0, \infty)$ and $X_0 = x_0$. Let $T = \inf\{t : X_t = 0\}$. If there exists $c > 0$ such that

$$\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] - X_t \leq -c$$

holds for all $t \geq 0$ as long as $X_t > 0$, then

$$\mathbb{E}[T] \leq \frac{x_0}{c} .$$

解釈

- X_t の条件付き期待値の減少量（ドリフト）が上から負の定数でバウンドされるなら，EFHT の上限が得られる
- 決定的な系列 $x_{t+1} - x_t \leq -c$ について得られる上限と対応

(補足) $\mathbb{R}^d$ 上の Filtration

$\mathcal{F}_t$ はステップ t までに得られる情報の集合

Definition (Filtration)

A **filtration** $\{\mathcal{F}_t\}_{t \geq 0}$ is a non-decreasing ($\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$) family of sub- σ -algebras of the Borel σ -algebra $\mathcal{B}(\mathbb{R}^d)$.

$\mathcal{F}_t^X$ は $X_0, X_1, \dots, X_t$ から得られる情報の集合

Definition (Natural Filtration)

For a sequence $\{X_t\}_{t \geq 0}$ of random variables in $\mathbb{R}^d$, a **natural filtration** $\{\mathcal{F}_t^X \subseteq \mathcal{B}(\mathbb{R}^d)\}_{t \geq 0}$ is a sequence where $\mathcal{F}_t^X$ is the σ -algebra generated by the random variables $X_0, X_1, \dots, X_t$, i.e., $\mathcal{F}_t^X = \sigma(X_0, X_1, \dots, X_t)$.

(補足) R^d 上の確率過程

Definition (Adapted Process)

A sequence $\{X_t\}_{t \geq 0}$ of random variables is adapted to a filtration $\{\mathcal{F}_t\}_{t \geq 0}$ if $\mathcal{F}_t^X \subseteq \mathcal{F}_t$ for every $t \geq 0$.

例えば, $X_t = \log \|m_t - x^*\|$ を考える場合,

- $\mathbb{E}[X_{t+1} \mid X_0, X_1, \dots, X_t]$ もしくは $\mathbb{E}[X_{t+1} \mid \mathcal{F}_t^X]$ を考える場合,
 $\sigma_0, \sigma_1, \dots, \sigma_t$ の情報を持っていないことになり, これらに条件付けられていない期待値を計算することになる.
- アルゴリズムが得られる情報を $\mathcal{F}_t$ とした場合, $\mathbb{E}[X_{t+1} \mid \mathcal{F}_t]$ は
 $\sigma_0, \sigma_1, \dots, \sigma_t$ の情報も持ったうえでの条件付き期待値となる.

証明の方針

- ・ 元々の確率過程の代わりに, $X_{T+1} = X_{T+2} = \dots = 0$ となる確率過程を考えても T は変わらないので, 一般性を失うことなくこれを仮定. (このようなプロセスを**ストッププロセス**と呼ぶ)
- ・ T は非負の (整数値を取る) 確率変数であるため,

$$\begin{aligned}\mathbb{E}[T] &= \sum_{t=0}^{\infty} t \Pr[T = t] \\ &= \sum_{t=0}^{\infty} \Pr[T > t]\end{aligned}$$

- ・ 条件式を用いることで, $\Pr[T > t]$ をバウンド

Additive Drift の証明 (2/4)

条件文

$$\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] - X_t \leq -c \quad \text{if } X_t > 0$$

$X_t > 0 \iff T > t$ なので, (ストッププロセスを考えていることに注意)

$$\iff (\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] - X_t) \mathbb{I}\{T > t\} \leq -c \mathbb{I}\{T > t\}$$

$$\iff \mathbb{E}[X_{t+1} \mathbb{I}\{T > t\} \mid \mathcal{F}_t] - X_t \mathbb{I}\{T > t\} \leq -c \mathbb{I}\{T > t\}$$

両辺について期待値をとれば,

$$\implies \mathbb{E}[\mathbb{E}[X_{t+1} \mathbb{I}\{T > t\} \mid \mathcal{F}_t]] - \mathbb{E}[X_t \mathbb{I}\{T > t\}] \leq -c \mathbb{E}[\mathbb{I}\{T > t\}]$$

$$\iff \mathbb{E}[X_{t+1} \mathbb{I}\{T > t\}] - \mathbb{E}[X_t \mathbb{I}\{T > t\}] \leq -c \Pr[T > t]$$

Additive Drift の証明 (3/4)

$T \leq t \implies X_t = 0$ より,

$$X_t = X_t \mathbb{I}\{T > t\} + X_t \mathbb{I}\{T \leq t\} = X_t \mathbb{I}\{T > t\}$$

期待値をとれば,

$$\implies \mathbb{E}[X_t] = \mathbb{E}[X_t \mathbb{I}\{T > t\}]$$

同様に, $T \leq t \implies X_{t+1} = 0$ なので,

$$\mathbb{E}[X_{t+1}] = \mathbb{E}[X_{t+1} \mathbb{I}\{T > t\}]$$

以上より, 任意の $t \geq 0$ において,

$$\mathbb{E}[X_{t+1}] - \mathbb{E}[X_t] = \mathbb{E}[X_{t+1} \mathbb{I}\{T > t\}] - \mathbb{E}[X_t \mathbb{I}\{T > t\}]$$

Additive Drift の証明 (4/4)

まとめると,

$$\mathbb{E}[X_{t+1}] - \mathbb{E}[X_t] \leq -c \Pr[T > t]$$

$\mathbb{E}[T]$ の式に代入すれば, $\mathbb{E}[X_0] = x_0$ および $\mathbb{E}[X_t] \geq 0$ に注意すれば,

$$\begin{aligned}\mathbb{E}[T] &= \sum_{t=0}^{\infty} \Pr[T > t] \\ &\leq \frac{1}{c} \sum_{t=0}^{\infty} \mathbb{E}[X_t] - \mathbb{E}[X_{t+1}] \\ &= \lim_{\tau \rightarrow \infty} \frac{1}{c} \sum_{t=0}^{\tau} \mathbb{E}[X_t] - \mathbb{E}[X_{t+1}] \\ &= \lim_{\tau \rightarrow \infty} \frac{1}{c} (\mathbb{E}[X_0] - \mathbb{E}[X_{\tau+1}]) \\ &\leq \lim_{\tau \rightarrow \infty} \frac{1}{c} \mathbb{E}[X_0] = \frac{x_0}{c}\end{aligned}$$

□

有界でない場合の困難さ

解析対象は $X_t = \log \|m_t - x^*\| \in (-\infty, \infty) \implies$ **Unbounded**

Unbounded な状態空間の場合、同じ条件では $\mathbb{E}[T]$ をバウンドできない

反例 : $p \in (0, 1)$ に対して,

$$X_{t+1} = \begin{cases} X_t & \text{w.p. } 1-p \\ X_t - 1/p & \text{w.p. } p \end{cases}$$

このときドリフトは p によらず

$$\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] - X_t = -1$$

直感的な理由 : $T = \inf\{t : X_t \leq \log \epsilon\}$ の期待値は p が小さくなるといくらでも大きくなり得る。(一度のジャンプで到達できるが、ジャンプが起きまでの期待ステップ数は $1/p$)

証明上の困難さ : 最終ステップの $\mathbb{E}[X_{\tau+1}]$ がバウンドできない.

Unbounded Domain での EFHT バウンドを得るメインツール

Theorem (Truncated Additive Drift)

Let $\{(X_t, \mathcal{F}_t) : t \geq 0\}$ be an integrable adapted process in $\mathbb{R}$ and $X_0 = \beta_0$. Let $T_\beta^X = \inf\{t : X_t \leq \beta\}$. If there exist $A \geq B > 0$ such that

$$\mathbb{E}[\max\{X_{t+1} - X_t, -A\} \mathbb{I}\{T_\beta^X > t\} \mid \mathcal{F}_t] \leq -B \mathbb{I}\{T_\beta^X > t\} ,$$

then

$$\mathbb{E}[T_\beta^X] \leq \frac{A + \beta_0 - \beta}{B} .$$

困難さへの対処

大きすぎるジャンプをクリップした上でのドリフトをバウンドすることで EFHT をバウンドできるようにする

証明の方針

- 元々の確率過程の代わりに, $Y_0 = X_0$ かつ

$$Y_{t+1} = Y_t + \max\{X_{t+1} - X_t, -A\} \mathbb{I}\{T_\beta^X > t\} - B \mathbb{I}\{T_\beta^X \leq t\}$$

で定義される別のプロセスを考える.

- $t \leq T_\beta^X$ である限り,

$$X_t \leq Y_t$$

が保証される $\implies T_\beta^X \leq T_\beta^Y$

- Y_t のストッププロセス ($Y_{T_\beta^Y+1} = Y_{T_\beta^Y+2} = \dots = Y_{T_\beta^Y}$) を考えると, $Y_{T_\beta^Y} \geq \beta - A$ が得られるため, 有界ドメインでの Additive Drift の証明の最終ステップにおいて, $\mathbb{E}[Y_{T_\beta^Y}] \geq \beta - A$ を使えばバウンドが得られる.

Theorem (Negative Drift (Akimoto, Auger, and Glasmachers, 2018))

Let $\{(X_t, \mathcal{F}_t) : t \geq 0\}$ be an integrable adapted process in $\mathbb{R}$. Suppose that $X_0 = \beta_0$, $X_{t+1} \leq X_t$, and

$$\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] - X_t \geq -C$$

for some $C > 0$. Let $T_\beta^X = \inf\{t : X_t \leq \beta\}$. Then

$$\mathbb{E}[T_\beta^X] \geq -\frac{1}{2} + \frac{\beta_0 - \beta}{4C} .$$

Negative Drift の証明 (1/2)

- ・ ドリフト条件より, $T = \left\lfloor \frac{\beta_0 - \beta}{2C} \right\rfloor$ ステップにおいて,

$$\begin{aligned}\mathbb{E}[X_0 - X_T] &= \mathbb{E} \left[\sum_{t=0}^{T-1} (X_t - X_{t+1}) \right] = \mathbb{E} \left[\sum_{t=0}^{T-1} \mathbb{E}[(X_t - X_{t+1}) \mid \mathcal{F}_t] \right] \\ &\leq \sum_{t=0}^{T-1} C = T \cdot C \leq \frac{\beta_0 - \beta}{2}\end{aligned}$$

- ・ 上の不等式に Markov の不等式 ($\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$) を用いれば,

$$\Pr[X_0 - X_T \geq \beta_0 - \beta] \leq \frac{1}{2} \implies \Pr[T_\beta^X \leq T] \geq \frac{1}{2}$$

Negative Drift の証明 (2/2)

- ・ もう一度 Markov の不等式を適用すれば,

$$\underbrace{\mathbb{E}[T_\beta^x]}_{(1)} \geq \Pr[T_\beta^x \leq T] \cdot T \underbrace{\geq}_{(2)} \underbrace{\frac{T}{2}}_{(3)} \underbrace{\geq}_{(4)} \frac{\beta_0 - \beta}{4C} - \frac{1}{2}$$

(1) Markov's inequality

(2) 前スライドの結論

(3) $T = \left\lfloor \frac{\beta_0 - \beta}{2C} \right\rfloor$ の代入

(4) フロア関数の処理

□

Theorem (Prop. 2 of Morinaga et al. (IEEE TEVC 2024))

Let $\{(X_t, \mathcal{F}_t) : t \geq 0\}$ be an integrable adapted process in $\mathbb{R}$. Consider the following conditions:

C1 $\exists B^{\text{upper}} > 0$ such that $\mathbb{E}[X_{t+1} - X_t \mid \mathcal{F}_t] \leq -B^{\text{upper}}$ for all $t \geq 0$;

C2 $\exists B^{\text{lower}} > 0$ such that $\mathbb{E}[X_{t+1} - X_t \mid \mathcal{F}_t] \geq -B^{\text{lower}}$ for all $t \geq 0$;

C3 $\sum_{t=1}^{\infty} \text{Var}[X_t]/t^2 < \infty$.

(I) If C1 and C3 hold, then we have

$$\Pr \left[\limsup_{t \rightarrow \infty} \frac{1}{t} (X_t - X_0) \leq -B^{\text{upper}} \right] = 1 .$$

(II) If C2 and C3 hold, then we have

$$\Pr \left[\liminf_{t \rightarrow \infty} \frac{1}{t} (X_t - X_0) \geq -B^{\text{lower}} \right] = 1 .$$

(補足) Martingale Sequence

Definition (Martingale Sequence)

Let $\{(X_t, \mathcal{F}_t) : t \geq 0\}$ be an integrable adapted process in $\mathbb{R}$. If

$$\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] = X_t,$$

for every $t \geq 0$, then it is called a martingale sequence.

Martingale 過程に対する大数の強法則を用いた証明 (1/3)

(I) の証明

- $Z_{t+1} = X_{t+1} - \mathbb{E}[X_{t+1} \mid \mathcal{F}_t]$ と定義すると, $\{Z_t\}_{t \geq 1}$ は martingale 差分系列 ($\mathbb{E}[Z_{t+1} \mid \mathcal{F}_t] = 0$ for all $t \geq 0$)
- $\{Z_t\}_{t \geq 1}$ を用いれば,

$$\begin{aligned}\frac{1}{t}(X_t - X_0) &= \frac{1}{t} \sum_{i=1}^t (X_i - X_{i-1}) \\ &= \frac{1}{t} \sum_{i=1}^t (Z_i + \mathbb{E}[X_i - X_{i-1} \mid \mathcal{F}_{i-1}]) \\ &= \frac{1}{t} \sum_{i=1}^t Z_i + \frac{1}{t} \sum_{i=1}^t \mathbb{E}[X_i - X_{i-1} \mid \mathcal{F}_{i-1}] \\ &\leq \frac{1}{t} \sum_{i=1}^t Z_i \underbrace{- B^{\text{upper}}}_{\leftarrow C1}\end{aligned}$$

Martingale 過程に対する大数の強法則を用いた証明 (2/3)

Theorem (Strong Law of Large Numbers for Martingale)

Let $\{Y_t\}$ be martingale and let $Z_t = Y_t - Y_{t-1}$. If

$$\sum_{t=1}^{\infty} \frac{\mathbb{E}[|Z_t|^{2\alpha}]}{t^{1+\alpha}} < \infty$$

for some $\alpha \geq 1$, then

$$\Pr \left[\lim_{t \rightarrow \infty} \frac{Y_t}{t} = 0 \right] = 1 .$$

- C3 が $\alpha = 1$ の場合の martingale についての SLLN の条件に対応

$$\begin{aligned} \frac{\text{Var}[X_t]}{t^2} &= \frac{\mathbb{E}[(X_t - \mathbb{E}[X_t | \mathcal{F}_{t-1}]) + (\mathbb{E}[X_t | \mathcal{F}_{t-1}] - \mathbb{E}[X_t])^2]}{t^2} \\ &= \frac{\text{Var}[Z_t^2] + \text{Var}[\mathbb{E}[X_t | \mathcal{F}_{t-1}]]}{t^2} \geq \frac{\text{Var}[Z_t^2]}{t^2} \end{aligned}$$

Martingale 過程に対する大数の強法則を用いた証明 (3/3)

- Martingale 過程についての SLLN ($\alpha = 1$) を用いれば,

$$\begin{aligned} & \Pr \left[\limsup_{t \rightarrow \infty} \frac{1}{t} (X_t - X_0) \leq -B^{\text{upper}} \right] \\ & \underbrace{\geq}_{(1)} \Pr \left[\limsup_{t \rightarrow \infty} \left(\frac{1}{t} \sum_{i=1}^t Z_i - B^{\text{upper}} \right) \leq -B^{\text{upper}} \right] \\ & = \Pr \left[\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t Z_i \leq 0 \right] \\ & \geq \Pr \left[\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t Z_i = 0 \right] \underbrace{=}_{(2)} 0 \end{aligned}$$

$$(1) \left(\frac{1}{t} \sum_{i=1}^t Z_i - B^{\text{upper}} \right) \leq -B^{\text{upper}} \implies \frac{1}{t} (X_t - X_0) \leq -B^{\text{upper}}$$

(2) Martingale 過程に対する SLLN

まとめ：(1+1)-ES の一次収束を示すための数学的ツール

- ・ EFHT の上限：Truncated Additive Drift
- ・ EFHT の下限：Negative Drift
- ・ 確率 1 収束レートの上下限：Martingale 過程についての大数の強法則

EFHT の利点：初期値の影響を評価できる (C_T に該当)

確率 1 収束レートの利点：タイトなバウンドが得られやすい

Spherical 関数のもとでの $(1+1)$ -ES の一次収束

(1+1)-ES の確率過程

- 1: sample $x_t \sim m_t + \sigma_t \mathcal{N}(0, I) \stackrel{d}{=} m_t + \sigma_t \cdot z_t$ ($z_t \sim \mathcal{N}(0, I)$)
- 2: **if** $f(x_t) \leq f(m_t)$ **then**
- 3: $m_{t+1} \leftarrow x_t$ ▷ move to the better solution
- 4: $\sigma_{t+1} \leftarrow \sigma_t \alpha_{\uparrow}$ ▷ increase the step size
- 5: **else**
- 6: $m_{t+1} \leftarrow m_t$ ▷ stay where we are
- 7: $\sigma_{t+1} \leftarrow \sigma_t \alpha_{\downarrow}$ ▷ decrease the step size
- 8: **end if**

(1+1)-ES の確率過程

$\theta_t = (m_t, \log(\sigma_t))$ を状態を表す変数とする. $\{z_t \in \mathbb{R}^d\}_{t \geq 0}$ を独立に $\mathcal{N}(0, I)$ に従う確率ベクトル列とする. このとき, $\theta_0 = (m_0, \log(\sigma_0))$ であり, 以降 θ_{t+1} は θ_t から $\theta_{t+1} = \theta_t + \mathcal{G}(\theta_t, z_t; f)$ と定まる. ここで,

$$\begin{aligned} \mathcal{G}(\theta, z; f) := & (\sigma \cdot z, \log(\alpha_{\uparrow})) \cdot \mathbb{I}\{f(m + \sigma \cdot z) \leq f(m)\} \\ & + (0, \log(\alpha_{\downarrow})) \cdot \mathbb{I}\{f(m + \sigma \cdot z) > f(m)\}. \end{aligned}$$

(1+1)-ES の EFHT の下限を Negative Drift を用いて示す

Negative Drift for (1+1)-ES

Let $\{X_t = \log\|m_t - x^*\| : t \geq 0\}$. Suppose that $X_0 = \beta_0$, $X_{t+1} \leq X_t$, and

$$\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] - X_t \geq -C$$

for some $C > 0$. Let $T_\beta^X = \inf\{t : X_t \leq \beta = \log \epsilon\}$. Then

$$\mathbb{E}[T_\beta^X] \geq -\frac{1}{2} + \frac{\beta_0 - \beta}{4C}.$$

m_t が x^* の ϵ 近傍にヒットするまでの EFHT の下限は,

1. $\log\|m_{t+1} - x^*\| \leq \log\|m_t - x^*\|$
2. $\mathbb{E}[\log\|m_{t+1} - x^*\| \mid \mathcal{F}_t] - \log\|m_t - x^*\| \geq -C$

を示せば求まる.

単調非増加であることは Spherical 関数では自明

Spherical 関数 $f(x) = g(\|x - x^*\|)$, $g : \mathbb{R} \rightarrow \mathbb{R}$ は単調増加関数

$$\begin{aligned} f(m_{t+1}) \leq f(m_t) &\iff g(\|m_{t+1} - x^*\|) \leq g(\|m_t - x^*\|) \\ &\iff \|m_{t+1} - x^*\| \leq \|m_t - x^*\| \\ &\iff \log \|m_{t+1} - x^*\| \leq \log \|m_t - x^*\| \end{aligned}$$

(1+1)-ES は目的関数の単調増加変換に対して不変なので、以下 $f(x) = \|x - x^*\|$ と仮定.

Hit-and-Run : 最適ステップサイズを用いた理想アルゴリズム

- 1: sample $x_t \sim m_t + \sigma_t \mathcal{N}(0, I) \stackrel{d}{=} m_t + \sigma_t \cdot z_t$ ($z_t \sim \mathcal{N}(0, I)$)
- 2: **if** $f(x_t) \leq f(m_t)$ **then**
- 3: $m_{t+1} \leftarrow x_t$ ▷ move to the better solution
- 4: $\sigma_{t+1} \leftarrow \sigma_t \alpha_{\uparrow}$ ▷ increase the step size
- 5: **else**
- 6: $m_{t+1} \leftarrow m_t$ ▷ stay where we are
- 7: $\sigma_{t+1} \leftarrow \sigma_t \alpha_{\downarrow}$ ▷ decrease the step size
- 8: **end if**

Hit-and-Run アルゴリズム

- 1: sample z_t ($z_t \sim \mathcal{N}(0, I)$)
- 2: $\sigma_t^* = \operatorname{argmin}_{\sigma \geq 0} \|m_t + \sigma z_t - x^*\|$
- 3: $x_t^* = m_t + \sigma_t^* z_t$
- 4: $m_{t+1} = x_t^*$

Hit-and-Run を用いてドリフトをバウンド

- ・ (1+1)-ES の m_{t+1} と Hit-and-Run の $m_{t+1} = x_t^*$ について

$$\begin{aligned}\|m_{t+1} - x^*\| &= \min\{\|m_t + \sigma_t Z_t - x^*\|, \|m_t - x^*\|\} \\ &\geq \min_{\sigma \geq 0} \|m_t + \sigma Z_t - x^*\| = \|m_t + \sigma_t^* Z_t - x^*\| = \|x_t^* - x^*\|\end{aligned}$$

が得られるので, $\log\|m_{t+1} - x^*\| \geq \log\|x_t^* - x^*\|$

- ・ これを用いてドリフトの下限を導出

$$\begin{aligned}\mathbb{E}[\log\|m_{t+1} - x^*\| \mid \mathcal{F}_t] - \log\|m_t - x^*\| \\ \geq \mathbb{E}[\log\|x_t^* - x^*\| \mid \mathcal{F}_t] - \log\|m_t - x^*\|\end{aligned}$$

意義: 左辺 ((1+1)-ES のドリフト) は σ_t に依存するが,
右辺 (Hit-and-Run のドリフト) は m_t のみに依存するため解析が容易

Hit-and-Run のドリフト下限の導出

Lemma (Lemma 4 in Akimoto, Auger, and Glasmachers, 2018)

For $d \geq 2$,

$$\mathbb{E}[\log \|x_t^* - x^*\| \mid \mathcal{F}_t] - \log \|m_t - x^*\| \geq -\frac{1}{d}$$

証明のスケッチ

- 幾何学的に考えると、最適なステップサイズは $\|x_t^* - x^*\|^2 + \|\sigma_t^* z_t\|^2 = \|m_t - x^*\|^2$ となる点
- これを考慮するとドリフトは

$$\frac{1}{2W_{d-1}} \int_0^{\pi/2} \log(\sin(\theta)) \sin^{d-2}(\theta) d\theta$$

- これを $d \geq 2$ のもので下からバウンド

(1+1)-ES の EFHT の下限は $\Omega\left(d \log\left(\frac{\|m_0 - x^*\|}{\epsilon}\right)\right)$

Theorem (Theorem 4.7 in Akimoto et al., 2022)

For $d \geq 2$,

$$\mathbb{E}[T_{\beta}^x] \geq -\frac{1}{2} + \frac{d}{4} \log\left(\frac{\|m_0 - x^*\|}{\epsilon}\right) .$$

メッセージ

どんな σ 更新を考えても

- 一次収束より速い収束は起こり得ない: $\Omega\left(\log\left(\frac{\|m_0 - x^*\|}{\epsilon}\right)\right)$
- EFHT は次元数に比例: $\Omega(d)$ (非勾配法の限界)

(1+1)-ES のドリフトの上限に関するクイズ

クイズ: 以下を満たすベストな定数 $B \geq 0$ は？

$$\mathbb{E}[\log \|m_{t+1} - x^*\| \mid \mathcal{F}_t] - \log \|m_t - x^*\| \leq -B$$

(1+1)-ES のドリフトの上限に関するクイズの答え

正解 : $B = 0$

- $d \rightarrow \infty$ における Progress Rate (再掲) :

$$\lim_{d \rightarrow \infty} \varphi_d(\bar{\sigma}) = -2\bar{\sigma} \mathbb{E} \left[\left(z_1 + \frac{\bar{\sigma}}{2} \right) \mathbb{I}\{z_1 \leq -\bar{\sigma}/2\} \right]$$

$\bar{\sigma} = \frac{d\sigma}{\|m - x^*\|} \rightarrow 0$ で 0 に漸近.

- $d \rightarrow \infty$ における Success Probability (再掲) :

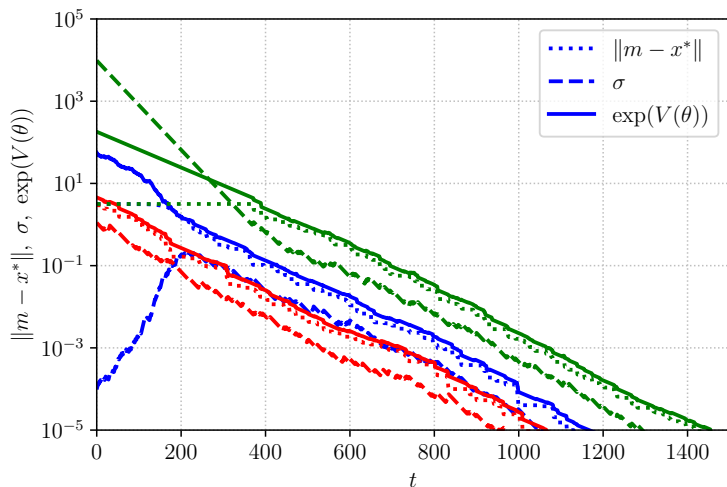
$$\lim_{d \rightarrow \infty} p_{0,d}^{\text{succ}}(\bar{\sigma}) = \Phi \left(-\frac{\bar{\sigma}}{2} \right)$$

$\bar{\sigma} \rightarrow \infty$ で 0 に漸近.

- $\sigma_t \ll \|m_t - x^*\|$ の場合, 改善量がいくらでも小さくなる
- $\sigma_t \gg \|m_t - x^*\|$ の場合, 改善確率がいくらでも小さくなる

(1+1)-ES の収束曲線

重要な観察: $\|m_t - x^*\|$ の傾きがほぼ 0 となる状況が存在



ドリフトを定数でバウンドできるポテンシャル関数を設計

アイデア

以下を満たすポテンシャル関数 $V(\theta)$ を設計

1. $\log\|m - x^*\| \leq V(\theta)$
2. $\mathbb{E}[V(\theta_{t+1}) \mid \mathcal{F}_t] - V(\theta_t) \leq -B$

T_β^x の代わりに, $T_\beta^V = \inf\{t : V(\theta_t) \leq \beta\}$ を解析

$$\cdot V(\theta_t) \leq \beta \implies \log\|m - x^*\| \leq \beta \text{ なので, } T_\beta^x \leq T_\beta^V$$

Spherical 関数における (1+1)-ES のポテンシャル関数

$$V(\theta) = \log\|m - x^*\| + \underbrace{v \log^+ \left[\frac{\alpha_{\uparrow} \ell \|m_t - x^*\|}{\sigma_t} \right]}_{\text{penalty for } \sigma_t \ll \|m_t - x^*\|} + \underbrace{v \log^+ \left[\frac{\sigma_t}{\alpha_{\downarrow} u \|m_t - x^*\|} \right]}_{\text{penalty for } \sigma_t \gg \|m_t - x^*\|}$$

$$\log^+(s) = \max\{0, \log(s)\}, v > 0, u > \ell > 0$$

Normalized Step-Size, Progress Rate, Success Probability

Normalized Step-Size (再掲)

$$\bar{\sigma} = \frac{d\sigma}{\|m - x^*\|}$$

Progress Rate (再掲)

$$\varphi(m, \sigma) = -\mathbb{E} \left[\log \frac{\|m_{t+1} - x^*\|}{\|m_t - x^*\|} \mid m_t = m, \sigma_t = \sigma \right]$$

Success Probability (少し拡張)

$$p_r^{\text{succ}}(m, \sigma) = \Pr_{z \sim \mathcal{N}(0, I)} [f(m + \sigma z) \leq (1 - r)f(m)]$$

Progress Rate と Success Probability は $\bar{\sigma}$ の関数

Progress Rate:

$$d\varphi(m, \sigma) = -d\mathbb{E} [\log \|e_1 + (\bar{\sigma}/d)z\| \mathbb{I}\{\|e_1 + (\bar{\sigma}/d)z\| \leq 1\}] =: \varphi_d(\bar{\sigma})$$

Success Probability:

$$p_r^{\text{succ}}(m, \sigma) = \Pr[\|e + d^{-1}\bar{\sigma}_t\mathcal{N}\| \leq 1 - r] =: p_r^{\text{succ}}(\bar{\sigma}_t)$$

Lemma (Lemma 1 of Akimoto, Auger, and Glasmachers, 2018)

- $p_r^{\text{succ}}(\bar{\sigma})$ is positive and continuous for $r \in [0, 1)$
- $p_{r=0}^{\text{succ}}(\bar{\sigma})$ is strictly decreasing (thus bijective)
- $p_{r=0}^{\text{succ}}(\bar{\sigma}) \rightarrow \frac{1}{2}$ as $\bar{\sigma} \rightarrow 0$ and $p_{r=0}^{\text{succ}}(\bar{\sigma}) \rightarrow 0$ as $\bar{\sigma} \rightarrow \infty$

各状況での振る舞い

	$\bar{\sigma}_t < \ell$	$\ell \leq \bar{\sigma}_t \leq u$	$u < \bar{\sigma}_t$
$p_{r=0}^{\text{succ}}(\bar{\sigma})$	≈ 0.5	positive	≈ 0
$\log \ m - x^*\ $	≈ 0	$\leq \log(1-r)$ w.p. $p_r^{\text{succ}}(\bar{\sigma})$	≈ 0
$\log^+ \left(\frac{\alpha_{\uparrow} \ell}{\bar{\sigma}} \right)$	$\begin{cases} -\log(\alpha_{\uparrow}) & \text{w.p. } \approx 0.5 \\ -\log(\alpha_{\downarrow}) & \text{w.p. } \approx 0.5 \end{cases}$	≈ 0	/
$\log^+ \left(\frac{\bar{\sigma}}{\alpha_{\downarrow} u} \right)$	/	≈ 0	$\log(\alpha_{\downarrow})$ w.p. ≈ 1

クリップされたドリフト $\mathbb{E}[\max\{V(\theta_{t+1}) - V(\theta_t), -A\} \mid \mathcal{F}_t] - V(\theta_t)$ はどのケースでも負

- $\bar{\sigma}_t < \ell$ の場合, $\log^+ \left(\frac{\alpha_{\uparrow} \ell}{\bar{\sigma}} \right)$ が減少 : $-\frac{1}{2} \log \left(\frac{\alpha_{\uparrow}}{\alpha_{\downarrow}} \right) < 0$
- $\ell \leq \bar{\sigma}_t \leq u$ の場合, $\log \|m - x^*\|$ が $-A$ よりも減少する確率が $p_{r=1-\exp(-A)}^{\text{succ}}(\bar{\sigma})$: ドリフト $\leq -A \min_{\ell \leq \bar{\sigma} \leq u} p_{r=1-\exp(-A)}^{\text{succ}}(\bar{\sigma})$
- $\bar{\sigma}_t > u$ の場合, $\log^+ \left(\frac{\bar{\sigma}}{\alpha_{\downarrow} u} \right)$ が減少 : $\log(\alpha_{\downarrow}) < 0$

(1+1)-ES の Spherical 関数のもとでの EFHT 上限

Theorem (Theorem 4.5 in Akimoto et al., 2022)

Let $T_\epsilon = \inf\{t : \|m_t - x^*\| \leq \epsilon\}$. Then,

$$\mathbb{E}[T_\epsilon] \leq C_T + C_R^{-1} \log \left(\frac{\|m_0 - x^*\|}{\epsilon} \right)$$

where

$$C_T = O \left(\max \left\{ 0, \log \left(\frac{\alpha_\uparrow \ell \|m_0 - x^*\|}{\sigma_0} \right), \log \left(\frac{\sigma_0}{\alpha_\downarrow u \|m_0 - x^*\|} \right) \right\} \right)$$

$$O \left(d \log \left(\frac{\|m_0 - x^*\|}{\epsilon} \right) \right) \text{ EFHT}$$

For $d \geq 2$, we can find coefficients of the potential function such that $C_R \in \Theta(1/d)$ as long as $\log(\alpha_\uparrow/\alpha_\downarrow) \in \omega(1/d)$.

まとめ : (1+1)-ES の Spherical 関数のもとでの EFHT

Spherical 関数 $f(x) = g(\|x - x^*\|)$ のもとでの最適解 x^* の ϵ 近傍までの FHT $T_\epsilon = \inf\{t : \|m_t - x^*\| \leq \epsilon\}$ の期待値は

$$\mathbb{E}[T_\epsilon] \in \Theta \left(d \log \left(\frac{\|m_0 - x^*\|}{\epsilon} \right) \right)$$

初期化の影響は $\mathbb{E}[T_\epsilon]$ のコンスタント c_T に以下のように表れる

$$c_T \in O \left(\max \left\{ 0, \log \left(\frac{\alpha_{\uparrow} \ell \|m_0 - x^*\|}{\sigma_0} \right), \log \left(\frac{\sigma_0}{\alpha_{\downarrow} u \|m_0 - x^*\|} \right) \right\} \right)$$

どんな σ 更新方法を用いたとしても,

$$\mathbb{E}[T_\epsilon] \in \Omega \left(d \log \left(\frac{\|m_0 - x^*\|}{\epsilon} \right) \right)$$

一般の関数クラスにおける $(1+1)$ -ES の一次収束

目的関数を単調変換に対して不変な指標に変換

Definition (Spatial Suboptimality Function)

For a measurable function $f: \mathbb{R}^d \rightarrow \mathbb{R}$, the spatial suboptimality function $f_\mu: \mathbb{R}^d \rightarrow [0, +\infty]$ is defined as (μ : Lebesgue measure)

$$f_\mu(x) = \sqrt[d]{\mu(f^{-1}((-\infty, f(x)]))} = \sqrt[d]{\mu(\{y \in \mathbb{R}^d \mid f(y) \leq f(x)\})}.$$

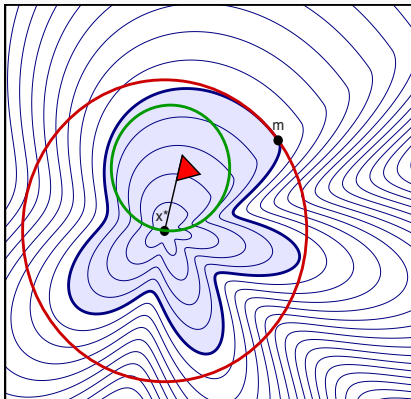
解釈:

- 目的関数を等高線内部の体積の d 乗根に変換
⇒ 同じ体積を持つ超球の半径（の定数倍）に変換
- 例えば, $f(x) = g(x^\top Hx)$ (g は任意の単調変換関数) は全て $\sqrt{x^\top Hx}$ (の定数倍) に変換

目的関数が満たすべき仮定 1 : 内包円と包含円

Assumption 1

There exist constants $C_u \geq C_\ell > 0$ such that for any $m \in \mathbb{R}^d$, an open ball $\mathcal{B}_\ell$ with radius $C_\ell f_\mu(m)$ and a closed ball $\bar{\mathcal{B}}_u$ with radius $C_u f_\mu(m)$ satisfy $\mathcal{B}_\ell \subseteq \{x \in \mathbb{R}^d \mid f_\mu(x) < f_\mu(m)\}$ and $\{x \in \mathbb{R}^d \mid f_\mu(x) \leq f_\mu(m)\} \subseteq \bar{\mathcal{B}}_u$.



Success Probability

Definition (Success Domain)

The r -success domain at m with $r \in [0, 1]$ is defined as

$$S_r(m) = \{x \in \mathbb{R}^d \mid f_\mu(x) \leq (1 - r)f_\mu(m)\} .$$

Definition (Success Probability)

The success probability with rate $r \in [0, 1]$ at $m \in S_0(m_0)$ under the normalized step-size $\bar{\sigma} = \sigma/f_\mu(m)$ is defined as

$$p_r^{\text{succ}}(\bar{\sigma}; m) = \Pr_{z \sim \mathcal{N}(0, I)} [m + f_\mu(m)\bar{\sigma}z \in S_r(m)] .$$

Note 1. $f(x) = g(\|x - x^*\|)$ の場合, これまでの定義と一致.

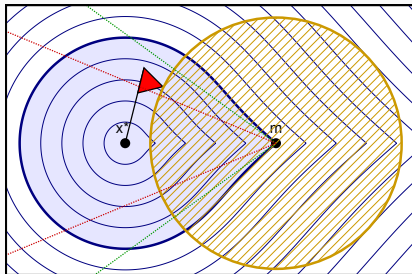
Note 2. Success Probability は一般の場合 m に依存.

Assumption 2 : p -improvability

Assumption 2

There exists p^{limit} such that for any $m \in \mathbb{R}^d$, the objective function f is p -improvable for some $p \geq p^{\text{limit}}$, i.e.,

$$\liminf_{\bar{\sigma} \downarrow 0} \inf_{m \in \mathbb{R}^d} p_0^{\text{succ}}(\bar{\sigma}; m) \geq p^{\text{limit}} .$$



例えば、微分可能な関数は $p^{\text{limit}} = \frac{1}{2}$ -improvable.

Assumptions を満たす関数の例 1 : L_ℓ -強凸かつ L_u -平滑な関数

Lemma

Let $f = g \circ h$ where g is a strictly increasing function and h is measurable, continuously differentiable with the unique critical point x^* , and quadratically bounded around x^* , i.e., for some $L_u \geq L_\ell > 0$,

$$(L_\ell/2)\|x - x^*\|^2 \leq h(x) - h(x^*) \leq (L_u/2)\|x - x^*\|^2 .$$

Then, Assumption 1 is satisfied with $C_\ell = \frac{1}{V_d} \sqrt{\frac{L_\ell}{L_u}}$ and $C_u = \frac{1}{V_d} \sqrt{\frac{L_u}{L_\ell}}$, where $V_d = \frac{\pi^{1/2}}{\Gamma^{1/d}(d/2+1)}$ is the d th root of the volume of the d -dimensional unit hyper-sphere

Lemma

Let $f = g \circ h$ where h is continuously differentiable and positively homogeneous with a unique optimum x^* , i.e., for some $\gamma > 0$

$$h(x^* + \gamma x) = h(x^*) + \gamma (h(x^* + x) - h(x^*)) \quad .$$

Then, Assumption 1 is satisfied with $C_u = \sup\{\|x - x^*\| : f_\mu(x) = 1\}$ and $C_\ell = \inf\{\|x - x^*\| : f_\mu(x) = 1\}$.

仮定が満たされると Success Probability のバウンドが得られる

Lemma (Proposition 4.1 in Akimoto et al., 2022)

Suppose that f satisfies Assumption 1. Then

$$\begin{aligned}\sup_m p_0^{\text{succ}}(\bar{\sigma}; m) &\leq \Phi\left(\bar{\mathcal{B}}\left(0, \frac{C_u}{\bar{\sigma}}\right)\right) \\ \inf_m p_0^{\text{succ}}(\bar{\sigma}; m) &\geq \Phi\left(\bar{\mathcal{B}}\left(\frac{(2C_u - C_\ell)}{\bar{\sigma}}e_1, \frac{C_\ell}{\bar{\sigma}}\right)\right),\end{aligned}$$

where $e_1 = (1, 0, \dots, 0)$ and Φ is the standard Gaussian measure.

加えて, Assumption 2 から $\liminf_{\bar{\sigma} \rightarrow 0} \inf_m p_0^{\text{succ}}(\bar{\sigma}; m) \geq p^{\text{limit}}$ なので,

- ・ $\bar{\sigma}$ が十分大きい : $p_0^{\text{succ}}(\bar{\sigma}; m) \approx 0$
- ・ $\bar{\sigma}$ が十分小さい : $p_0^{\text{succ}}(\bar{\sigma}; m) \gtrsim p^{\text{limit}}$

丁度良い Normalized Step-Size の範囲が存在

Lemma (Lemma 4.2 in Akimoto et al., 2022)

For any p_u and p_ℓ satisfying $0 < p_u < \frac{\log(1/\alpha_\downarrow)}{\log(\alpha_\uparrow/\alpha_\downarrow)} < p_\ell < p^{\text{limit}}$, the constants

$$\begin{aligned}\bar{\sigma}_\ell &= \sup \left\{ \bar{\sigma} > 0 : \inf_m p_0^{\text{succ}}(\bar{\sigma}; m) \geq p_\ell \right\} \\ \bar{\sigma}_u &= \inf \left\{ \bar{\sigma} > 0 : \sup_m p_0^{\text{succ}}(\bar{\sigma}; m) \leq p_u \right\}\end{aligned}$$

exist as positive finite values.

Let $\ell \leq \bar{\sigma}_\ell$ and $u \geq \bar{\sigma}_u$ such that $u/\ell \geq \alpha_\uparrow/\alpha_\downarrow$. Then, for $r \in [0, 1]$, p_r^* defined as

$$p_r^* := \inf_{\ell \leq \bar{\sigma} \leq u} \inf_m p_r^{\text{succ}}(\bar{\sigma}; m)$$

is lower bounded by a positive number defined by

$$\min_{\ell \leq \bar{\sigma} \leq u} \Phi \left(\mathcal{B} \left(\left(\frac{(2C_u - (1-r)C_\ell)}{\bar{\sigma}} \right) e_1, \frac{(1-r)C_\ell}{\bar{\sigma}} \right) \right) .$$

Assumptions 1 & 2 のもとでのポテンシャル関数

(1+1)-ES のポテンシャル関数

前スライドの条件を満たす u, ℓ について,

$$V(\theta) = \log(f_\mu(m)) + v \log^+ \left[\frac{\alpha_\uparrow \ell f_\mu(m)}{\sigma} \right] + v \log^+ \left[\frac{\sigma}{\alpha_\downarrow u f_\mu(m)} \right]$$

Spherical 関数のもとでのポテンシャル関数との違いは,

- ・ 進捗の指標が $\|m - x^*\|$ の代わりに $f_\mu(m)$

Note: $C_\ell f_\mu(x) \leq \|x - x^*\| \leq C_u f_\mu(x)$

- ・ Normalized Step-Size の定義が $\bar{\sigma} = \frac{d\sigma}{\|m - x^*\|}$ の代わりに $\bar{\sigma} = \frac{\sigma}{f_\mu(m)}$

各状況での振る舞い (Assumptions 1 & 2 の場合)

	$\bar{\sigma}_t < \ell$	$\ell \leq \bar{\sigma}_t \leq u$	$u < \bar{\sigma}_t$
$p_{r=0}^{\text{succ}}(\bar{\sigma}; m)$	$\geq p_\ell$	positive	$\leq p_u$
$\log f_\mu(m)$	≈ 0	$\leq \log(1-r)$ w.p. $p_r^{\text{succ}}(\bar{\sigma}; m) \geq p_r^*$	≈ 0
$\log^+ \left(\frac{\alpha_\uparrow \ell}{\bar{\sigma}} \right)$	$\begin{cases} -\log(\alpha_\uparrow) & \text{w.p. } \geq p_\ell \\ -\log(\alpha_\downarrow) & \text{w.p. } \leq 1 - p_\ell \end{cases}$	≈ 0	$/$
$\log^+ \left(\frac{\bar{\sigma}}{\alpha_\downarrow u} \right)$	$/$	≈ 0	$\begin{cases} -\log(\alpha_\downarrow) & \text{w.p. } \geq 1 - p_u \\ -\log(\alpha_\uparrow) & \text{w.p. } \leq p_u \end{cases}$

クリップされたドリフト $\mathbb{E}[\max\{V(\theta_{t+1}) - V(\theta_t), -A\} \mid \mathcal{F}_t] - V(\theta_t)$ は, Spherical 関数とほとんど同じ議論により, どのケースでも負

補足 : ペナルティ部分のドリフト

$$\leq -\log(\alpha_\uparrow)p_\ell - \log(\alpha_\downarrow)(1 - p_\ell) = -\log(\alpha_\downarrow) - p_\ell \log(\alpha_\uparrow/\alpha_\downarrow)$$

$$< -\log(\alpha_\downarrow) - \frac{\log(1/\alpha_\downarrow)}{\log(\alpha_\uparrow/\alpha_\downarrow)} \log(\alpha_\uparrow/\alpha_\downarrow) = 0$$

まとめ : Assumptions 1 & 2 のもとでの (1+1)-ES の EFHT 上限

Assumptions 1 & 2 のもとで (1+1)-ES は一次収束

Suppose that f satisfies Assumptions 1 & 2 with $p^{\text{limit}} > \frac{\log(1/\alpha_{\downarrow})}{\log(\alpha_{\uparrow}/\alpha_{\downarrow})}$.
Then, the EFHT of (1+1)-ES is

$$\mathbb{E}[T_{\epsilon}] \in \Omega \left(\log \left(\frac{\|m_0 - x^*\|}{\epsilon} \right) \right) .$$

Note 1. 成功確率のターゲット $\frac{\log(1/\alpha_{\downarrow})}{\log(\alpha_{\uparrow}/\alpha_{\downarrow})}$ は等高線が滑らかであるかぎり $p^{\text{limit}} = 1/2$ 未満なら良い. ($\alpha_{\downarrow}$ や $\alpha_{\uparrow}$ の調整が不要)

Note 2. 逆に, 等高線が滑らかでない場合, 一般に $p^{\text{limit}} < 1/2$. 例:
 $f(x) = \|x - x^*\|_{\infty} = \max_{i=1, \dots, d} |x_i|$ においては $p^{\text{limit}} = 2^{-d}$. ($\alpha_{\downarrow}$ や $\alpha_{\uparrow}$ の調整が必要不可欠)

Assumptions 1 & 2 のもとでの (1+1)-ES の EFHT 下限

下限は Hit-and-Run アルゴリズムを考えることで, Spherical 関数の場合と同じものが得られる.

Theorem (Theorem 4.7 in Akimoto et al., 2022)

For $d \geq 2$,

$$\mathbb{E}[T_{\beta}^x] \geq -\frac{1}{2} + \frac{d}{4} \log \left(\frac{\|m_0 - x^*\|}{\epsilon} \right) .$$

収束レートを示すにはよりタイトな上限の評価が必要

1. より強い仮定（目的関数を限定する）
2. 限定された目的関数でのより良いポテンシャル関数の設計
3. ドリフトのよりタイトな評価

仮定：強凸性とリプシッツ平滑性

Assumption 3: Strongly Convexity and Lipschitz Smoothness

A objective function is $f(x) = g(h(x))$, where g is strictly increasing and h is measurable, continuously differentiable and L -strongly convex and U -Lipschitz smooth, i.e.,

$$\frac{L}{2}\|x - y\|^2 \leq h(y) - h(x) - \langle y - x, \nabla h(x) \rangle \leq \frac{U}{2}\|x - y\|^2 .$$

The unique optimum is located at x^* .

強凸かつリプシッツ平滑な関数のもとでの (1+1)-ES のポテンシャル関数

$$V(\theta) = \log(h(m)) \\ + v \cdot \log^+ \left(\frac{s \cdot \sqrt{Lh(m)}}{\sigma E_Q} \right) + v \cdot \log^+ \left(\frac{\sigma E_Q}{\ell \cdot \sqrt{Lh(m)}} \right)$$

where $E_Q = \sup_{m, \sigma} \mathbb{E}[Q_Z]$ and

$$Q_Z = \frac{2}{\sigma^2} \mathbb{E}[h(m + \sigma z) - h(m) - \langle \nabla h(m), \sigma z \rangle]$$

タイトなドリフトのバウンドを得るために, 以下のように条件分岐

- $\sigma < \frac{s \cdot \sqrt{Lh(m)}}{E_Q};$
- $\sigma > \frac{\ell \cdot \sqrt{Lh(m)}}{\sqrt{2\alpha_{\downarrow}} \mathbb{E}[Q_Z]};$
- Otherwise.

(1+1)-ES の収束レート

Theorem

If $\alpha_{\uparrow}$ and $\alpha_{\downarrow}$ are chosen so that $\frac{\log(1/\alpha_{\downarrow})}{\log(\alpha_{\uparrow}/\alpha_{\downarrow})} > \Phi\left(-\sqrt{\frac{2}{\pi}}\right) \approx 0.212$, where Φ is the standard Gaussian CDF, then

$$\frac{1}{d} \geq \text{CR} \in \Omega_{d \rightarrow \infty} \left(\min \left\{ \frac{L}{E_Q}, \log \left(\frac{\alpha_{\uparrow}}{\alpha_{\downarrow}} \right) \right\} \right)$$

具体例

- 一般の L -強凸かつ U -平滑な関数の場合: $\frac{L}{E_Q} = \frac{L}{dU}$
 d に反比例, 条件数 $\frac{U}{L}$ に反比例
- 凸二次関数 $\frac{1}{2}x^T Hx$ の場合: $\frac{L}{E_Q} = \frac{L}{\text{Tr}(H)}$
同じ条件数であっても, レートが $\frac{L}{dU-L}$ から $\frac{1}{d+(U/L-1)}$ まであり得る

凸二次関数における実験値との比較

三つの同じ条件数を持つ凸二次関数での収束レート（実験値）の比較

- $H_1 = \text{diag}(1, 10^\kappa, \dots, 10^\kappa)$
- $H_2 = \text{diag}\left(10^{\kappa \cdot \frac{0}{d-1}}, 10^{\kappa \cdot \frac{1}{d-1}}, \dots, 10^{\kappa \cdot \frac{d-1}{d-1}}\right)$
- $H_3 = \text{diag}(1, \dots, 1, 10^\kappa)$

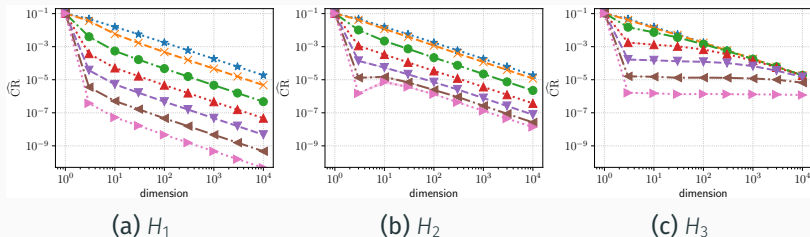


Figure 1: Results on convex quadratic functions with Hessian matrix H_1 , H_2 , and H_3 . Average and standard error of $\widehat{CR}$ with varying d and varying κ : ($\star$) 0, ($\times$) 1, ($\circ$) 2, ($\triangle$) 3, (∇) 4, ($\triangleleft$) 5, ($\triangleright$) 6. $\alpha_\uparrow = \exp(1/\sqrt{d})$ and $\alpha_\downarrow = \alpha_\uparrow^{-\frac{1}{4}}$.

まとめ：一般の場合の (1+1)-ES の一次収束

- Assumption 1 と 2 を満たす関数では, $p^{\text{limit}} > \frac{\log(1/\alpha_{\downarrow})}{\log(\alpha_{\uparrow}/\alpha_{\downarrow})}$ の場合に一次収束
- 等高線が滑らかな関数では $p^{\text{limit}} = 1/2$, 一般には $p^{\text{limit}} < 1/2$.
 $p^{\text{limit}} = 0$ なら条件を満たせない (実際, Assumption 2 は最適解への収束のための必要条件 (Glasmachers, 2020))
- 強凸かつリプシッツ平滑な関数 (とその単調変換) のもとでの収束レートは $O\left(\frac{L}{dU}\right)$
- 凸二次関数の場合, 収束レートは $O\left(\frac{L}{\text{Tr}(H)}\right)$

References



Akimoto, Y. and N. Hansen (Sept. 2020). “Diagonal Acceleration for Covariance Matrix Adaptation Evolution Strategies”. In: *Evolutionary Computation* 28.3, pp. 405–435. DOI: [10.1162/evco_a_00260](https://doi.org/10.1162/evco_a_00260).



Akimoto, Youhei, Anne Auger, and Tobias Glasmachers (2018). “Drift theory in continuous search spaces: expected hitting time of the $(1+1)$ -ES with $1/5$ success rule”. In: *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 801–808.



Akimoto, Youhei et al. (Jan. 2022). “Global Linear Convergence of Evolution Strategies on More than Smooth Strongly Convex Functions”. In: *SIAM J. on Optimization* 32.2, pp. 1402–1429. DOI: [10.1137/20M1373815](https://doi.org/10.1137/20M1373815).



Glasmachers, Tobias (Mar. 2020). “Global Convergence of the $(1+1)$ Evolution Strategy to a Critical Point”. In: *Evolutionary Computation* 28.1, pp. 27–53. DOI: [10.1162/evco_a_00248](https://doi.org/10.1162/evco_a_00248).



Hajek, Bruce (1982). “Hitting-time and occupation-time bounds implied by drift analysis with applications”. In: *Advances in Applied probability* 14.3, pp. 502–525.



Hansen, N. and A. Ostermeier (2001). “Completely derandomized self-adaptation in evolution strategies”. In: *Evolutionary Computation* 9.2, pp. 159–195.



Hansen, Nikolaus and Stefan Kern (2004). “Evaluating the CMA Evolution Strategy on Multimodal Test Functions”. In: *Parallel Problem Solving from Nature - PPSN VIII*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 282–291.



Hansen, Nikolaus, Sibylle D. Müller, and Petros Koumoutsakos (Mar. 2003). “Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES)”. In: *Evolutionary Computation* 11.1, pp. 1–18. ISSN: 1063-6560. DOI: [10.1162/106365603321828970](https://doi.org/10.1162/106365603321828970).



Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan Press. ISBN: 9780472084609. URL: <https://books.google.co.jp/books?id=JE5RAAAAMAAJ>.



Jastrebski, G.A. and D.V. Arnold (2006). “Improving Evolution Strategies through Active Covariance Matrix Adaptation”. In: *2006*

IEEE International Conference on Evolutionary Computation, pp. 2814–2821. DOI: [10.1109/CEC.2006.1688662](https://doi.org/10.1109/CEC.2006.1688662).



Lengler, Johannes and Angelika Steger (2016). *Drift Analysis and Evolutionary Algorithms Revisited*. Tech. rep. arXiv:1608.03226.



Ollivier, Yann et al. (2017). “Information-geometric optimization algorithms: A unifying picture via invariance principles”. In: *The Journal of Machine Learning Research* 18.1, pp. 564–628.



Rechenberg, Ingo (1973). *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog.