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

Anne Auger<br>forename.lastname@inria.fr<br>Nikolaus Hansen<br>forename.lastname@inria.fr<br>Inria and CMAP, Ecole Polytechnique, Inria and CMAP, Ecole Polytechnique,<br>IP Paris, CNRS<br>Palaiseau, France<br>IP Paris, CNRS<br>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 $\kappa_{t+1}=\bar{\kappa}\left(X_{t}, \alpha\left\|X_{t}\right\| \sum_{i=1}^{\mu} w_{i} U_{t+1}^{\varphi(i)}\right)$. Then, by (A1)

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

But then, according to (A2), we have that $Z_{t}$ has the same distribution than $\left\|e_{1}+\alpha \bar{\kappa}\left(e_{1}, \alpha S_{V}^{\varphi}\right) S_{V}^{\varphi}\right\|$. 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 $\psi^{s}\left(X_{t}\right)=\mathbb{E}\left[\exp i s Z_{t} \mid X_{t}\right]$. Then, for $x \in \mathbb{R}$, according to (A2) applied to $X=x$

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

Thus, $\mathbb{E}\left[\exp\right.$ is $\left.Z_{t} \mid X_{t}\right]=\psi^{s}\left(X_{t}\right)=\psi^{s}\left(e_{1}\right)$. Define $\mathcal{F}_{t}$ the filtration induced by $X_{0}, \ldots, X_{t}$. We get then

$$
\begin{aligned}
\mathbb{E} & {\left[\exp i\left(s_{0} Z_{0}+s_{1} Z_{1}+\cdots+s_{t} Z_{t}\right)\right] } \\
& =\mathbb{E}\left[\mathbb{E}\left[\exp i\left(s_{0} Z_{0}+s_{1} Z_{1}+\cdots+s_{t} Z_{t}\right) \mid \mathcal{F}_{t}\right]\right] \\
& =\mathbb{E}\left[\exp i s\left(s_{0} Z_{0}+s_{1} Z_{1}+\cdots+s_{t-1} Z_{t-1}\right) \times \mathbb{E}\left[\exp i s_{t} Z_{t} \mid X_{t}\right]\right] \\
& =\mathbb{E}\left[\exp i\left(s_{0} Z_{0}+s_{1} Z_{1}+\cdots+s_{t-1} Z_{t-1}\right)\right] \times \psi^{s_{t}}\left(e_{1}\right) \\
& =\mathbb{E}\left[\exp i\left(s_{0} Z_{0}+s_{1} Z_{1}+\cdots+s_{t-1} Z_{t-1}\right)\right] \times \mathbb{E}\left[\exp i s_{t} Z_{t}\right]
\end{aligned}
$$

Hence, $Z_{t}$ is independent from $\left(Z_{0}, Z_{1}, \ldots, Z_{t-1}\right)$. By induction, we get that $\left(Z_{0}, Z_{1}, \ldots, Z_{t}\right)$ is mutually independent.

Proof of Lemma 3.2 in [1]. Assumption (A1) holds by [1, Lemma 2.1].

Let $R$ be a rotation matrix. Then $\bar{\kappa}_{\text {PLS }}(R x, R v)=\arg \min _{\kappa \geqslant 0} \| R x+$ $\kappa R v\left\|=\arg \min _{\kappa \geqslant 0}\right\| R(x+\kappa v)\left\|=\arg \min _{\kappa \geqslant 0}\right\| x+\kappa v \|=\bar{\kappa}_{\text {PLS }}(x, v)$. 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 $U^{1}, \ldots, U^{\lambda} \in \mathbb{R}^{\mathbb{N}}$ be i.i.d. infinite dimensional standard Gaussian vectors. Then, when $n \rightarrow \infty$, the following limit holds a.s. by the strong LLN

$$
\frac{1}{n}\left\|\sum_{i=1}^{\mu} w_{i}\left[U^{i}\right]_{\leqslant n}\right\|^{2}=\frac{1}{n} \sum_{k=1}^{n}\left|\sum_{i=1}^{\mu} w_{i}\left[U^{i}\right]_{k}\right|^{2} \rightarrow \mu_{\mathrm{w}}^{-1}
$$

thus by [1, Lemma 3.1]

$$
\begin{aligned}
& \bar{\kappa}_{\mathrm{PLS}}\left(e_{1}, \frac{\alpha}{n} \sum_{i=1}^{\mu} w_{i}\left[U^{i}\right]_{\leqslant n}\right)=-1 \sum_{i=1}^{\mu} w_{i}\left[U^{i}\right]_{1}<0 \frac{\sum_{i=1}^{\mu} w_{i}\left[U^{i}\right]_{1}}{\frac{\alpha}{n}\left\|\sum_{i=1}^{\mu} w_{i}\left[U^{i}\right]_{\leqslant n}\right\|^{2}} \\
& \rightarrow \mathbf{1}_{\sum_{i=1}^{\mu} w_{i}\left[U^{i}\right]_{1}<0} \alpha^{-1} \mu_{\mathrm{w}} \sum_{i=1}^{\mu} w_{i}\left[U^{i}\right]_{1} .
\end{aligned}
$$

Thus (A4) holds.

Proof of Lemma 4.2 in [1]. First, let us prove the scaling-invariance condition (A1) for $\bar{\kappa}_{\text {DLS }}^{\varepsilon, \beta}$. Consider $\bar{\kappa}_{\text {DLS }}^{\varepsilon, \beta}\left(x / r, v / r, \kappa_{\text {init }}\right)$. Then, as in line 4 of [1, Algorithm 2], the condition $f\left(x+\kappa^{0} v\right)<f\left(x+\kappa^{1} v\right)$ is equivalent to $f\left(x / r+\kappa^{0} v / r\right)<f\left(x / r+\kappa^{1} v / r\right)$, as $f$ is the sphere function is scaling-invariant, then $\bar{\kappa}_{\text {DLS }}^{\varepsilon, \beta}\left(x / r, v / r, \kappa_{\text {init }}\right)=$ $\bar{\kappa}_{\text {DLS }}^{\varepsilon, \beta}\left(x, v, \kappa_{\text {init }}\right)$. 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 $f(x+$ $\left.\kappa^{0} v\right)<f\left(x+\kappa^{1} v\right)$ is equivalent to $f\left(R x+\kappa^{0} R v\right)<f\left(R x+\kappa^{1} R v\right)$, as $f$ is the sphere function is invariant by rotation, then this implies that $\bar{\kappa}_{\text {DLS }}^{\varepsilon, \beta}\left(R x, R v, \kappa_{\text {init }}\right)=\bar{\kappa}_{\text {DLS }}^{\varepsilon, \beta}\left(x, v, \kappa_{\text {init }}\right)$. Thus, by [1, Lemma 2.2], (A2) holds.
We prove now that (A4) is satisfied. Consider $U^{1}, \ldots, U^{\mu} \in \mathbb{R}^{\mathbb{N}}$ $\mu$ i.i.d. infinite dimensional standard Gaussian vectors, and denote $S_{U}^{n}=\sum_{i=1}^{\mu} w_{i}\left[U^{i}\right] \leqslant n$.

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 $\left(\kappa^{i, 0}, \kappa^{i, 1}\right)_{i=0, \cdots, C(\beta, \varepsilon)-2}$ the value of $\kappa^{1}$ and $\kappa^{0}$ over the iterations of this line search.

We prove now by induction that for all $i=0,1, \ldots$, 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 $X=\alpha^{-1} \mu_{\mathrm{w}} S_{U}^{1} \in \mathbb{R}^{1}, v=$ $1 \in \mathbb{R}^{1}$, 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, $\kappa^{i+1,0}=\kappa^{i, 0}+(1-$ $\beta)\left(\kappa^{i, 1}-\kappa^{i, 0}\right) 1_{\left\{h_{\alpha / n}\left(\kappa^{i, 1} S_{U}^{n}\right)<h_{\alpha / n}\left(\kappa^{i, 0} S_{U}^{n}\right)\right\}}$, where

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

Hence, a.s. $h_{\alpha / n}\left(\kappa^{i, j} S_{U}^{n}\right)$ tends to $2 \kappa^{\infty, i, j} S_{U}^{1}+\alpha\left(\kappa^{\infty, i, j}\right)^{2} \mu_{\mathrm{w}}^{-1}$ when $n \rightarrow \infty$, so that

$$
\begin{aligned}
\lim _{n \rightarrow \infty} 1_{\left\{h_{\alpha / n}\left(\kappa^{i, 1} S_{U}^{n}\right)<h_{\alpha / n}\left(\kappa^{i, 0} S_{U}^{n}\right)\right\}} \\
\quad=1_{\left\{\left(\kappa^{\infty, i, 1}-\kappa^{\infty, i, 0}\right)\left(2 S_{U}^{n}+\left(\kappa^{\infty, i, 0}+\kappa^{\infty, i, 1}\right) \alpha \sum_{i=1}^{\mu} w_{i}^{2}\right)>0\right\}}
\end{aligned}
$$

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

$$
\lim _{n \rightarrow \infty} \bar{\kappa}\left(e_{1}, \frac{\alpha}{n} S_{U}^{n}, \kappa_{\text {init }}\right)=\bar{\kappa}^{\infty}\left(\left(\left[U^{i}\right]_{1}\right)_{i=1, \ldots, \lambda}, \kappa_{\text {init }}\right)
$$

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

Lemma A.1. For $n \in \mathbb{N}^{*}$, let $U_{n}^{1}, \ldots, U_{n}^{\mu}$ be i.i.d. standard multivariate 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 $\bar{\kappa}_{\mathrm{PLS}}$ or all upperbounded by a positive constant $\kappa^{M}$. 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

$$
\begin{aligned}
n \ln \| e_{1}+ & \bar{\kappa}_{n}\left(e_{1}, \alpha / n \sum_{i=1}^{\mu} w_{i} U_{n}^{i}\right) \frac{\alpha}{n} \sum_{i=1}^{n} w_{i} U_{n}^{i} \|^{2} \\
& =-8 \ln \left[\left(1-\frac{\left[\sum_{i=1}^{\mu} w_{i} U_{n}^{i}\right]_{1}^{2}}{\left\|\sum_{i=1}^{\mu} w_{i} U_{n}^{i}\right\|^{2}}\right)^{-n / 8}\right] \mathbf{1}_{\left[\sum_{i=1}^{\mu} w_{i} U_{n}^{i}\right]_{1}<0}
\end{aligned}
$$

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

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

is uniformly integrable.
Suppose now that there exists $\kappa^{M}>0$ such that for all $n \in \mathbb{N}^{*}$, $\bar{\kappa}_{n} \leqslant \kappa^{M}$. 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 $\bar{\kappa}_{\mathrm{PLS}}$, then

$$
\begin{aligned}
& \left\|e_{1}+\bar{\kappa}_{n}\left(e_{1}, \alpha / n \sum_{i=1}^{\mu} w_{i} U_{n}^{i}\right) \alpha / n \sum_{i=1}^{n} w_{i} U_{n}^{i}\right\|^{2} \\
& \geqslant\left\|e_{1}+\bar{\kappa}_{\mathrm{PLS}}\left(e_{1}, \alpha / n \sum_{i=1}^{\mu} w_{i} U_{n}^{i}\right) \alpha / n \sum_{i=1}^{n} w_{i} U_{n}^{i}\right\|^{2}
\end{aligned}
$$

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

$$
\begin{align*}
& n \ln ^{-}\left\|e_{1}+\bar{\kappa}_{n}\left(e_{1}, \alpha / n \sum_{i=1}^{\mu} w_{i} U_{n}^{i}\right) \alpha / n \sum_{i=1}^{n} w_{i} U_{n}^{i}\right\|^{2} \\
& \quad \leqslant n \ln ^{-}\left\|e_{1}+\bar{\kappa}_{\mathrm{PLS}}\left(e_{1}, \alpha / n \sum_{i=1}^{\mu} w_{i} U_{n}^{i}\right) \alpha / n \sum_{i=1}^{n} w_{i} U_{n}^{i}\right\|^{2} \tag{1}
\end{align*}
$$

Note that, as $\left\|e_{1}+\bar{\kappa}_{\mathrm{PLS}}\left(e_{1}, \alpha / n \sum_{i=1}^{\mu} w_{i} U_{n}^{i}\right) \alpha / n \sum_{i=1}^{n} w_{i} U_{n}^{i}\right\| \leqslant \| e_{1}+$ $0 \|=1$, then $\ln \left\|e_{1}+\bar{\kappa}_{\text {PLS }}\left(e_{1}, \alpha / n \sum_{i=1}^{\mu} w_{i} U_{n}^{i}\right) \alpha / n \sum_{i=1}^{n} w_{i} U_{n}^{i}\right\| \leqslant$ $\left\|e_{1}+0\right\| \leqslant 0$, thus is equal in absolute value to its negative part.

However, as

$$
\left(n \ln ^{-}\left\|e_{1}+\bar{\kappa}_{\mathrm{PLS}}\left(e_{1}, \alpha / n \sum_{i=1}^{\mu} w_{i} U_{n}^{i}\right) \alpha / n \sum_{i=1}^{n} w_{i} U_{n}^{i}\right\|^{2}\right)_{n \in \mathbb{N}^{*}}
$$

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 $0 \leqslant \kappa \leqslant \kappa^{M}$, then

$$
\ln ^{+}\left\|e_{1}+\kappa x\right\| \leqslant \ln ^{+}\left\|e_{1}+\kappa^{M} x\right\|
$$

Indeed, the above equation is equivalent to

$$
\left\|e_{1}+\kappa x\right\|^{2} \mathbf{1}_{\left\|e_{1}+\kappa x\right\| \geq 1} \leqslant\left\|e_{1}+\kappa^{M} x\right\|^{2} \mathbf{1}_{\left\|e_{1}+\kappa^{M} x\right\| \geq 1}
$$

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

For the positive part, we have then

$$
\begin{aligned}
n \ln ^{+} \| e_{1}+\bar{\kappa}_{n}\left(e_{1}, \frac{\alpha}{n} \sum_{i=1}^{\mu} w_{i} U_{n}^{i}\right) & \frac{\alpha}{n} \sum_{i=1}^{\mu} w_{i} U_{n}^{i} \|^{2} \\
& \leqslant n \ln ^{+}\left\|e_{1}+\frac{\kappa^{M} \alpha}{n} \sum_{i=1}^{\mu} w_{i} U_{n}^{i}\right\|^{2}
\end{aligned}
$$

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.

## REFERENCES

[1] Armand Gissler, Anne Auger, and Nikolaus Hansen. Learning rate adapatation 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 $\left(\mu / \mu_{w}, \lambda\right)$-ES and optimal $\mu$ for intermediate recombination for large population sizes. In International Conference on Parallel Problem Solving from Nature, pages 52-62. Springer, 2010.

[^0]
[^0]:    ${ }^{1}$ replace $\sigma^{*}$ by $\alpha \kappa{ }^{M}$ in the proof of Proposition 4 of [2]

