



HAL
open science

Supplementary Material for Learning Rate Adaptation by Line Search in Evolution Strategies with Recombination

Armand Gissler, Anne Auger, Nikolaus Hansen

► **To cite this version:**

Armand Gissler, Anne Auger, Nikolaus Hansen. Supplementary Material for Learning Rate Adaptation by Line Search in Evolution Strategies with Recombination. 2022. hal-03626292

HAL Id: hal-03626292

<https://inria.hal.science/hal-03626292>

Submitted on 11 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Supplementary Material for Learning Rate Adaptation by Line Search in Evolution Strategies with Recombination

Armand Gissler

forename.lastname@polytechnique.edu
Inria and CMAP, Ecole Polytechnique,
IP Paris, CNRS
Palaiseau, France

Anne Auger

forename.lastname@inria.fr
Inria and CMAP, Ecole Polytechnique,
IP Paris, CNRS
Palaiseau, France

Nikolaus Hansen

forename.lastname@inria.fr
Inria and CMAP, Ecole Polytechnique,
IP Paris, CNRS
Palaiseau, France

ABSTRACT

This is the supplementary material (proofs) for [1].

PROOF OF LEMMA 2.3 IN [1]. Let $t \in \mathbb{N}$. We know that $X_{t+1} = X_t + \kappa_{t+1} \sum_{i=1}^{\mu} w_i U_{t+1}^{\varphi(i)}$, with $\kappa_{t+1} = \bar{\kappa}(X_t, \alpha \|X_t\| \sum_{i=1}^{\mu} w_i U_{t+1}^{\varphi(i)})$. Then, by (A1)

$$Z_t = \frac{\|X_{t+1}\|}{\|X_t\|} = \left\| \frac{X_t}{\|X_t\|} + \alpha \bar{\kappa} \left(\frac{X_t}{\|X_t\|}, \alpha S_{U_{t+1}}^{\varphi} \right) S_{U_{t+1}}^{\varphi} \right\|.$$

But then, according to (A2), we have that Z_t has the same distribution than $\|e_1 + \alpha \bar{\kappa}(e_1, \alpha S_V^{\varphi}) S_V^{\varphi}\|$. Thus, all the $Z_t, t \in \mathbb{N}$ are identically distributed.

Let $s \in \mathbb{R}$, and denote $\psi^s: \mathbb{R}^n \rightarrow \mathbb{C}$ the measurable function such that $\psi^s(X_t) = \mathbb{E}[\exp is Z_t \mid X_t]$. Then, for $x \in \mathbb{R}$, according to (A2) applied to $X = x$

$$\begin{aligned} \psi^s(x) &= \mathbb{E} \exp is \left\| \frac{x}{\|x\|} + \alpha \bar{\kappa} \left(\frac{x}{\|x\|}, \alpha S_{U_{t+1}}^{\varphi} \right) S_{U_{t+1}}^{\varphi} \right\| \\ &= \psi^s(e_1). \end{aligned}$$

Thus, $\mathbb{E}[\exp is Z_t \mid X_t] = \psi^s(X_t) = \psi^s(e_1)$. Define \mathcal{F}_t the filtration induced by X_0, \dots, X_t . We get then

$$\begin{aligned} &\mathbb{E}[\exp i(s_0 Z_0 + s_1 Z_1 + \dots + s_t Z_t)] \\ &= \mathbb{E}[\mathbb{E}[\exp i(s_0 Z_0 + s_1 Z_1 + \dots + s_t Z_t) \mid \mathcal{F}_t]] \\ &= \mathbb{E}[\exp is(s_0 Z_0 + s_1 Z_1 + \dots + s_{t-1} Z_{t-1}) \times \mathbb{E}[\exp is_t Z_t \mid X_t]] \\ &= \mathbb{E}[\exp i(s_0 Z_0 + s_1 Z_1 + \dots + s_{t-1} Z_{t-1})] \times \psi^{s_t}(e_1) \\ &= \mathbb{E}[\exp i(s_0 Z_0 + s_1 Z_1 + \dots + s_{t-1} Z_{t-1})] \times \mathbb{E}[\exp is_t Z_t] \end{aligned}$$

Hence, Z_t is independent from $(Z_0, Z_1, \dots, Z_{t-1})$. By induction, we get that (Z_0, Z_1, \dots, Z_t) is mutually independent. \square

PROOF OF LEMMA 3.2 IN [1]. Assumption (A1) holds by [1, Lemma 2.1].

Let R be a rotation matrix. Then $\bar{\kappa}_{\text{PLS}}(Rx, Rv) = \arg \min_{\kappa \geq 0} \|Rx + \kappa Rv\| = \arg \min_{\kappa \geq 0} \|R(x + \kappa v)\| = \arg \min_{\kappa \geq 0} \|x + \kappa v\| = \bar{\kappa}_{\text{PLS}}(x, v)$. Hence $\bar{\kappa}_{\text{PLS}}$ is rotation-invariant and [1, Lemma 2.2] implies that (A2) holds.

Now, let us show that (A4) is satisfied. Let $U^1, \dots, U^\mu \in \mathbb{R}^n$ be i.i.d. infinite dimensional standard Gaussian vectors. Then, when $n \rightarrow \infty$, the following limit holds a.s. by the strong LLN

$$\frac{1}{n} \left\| \sum_{i=1}^{\mu} w_i [U^i]_{\leq n} \right\|^2 = \frac{1}{n} \sum_{k=1}^n \left| \sum_{i=1}^{\mu} w_i [U^i]_k \right|^2 \rightarrow \mu_w^{-1},$$

thus by [1, Lemma 3.1]

$$\begin{aligned} \bar{\kappa}_{\text{PLS}} \left(e_1, \frac{\alpha}{n} \sum_{i=1}^{\mu} w_i [U^i]_{\leq n} \right) &= -1_{\sum_{i=1}^{\mu} w_i [U^i]_1 < 0} \frac{\sum_{i=1}^{\mu} w_i [U^i]_1}{\frac{\alpha}{n} \left\| \sum_{i=1}^{\mu} w_i [U^i]_{\leq n} \right\|^2} \\ &\rightarrow -1_{\sum_{i=1}^{\mu} w_i [U^i]_1 < 0} \alpha^{-1} \mu_w \sum_{i=1}^{\mu} w_i [U^i]_1. \end{aligned}$$

Thus (A4) holds. \square

PROOF OF LEMMA 4.2 IN [1]. First, let us prove the scaling-invariance condition (A1) for $\bar{\kappa}_{\text{DLS}}^{\varepsilon, \beta}$. Consider $\bar{\kappa}_{\text{DLS}}^{\varepsilon, \beta}(x/r, v/r, \kappa_{\text{init}})$. Then, as in line 4 of [1, Algorithm 2], the condition $f(x + \kappa^0 v) < f(x + \kappa^1 v)$ is equivalent to $f(x/r + \kappa^0 v/r) < f(x/r + \kappa^1 v/r)$, as f is the sphere function is scaling-invariant, then $\bar{\kappa}_{\text{DLS}}^{\varepsilon, \beta}(x/r, v/r, \kappa_{\text{init}}) = \bar{\kappa}_{\text{DLS}}^{\varepsilon, \beta}(x, v, \kappa_{\text{init}})$. Thus, (A1) holds.

The function $\bar{\kappa}_{\text{DLS}}^{\varepsilon, \beta}$ is rotation-invariant. Indeed if R is a rotation matrix, then, as in line 4 of [1, Algorithm 2], the condition $f(x + \kappa^0 v) < f(x + \kappa^1 v)$ is equivalent to $f(Rx + \kappa^0 Rv) < f(Rx + \kappa^1 Rv)$, as f is the sphere function is invariant by rotation, then this implies that $\bar{\kappa}_{\text{DLS}}^{\varepsilon, \beta}(Rx, Rv, \kappa_{\text{init}}) = \bar{\kappa}_{\text{DLS}}^{\varepsilon, \beta}(x, v, \kappa_{\text{init}})$. Thus, by [1, Lemma 2.2], (A2) holds.

We prove now that (A4) is satisfied. Consider $U^1, \dots, U^\mu \in \mathbb{R}^n$ i.i.d. infinite dimensional standard Gaussian vectors, and denote $S_U^n = \sum_{i=1}^{\mu} w_i [U^i]_{\leq n}$.

Consider the line search obtained with $\bar{\kappa}_{\text{DLS}}^{\varepsilon, \beta} \left(e_1, \frac{\alpha}{n} S_U^n, \kappa_{\text{init}} \right)$. We denote $(\kappa^{i,0}, \kappa^{i,1})_{i=0, \dots, C(\beta, \varepsilon) - 2}$ the value of κ^1 and κ^0 over the iterations of this line search.

We prove now by induction that for all $i = 0, 1, \dots$, the following limits when $n \rightarrow \infty$ hold a.s. $\kappa^{i,0} \rightarrow \kappa^{\infty, i, 0}$ and $\kappa^{i,1} \rightarrow \kappa^{\infty, i, 1}$, where $\kappa^{\infty, i, 0}, \kappa^{\infty, i, 1}$ are given in the line 5 and line 7 of [1, Algorithm 2] over iteration initialized with parameters $X = \alpha^{-1} \mu_w S_U^1 \in \mathbb{R}^1, v = 1 \in \mathbb{R}^1$, and given $\kappa_{\text{init}}, \varepsilon, \beta$. At iteration $i = 0$, this is trivially true as $\kappa^{i,0} = \kappa^{\infty, i, 0} = \kappa_{\text{init}}/2$, and $\kappa^{i,1} = \kappa^{\infty, i, 1} = 2\kappa_{\text{init}}$.

Now consider an arbitrary step i of the line search, and assume that $\kappa^{i,0} \rightarrow \kappa^{\infty, i, 0}$ and $\kappa^{i,1} \rightarrow \kappa^{\infty, i, 1}$. Then, $\kappa^{i+1,0} = \kappa^{i,0} + (1 - \beta) (\kappa^{i,1} - \kappa^{i,0}) \mathbf{1}_{\{h_{\alpha/n}(\kappa^{i,1} S_U^n) < h_{\alpha/n}(\kappa^{i,0} S_U^n)\}}$, where

$$h_{\alpha/n}(\kappa^{i,j} S_U^n) = 2 \underbrace{\kappa^{i,j}}_{\rightarrow \kappa^{\infty, i, j}} S_U^1 + \alpha \left(\kappa^{i,j} \right)^2 \underbrace{\frac{\|S_U^n\|^2}{n}}_{\rightarrow \sum_{i=1}^{\mu} w_i^2}$$

Hence, a.s. $h_{\alpha/n}(\kappa^{i,j} S_U^n)$ tends to $2\kappa^{\infty,i,j} S_U^1 + \alpha (\kappa^{\infty,i,j})^2 \mu_w^{-1}$ when $n \rightarrow \infty$, so that

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{1}_{\{h_{\alpha/n}(\kappa^{i,1} S_U^n) < h_{\alpha/n}(\kappa^{i,0} S_U^n)\}} \\ = \mathbf{1}_{\{(\kappa^{\infty,i,1} - \kappa^{\infty,i,0}) (2S_U^n + (\kappa^{\infty,i,0} + \kappa^{\infty,i,1}) \alpha \sum_{i=1}^{\mu} w_i^2) > 0\}}, \end{aligned}$$

hence $\kappa^{i+1,0} \rightarrow \kappa^{\infty,i+1,0}$. Similarly, $\lim_{n \rightarrow \infty} \kappa^{i+1,1} = \kappa^{\infty,i+1,1}$.

In the end, by induction, we get that

$$\lim_{n \rightarrow \infty} \bar{\kappa} \left(e_1, \frac{\alpha}{n} S_U^n, \kappa_{\text{init}} \right) = \bar{\kappa}^{\infty} \left(\left([U^i]_1 \right)_{i=1, \dots, \lambda}, \kappa_{\text{init}} \right).$$

where $\bar{\kappa}^{\infty}$ is defined in [1, Lemma 4.2]. Hence (A4) holds. \square

LEMMA A.1. For $n \in \mathbb{N}^*$, let U_1^n, \dots, U_n^n be i.i.d. standard multi-variate normal distribution of dimension n . Suppose that the functions $\bar{\kappa}_n: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+$ are either all equal to $\bar{\kappa}_{\text{PLS}}$ or all upperbounded by a positive constant κ^M . Then,

$$\left(n \ln \left\| e_1 + \bar{\kappa}_n \left(e_1, \alpha/n \sum_{i=1}^{\mu} w_i U_n^i \right) \alpha/n \sum_{i=1}^n w_i U_n^i \right\|^2 \right)_{n \in \mathbb{N}^*}$$

is uniformly integrable.

PROOF. Suppose first that for all $n \in \mathbb{N}^*$, $\bar{\kappa}_n = \bar{\kappa}_{\text{PLS}}$. Note then that, as in the proof of [1, Theorem 3.3], we have

$$\begin{aligned} n \ln \left\| e_1 + \bar{\kappa}_n \left(e_1, \alpha/n \sum_{i=1}^{\mu} w_i U_n^i \right) \alpha/n \sum_{i=1}^n w_i U_n^i \right\|^2 \\ = -8 \ln \left[\left(1 - \frac{[\sum_{i=1}^{\mu} w_i U_n^i]_1^2}{\|\sum_{i=1}^{\mu} w_i U_n^i\|^2} \right)^{-n/8} \right] \mathbf{1}_{[\sum_{i=1}^{\mu} w_i U_n^i]_1 < 0}. \end{aligned}$$

But it is proven in [2, Proof of Proposition 4, page 23] that

$$\left(\ln \left[\left(1 - \frac{[\sum_{i=1}^{\mu} w_i U_n^i]_1^2}{\|\sum_{i=1}^{\mu} w_i U_n^i\|^2} \right)^{-n/8} \right] \mathbf{1}_{[\sum_{i=1}^{\mu} w_i U_n^i]_1 < 0} \right)_{n \in \mathbb{N}^*}$$

is uniformly integrable.

Suppose now that there exists $\kappa^M > 0$ such that for all $n \in \mathbb{N}^*$, $\bar{\kappa}_n \leq \kappa^M$. We will prove the uniform integrability of the positive part and of the negative part of

$$\left(n \ln \left\| e_1 + \bar{\kappa}_n \left(e_1, \alpha/n \sum_{i=1}^{\mu} w_i U_n^i \right) \alpha/n \sum_{i=1}^n w_i U_n^i \right\|^2 \right)_{n \in \mathbb{N}^*}.$$

For the negative part, note that, by definition of $\bar{\kappa}_{\text{PLS}}$, then

$$\begin{aligned} \left\| e_1 + \bar{\kappa}_n \left(e_1, \alpha/n \sum_{i=1}^{\mu} w_i U_n^i \right) \alpha/n \sum_{i=1}^n w_i U_n^i \right\|^2 \\ \geq \left\| e_1 + \bar{\kappa}_{\text{PLS}} \left(e_1, \alpha/n \sum_{i=1}^{\mu} w_i U_n^i \right) \alpha/n \sum_{i=1}^n w_i U_n^i \right\|^2. \end{aligned}$$

In addition, the negative part of the logarithm \ln^- is a decreasing function, thus

$$\begin{aligned} n \ln^- \left\| e_1 + \bar{\kappa}_n \left(e_1, \alpha/n \sum_{i=1}^{\mu} w_i U_n^i \right) \alpha/n \sum_{i=1}^n w_i U_n^i \right\|^2 \\ \leq n \ln^- \left\| e_1 + \bar{\kappa}_{\text{PLS}} \left(e_1, \alpha/n \sum_{i=1}^{\mu} w_i U_n^i \right) \alpha/n \sum_{i=1}^n w_i U_n^i \right\|^2. \end{aligned} \quad (1)$$

Note that, as $\|e_1 + \bar{\kappa}_{\text{PLS}}(e_1, \alpha/n \sum_{i=1}^{\mu} w_i U_n^i) \alpha/n \sum_{i=1}^n w_i U_n^i\| \leq \|e_1 + 0\| = 1$, then $\ln \|e_1 + \bar{\kappa}_{\text{PLS}}(e_1, \alpha/n \sum_{i=1}^{\mu} w_i U_n^i) \alpha/n \sum_{i=1}^n w_i U_n^i\| \leq \|e_1 + 0\| \leq 0$, thus is equal in absolute value to its negative part.

However, as

$$\left(n \ln^- \left\| e_1 + \bar{\kappa}_{\text{PLS}} \left(e_1, \alpha/n \sum_{i=1}^{\mu} w_i U_n^i \right) \alpha/n \sum_{i=1}^n w_i U_n^i \right\|^2 \right)_{n \in \mathbb{N}^*}$$

is uniformly integrable (since it is equal to the sequence with \ln instead of \ln^- which is uniformly integrable as seen above), then according to Eq. (1) so is

$$\left(n \ln^- \left\| e_1 + \bar{\kappa}_n \left(e_1, \alpha/n \sum_{i=1}^{\mu} w_i U_n^i \right) \alpha/n \sum_{i=1}^n w_i U_n^i \right\|^2 \right)_{n \in \mathbb{N}^*}.$$

We claim that for all $x \in \mathbb{R}^n$, and $0 \leq \kappa \leq \kappa^M$, then

$$\ln^+ \|e_1 + \kappa x\| \leq \ln^+ \|e_1 + \kappa^M x\|.$$

Indeed, the above equation is equivalent to

$$\|e_1 + \kappa x\|^2 \mathbf{1}_{\|e_1 + \kappa x\| \geq 1} \leq \|e_1 + \kappa^M x\|^2 \mathbf{1}_{\|e_1 + \kappa^M x\| \geq 1}.$$

However, the derivative of the function $\kappa \mapsto \|e_1 + \kappa x\|^2$ is equal to $\kappa \mapsto 2[x]_1 + \kappa \langle x, x \rangle$, hence is nonnegative for any $\kappa \geq 0$ that also satisfies that $\|e_1 + \kappa x\|^2 = 1 + \kappa(2[x]_1 + \kappa \langle x, x \rangle)$ is greater than or equal to 1. Thus, the above condition is satisfied.

For the positive part, we have then

$$\begin{aligned} n \ln^+ \left\| e_1 + \bar{\kappa}_n \left(e_1, \alpha/n \sum_{i=1}^{\mu} w_i U_n^i \right) \alpha/n \sum_{i=1}^n w_i U_n^i \right\|^2 \\ \leq n \ln^+ \left\| e_1 + \frac{\kappa^M \alpha}{n} \sum_{i=1}^{\mu} w_i U_n^i \right\|^2. \end{aligned}$$

But it is proven in [2]¹ that the RHS of the above equation is uniformly integrable. All in all, we get that

$$\left(n \ln \left\| e_1 + \bar{\kappa}_n \left(e_1, \alpha/n \sum_{i=1}^{\mu} w_i U_n^i \right) \alpha/n \sum_{i=1}^n w_i U_n^i \right\|^2 \right)_{n \in \mathbb{N}^*}$$

is uniformly integrable. \square

REFERENCES

- [1] Armand Gissler, Anne Auger, and Nikolaus Hansen. Learning rate adaptation by line search in evolution strategies with recombination. In *Genetic and Evolutionary Computation Conference (GECCO 2022)*, Boston, United States, July 2022.
- [2] Mohamed Jebalia and Anne Auger. Log-linear convergence of the scale-invariant $(\mu/\mu_w, \lambda)$ -ES and optimal μ for intermediate recombination for large population sizes. In *International Conference on Parallel Problem Solving from Nature*, pages 52–62. Springer, 2010.

¹replace σ^* by $\alpha \kappa^M$ in the proof of Proposition 4 of [2]