Analysis of the Effect of Multiplicative Noise on the $(\mu/\mu_w, \lambda)$-ES under Two Noise Models

Alexander Jungeilges[1][0009−0007−3131−6314], Alexandre Chotard[2][0009−0009−3606−4237], and Anne Auger[3][0009−0008−0912−2764]

[1] Institute for Neural Computation, Faculty of Computer Science, Ruhr-University Bochum, Germany
alexander.jungeilges@ini.rub.de
[2] Université du Littoral Côte d'Opale, France
alexandre.chotard@univ-littoral.fr
[3] Inria, CMAP Ecole Polytechnique, France
{firstname.lastame}@inria.fr

Abstract. Evolution strategies (ES) are effective optimizers in noisy settings, where their selection mechanism provides an inherent strength against uncertainty in the objective function. In the setting of multiplicative noise, this paper analyzes the impact of two different noise formulations on the selection mechanism of the $(\mu/\mu_w, \lambda)$-ES. In particular, we compare a log-normal noise model with its first-order approximation and show that their selection dynamics differ in the limit of large noise strength. Using a step-size proportional to the distance to the optimum, we relate the expected logarithmic progress to almost sure linear convergence and analyze the resulting convergence rates. Although the models differ in finite dimensions, we show that their asymptotic convergence rates coincide as the dimension tends to infinity.

Keywords: Evolution Strategy · Multiplicative Noise · Linear Convergence.

1 Introduction

Noise appears in many real-world optimization problems, arising from sources such as measurement errors in physical experiments, randomness in simulation-based evaluations, or stochastic effects in data-driven models. As a result, objective function evaluations are distorted and may produce a range of possible fitness values. Several noise models have been proposed in the literature often classified into additive and multiplicative noise. Let $f(x)$ denote the noiseless objective function and $\tilde{f}(x, \eta)$ one noisy instance of f where η denotes a one-dimensional random variable. In the case of additive noise, $\tilde{f}(x, \eta) = f(x) + \eta$, i.e. the fluctuations are added independently of the current function value. The noise is often modeled as a zero-mean Gaussian random variable with constant variance, allowing the observed function value to be perturbed positively or negatively. Since the noise amplitude remains constant while the true function

value decreases during optimization, the signal-to-noise ratio deteriorates as the optimum is approached.

In contrast, multiplicative noise, $\tilde{f}(x,\eta) = f(x)\eta$, scales with the value of the objective function. Consequently, as the optimization process approaches the optimum and $f(x)$ decreases, the magnitude of the noise diminishes as well. This property can preserve the relative scale between signal and noise throughout the search. In our setting, we assume that $f(x) \geq 0$ and that the optimal objective value is 0. Under these assumptions, the noise magnitude vanishes near the optimum.

Evolution strategies (ES) have been shown to be a competitive choice in noisy environments, as their selection mechanism provides an inherent robustness to uncertainty in the objective function [4, 16]. In the ES literature, the following additive noise model, $f(x) + \tilde{\sigma}_\varepsilon \mathcal{N}(0,1)$, has been motivating theoretical studies on noise where $\tilde{\sigma}_\varepsilon > 0$ denotes the noise strength and $\mathcal{N}(0,1)$ a one-dimensional standard normal distribution [3, 6, 14]. Subsequently, the authors introduce a normalized noise strength $\sigma_\varepsilon = \tilde{\sigma}_\varepsilon d/R^2$, where $d \in \mathbb{N}$ denotes the dimension and R the distance from the optimal solution[4]. On the sphere function, it holds $f(x) = R^2 = \|x\|^2$, which in turn leads to $\tilde{\sigma}_\varepsilon = \sigma_\varepsilon f(x)/d$ after rearranging. The noise model can then be written as

$$f(x) + \frac{\sigma_\varepsilon}{d} f(x)\mathcal{N}(0,1) = f(x)\left(1 + \frac{\sigma_\varepsilon}{d}\mathcal{N}(0,1)\right), \tag{1}$$

which coincides with a multiplicative noise model. Consequently, the corresponding results [3,6,14] effectively describe a multiplicative noise setting, in which the noise strength scales proportionally with the function value and thus vanishes as the optimum is approached.

This scaling relies on the non-negativity of the noise realizations. However, the noise model (1) admits negative fitness values with probability $\mathbb{P}(\mathcal{N}(0,1) < -\sigma_\varepsilon/d) > 0$. For objective functions that are strictly non-negative, such as the sphere, this implies arbitrary good candidate solutions can be observed, even when the true function value is positive and corresponds to solutions far from the noiseless optimum. This can be seen as a defect of the noise model rather than a realistic modeling of a practical situation. Independently of the possibility of negative noisy evaluations, as shown in Section 3, infinite noise strength in the noise model (1) does not lead to a uniform ranking of noisy samples. This includes noise distributions with positive distributions, and is a property of noise model (1). In the context of a $(1+1)$-ES, it has been shown that replacing the normal distribution in the noise model (1) with a uniform distribution can lead to non-convergence to the optimum on the sphere function [20]. Furthermore, the noise model (1) is a first-order approximation of the log-normal model $f(x)\exp((\sigma_\varepsilon/d)\mathcal{N}(0,1))$, for which non-negativity is guaranteed, and for which infinite noise strength does lead to a uniform ranking of noisy samples.

[4] Arnold et al. define the normalized noise strength as $\sigma_\varepsilon = \tilde{\sigma}_\varepsilon d/(2R^2)$. For simplicity, the constant factor 2 is omitted here, since it does not influence the subsequent analysis.

The purpose of this paper is to theoretically study the differences and similarities of both noise models and evaluate the impact of their differences in the context of a $(\mu/\mu_w, \lambda)$-ES. Throughout this paper we primarily consider the sphere function $f(x) = \|x\|^2$ as the noiseless objective function. Therefore, we will study the noisy sphere with multiplicative noise denoted in a generic way as in (2), specified with a log-ξ multiplicative noise (3) and its first-order approximation (4):

$$\|x\|^2\eta \quad (2) \qquad \|x\|^2 \exp\left(\frac{\sigma_\varepsilon}{d}\xi\right) \quad (3) \qquad \|x\|^2\left(1+\frac{\sigma_\varepsilon}{d}\xi\right) \quad (4)$$

where η and ξ are random variables sampled each time an evaluation is asked on the objective function. To instantiate the noise models discussed above, ξ follows a standard normal distribution in which case for (3) we talk about log-normal multiplicative noise.

The paper is organized as follows. The next section introduces the $(\mu/\mu_w, \lambda)$-ES. Afterwards, the selection dynamics with respect to both noise models are analyzed, with particular emphasis on the ranking of candidate solutions as the noise strength tends to infinity. Section 4 investigates the convergence behavior in a noisy setting. Using a step-size proportional to the distance to the optimum, the expected logarithmic progress is related to almost sure linear convergence. Furthermore, both noise models are connected through their common asymptotic limit of the convergence rate as the dimension tends to infinity. Finally, Section 5 presents an experimental study comparing the resulting convergence rates for the two noise models.

2 The $(\mu/\mu_w, \lambda)$-Evolution Strategy Optimizing the Sphere with Multiplicative Noise

We consider the isotropic $(\mu/\mu_w, \lambda)$-ES to minimize a d-dimensional noisy function $\tilde{f} : \mathbb{R}^d \times \Omega \to \mathbb{R}$, with Ω denoting the sample space of the random variable η. In iteration $t \geq 0$, given $X_t \in \mathbb{R}^d$ representing the mean of the Gaussian sampling distribution and σ_t its standard deviation, it samples λ independent random vectors $U_{t+1}^1, .., U_{t+1}^\lambda$ from the d-dimensional standard normal distribution $\mathcal{N}(0, I_d)$, where $0 \in \mathbb{R}^d$ is the zero vector and I_d is the d-dimensional identity matrix. The $\lambda \in \mathbb{N}$ candidate solutions $X_{t+1}^1, .., X_{t+1}^\lambda$, given by $X_{t+1}^i = X_t + \sigma_t U_{t+1}^i$, with $\sigma_t > 0$, are then evaluated on the noisy objective function. The noise is sampled independently for each evaluation, yielding $\tilde{f}(X_{t+1}^1, \eta_{t+1}^1), \ldots, \tilde{f}(X_{t+1}^\lambda, \eta_{t+1}^\lambda)$. Here, $\eta_{t+1}^1, \ldots, \eta_{t+1}^\lambda$ are independent random variables drawn from a density denoted p_{noise}. The algorithm then ranks the candidate solutions as

$$\tilde{f}(X_{t+1}^{s(1)}, \eta_{t+1}^{s(1)}) \leq \ldots \leq \tilde{f}(X_{t+1}^{s(\lambda)}, \eta_{t+1}^{s(\lambda)}), \tag{5}$$

where $s(\cdot)$ denotes the permutation of $\{1, \ldots, \lambda\}$ that orders the offspring according to their noisy objective function values. Let $\mu \in \mathbb{N}$, $0 < \mu < \lambda$ and

$w_1 \geq \cdots \geq w_\mu$ denote some real-valued recombination weights. Various formulations for these weights exist; an overview can be found in [17]. For now, the weights are kept general and will be specified when required, in particular for the experimental results presented in Section 5.

Finally, the state variables of the algorithm are updated. The mean vector of the search distribution for the next iteration, X_{t+1}, is calculated as

$$X_{t+1} = X_t + \sigma_t \sum_{i=1}^{\mu} w_i U_{t+1}^{s(i)}. \tag{6}$$

There exist several update mechanisms of the step size σ_{t+1}. For now, we treat it as an oracle; a specific adaptation mechanism will be introduced later. Given $X_t = x$ and $\sigma_t = \sigma$, we denote by $\alpha_{x,\sigma}(U_{t+1}, \eta_{t+1})$ the random vector containing the μ best samples according to the objective function. We say that this vector is composed of selected samples together with the noise associated as they are selected to update the state-variables of the algorithm, i.e,

$$\alpha_{x,\sigma}(U_{t+1}, \eta_{t+1}) = \left((U_{t+1}^{s(1)}, \eta_{t+1}^{s(1)}), \ldots, (U_{t+1}^{s(\mu)}, \eta_{t+1}^{s(\mu)}) \right). \tag{7}$$

Given $X_t = x$ and $\sigma_t = \sigma$, following the same line as in [19, Lemma 1][5], we derive that the distribution of $\alpha_{x,\sigma}(U_{t+1}, \eta_{t+1})$ on the multiplicative noisy sphere (2) follows

$$p_{x,\sigma}^{\text{sel,noise}}(u, \eta) = \frac{\lambda! Q_{x,\sigma}(u^\mu, \eta^\mu)^{\lambda-\mu}}{(\lambda-\mu)!} \mathbb{1}_{\{\|x+\sigma u^1\|^2 \eta^1 \leq \ldots \leq \|x+\sigma u^\mu\|^2 \eta^\mu\}} \prod_{i=1}^{\mu} p_{\mathcal{N}}(u^i) p_{\text{noise}}(\eta^i) \tag{8}$$

where $u = (u^1, \ldots, u^\mu) \in (\mathbb{R}^d)^\mu$, $\eta = (\eta^1, \ldots, \eta^\mu) \in \mathbb{R}^\mu$, $p_{\mathcal{N}}$ is the density of a d-dimensional standard normal distribution, p_{noise} the density followed by each sampled noise random variable and Q the function defined as

$$Q_{x,\sigma}(u, \eta) := \mathbb{P}_{N \sim \mathcal{N}(0, I_d), \nu \sim p_{\text{noise}}} \left(\|x + \sigma u\|^2 \eta < \|x + \sigma N\|^2 \nu \right). \tag{9}$$

3 When Both Models Differ: Noise Strength to Infinity

From a theoretical perspective, it is natural to assume that for $\sigma_\varepsilon \to \infty$, the noise drowns out any signal from the objective function, and that the noisy ranking should converge to a uniform distribution over the λ offspring. In the absence of signal, selection does not change the sample distribution. Hence, the joint distribution of the selected samples (7) should converge when σ_ε goes to ∞ to the distribution of the samples prior to selection. This property holds for the log-ξ model, which can be seen in the inequality $f(x_1)\eta_1 < f(x_2)\eta_2$: indeed

[5] See in particular the HAL version with proofs.

assume two independent candidate solutions equal to x_1 and x_2, with associated noise realizations $\eta_1 \neq \eta_2$. The following inequalities determining the best among the two samples x_1 and x_2 are equivalent:

$$\begin{aligned} f(x_1)\eta_1 < f(x_2)\eta_2 &\Leftrightarrow f(x_1)\exp\left(\frac{\sigma_\varepsilon}{d}\xi_1\right) < f(x_2)\exp\left(\frac{\sigma_\varepsilon}{d}\xi_2\right) \\ &\Leftrightarrow \xi_1 < \xi_2 + \frac{d}{\sigma_\varepsilon}\log\left(\frac{f(x_2)}{f(x_1)}\right) \quad \underset{\sigma_\varepsilon\to\infty}{\Leftrightarrow} \xi_1 < \xi_2 \ . \end{aligned} \tag{10}$$

In the limit of infinite noise strength, the ranking of x_1 and x_2 is independent of $f(x_1)$ and $f(x_2)$, and is entirely determined by the noise. The situation is different for the first-order model. Indeed, in this case, assuming $f(x_1)\xi_1 \neq f(x_2)\xi_2$, the ranking of x_1 and x_2 is determined via the following equivalent inequalities:

$$\begin{aligned} f(x_1)\eta_1 < f(x_2)\eta_2 &\Leftrightarrow f(x_1)\left(1+\frac{\sigma_\varepsilon}{d}\xi_1\right) < f(x_2)\left(1+\frac{\sigma_\varepsilon}{d}\xi_2\right) \\ &\Leftrightarrow f(x_1)\xi_1 < f(x_2)\xi_2 + \frac{d(f(x_2)-f(x_1))}{\sigma_\varepsilon} \\ &\underset{\sigma_\varepsilon\to\infty}{\Leftrightarrow} f(x_1)\xi_1 < f(x_2)\xi_2 \ . \end{aligned} \tag{11}$$

The f-values of x_1 and x_2 influence the ranking even when σ_ε is infinite, so in this model σ_ε to infinity does not erase signal. Note that no particular assumption was used here on either f or the distribution of ξ (except avoiding events where $\xi_1 = \xi_2$ or $f(x_1)\xi_1 = f(x_2)\xi_2$, by say assuming absolute continuity of the distribution of ξ with respect to the Lebesgue measure and $f > 0$)[6]. The difference in ranking for σ_ε to infinity between the two models is formalized in the following proposition.

Proposition 1. *Consider the two different noise models from* (3) *and* (4) *where* $x \mapsto \|x\|^2$ *is replaced by any* f *non-negative measurable function, applied to* λ *candidate solutions* $y := (y^i)_{i\in[1..\lambda]} \in (\mathbb{R}^d)^\lambda$ *with non-zero* f*-values. Assume that the distribution of the noise* ξ *is absolutely continuous with respect to the Lebesgue measure, with density* p_ξ*. Let* $s_y^{\sigma_\varepsilon}$ *denote the random permutation for which the noisy ordering* (5) *holds, conditionally to* $X_{t+1}^i = y^i$ *for each* $i \in [1..\lambda]$*. When* σ_ε *tends to infinity, in the log-*ξ *model* $s_y^{\sigma_\varepsilon}$ *converges in distribution to a random permutation* s^∞ *which is independent of* y *and uniformly distributed on the set of permutations* $\mathfrak{S}_\lambda$*. That is, for any* $s \in \mathfrak{S}_\lambda$, $\Pr(s^\infty = s) = 1/\lambda!$.

In the case of the first-order model, $s_y^{\sigma_\varepsilon}$ *converges in distribution to a random permutation* s_y^{FO}*, which is distributed as follows: for all* $s \in \mathfrak{S}_\lambda$

$$\Pr(s_y^{FO} = s) = \Pr_{\xi\sim p_\xi^\lambda}\left(f(y^{s(1)})\xi^{s(1)} < \ldots < f(y^{s(\lambda)})\xi^{s(\lambda)}\right) \ .$$

Proof. Let s be a permutation of $\mathfrak{S}_\lambda$. In the log-ξ model, the probability that $s_y^{\sigma_\varepsilon}$ equals s is $\Pr_{\xi\sim p_\xi^\lambda}(f(y^{s(1)})\exp(\sigma_\varepsilon/d\xi^{s(1)}) < \ldots < f(y^{s(\lambda)})\exp(\sigma_\varepsilon/d\xi^{s(\lambda)}))$.

[6] Note also that in an n^{th} order model, the inequality (11) would come to $\sqrt[n]{f(x_1)}\xi_1 < \sqrt[n]{f(x_2)}\xi_2$: the relative variance of $\sqrt[n]{f(x)}$ flattens out with higher order models.

By Lebesgue's Dominated Convergence Theorem, using the same argument as for inequalities (10), $\Pr(s_y^{\sigma_\varepsilon} = s)$ tends to $\Pr_{\xi \sim p_\xi^\lambda}(\xi^{s(1)} < \ldots < \xi^{s(\lambda)})$, which equals $1/\lambda!$ as the ξ^i are i.i.d. Note that the inequalities hold in the limit in distribution, as the event $\xi^{s(i)} = \xi^{s(i+1)}$ has negligible probability. By the same reasoning, for the first-order model $\Pr(s_y^{\sigma_\varepsilon} = s)$ tends to $\Pr_{\xi \sim p_\xi^\lambda}(f(y^{s(1)})\xi^{s(1)} < \ldots < f(y^{s(\lambda)})\xi^{s(\lambda)})$. □

The previous proposition highlights the difference in how both noise models (3) and (4) rank candidates solutions. The probability $p_{i,j}^{\sigma_\varepsilon}$ that the candidate with i^{th} best noiseless f-value is ranked as j^{th} best noisy value is estimated in Figure 1 for different noise strength. With no noise ($\sigma_\varepsilon = 0$), $p_{i,i}^{\sigma_\varepsilon} = 1$, and as σ_ε increases, $p_{i,i}^{\sigma_\varepsilon}$ decreases and spreads its mass to the neighbouring events. In the case of the log-normal model, as predicted by Proposition 1, the probability mass spreads to a uniform distribution. When ξ follows a standard normal distribution, the first-order model, relatively to the log-normal model, biases worse candidates (in terms of noiseless f-values) to extreme (best and worst) noisy ranks, and best candidates (for noiseless f) to middle noisy ranks: the influence of the function values does not vanish with infinite σ_ε. The magnitude of this bias depends on the relative variations in f-values among the population: if σ is small or the dimension d is high, then as shown in Figure 2, the difference between the two models is small. The opposite holds for large values of σ, or large population size λ. This effect also depends on the steepness of f, and would for example be magnified on the higher scaling sphere function $\|x\|^4$. In the case of the sphere function $\|x\|^2$ and for $\sigma = 0.1$, as shown in Figure 1, the difference in ranking between the two models is important for noise levels σ_ε/d larger than 2.

In summary, we showed that the two models scale differently with σ_ε, and that contrarily to what we may expect, in the first-order model, the influence of the candidates' f-values does not disappear even under infinite noise strength. This is not the case for the log-ξ model. When ξ is normally distributed, as noisy evaluations $f(x)(1 + \sigma_\varepsilon/d\mathcal{N}(0,1))$ are likely to be negative for high σ_ε, it may seem obvious that problems are bound to occur. However, we showed that the difference in limit is not due to possible negative values: it exists irrespectively of the choice of p_ξ, including if ξ were a positive random variable. We can also infer that the difference between the two models under high noise strength decreases with smaller σ and in high dimension, which is helpful to estimate when the first-order model is a good approximation to the log-ξ model. Additional simulations on the convergence rate of the algorithm, presented in Section 5, illustrate the resulting differences in convergence behavior between the two models. Before turning to these experiments, the next section investigates the convergence mechanisms of the $(\mu/\mu_w, \lambda)$-ES with respect to both noise models.

4 Linear Convergence Bounds Associated to Both Models

Linear convergence plays a central role in optimization theory. It corresponds to the convergence of many first-order methods including gradient descent on

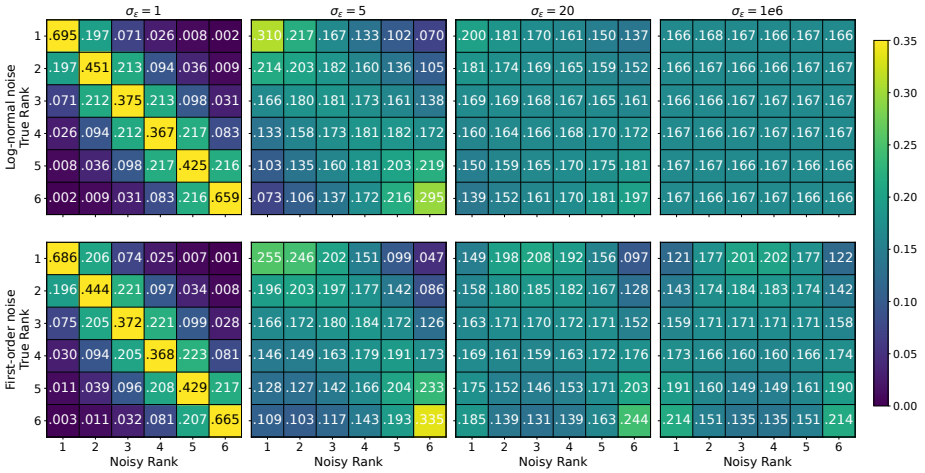

Fig. 1. Probability $p_{i,j}^{\sigma_\varepsilon}$ for six candidate solutions drawn from $\|e_1 + \sigma N\|^2 \eta$, where $N \sim \mathcal{N}(0, I_d)$. The noise η is modelled according to the log-normal (upper row) and first-order (lower row) noise model. Each column depicts a different noise strength $\sigma_\varepsilon \in \{1, 5, 20, 10^6\}$. The ranking is averaged through Monte Carlo simulations with 10^6 samples and parameters: $d = 10, \sigma = 0.1$. The intensity of color indicates the probability that a candidate with true rank $i \in \{1, .., 6\}$ is ranked at position $j \in \{1, .., 6\}$ after the noise is applied.

L-smooth and μ-strongly convex objectives [12]. Although ES are stochastic and gradient-free methods, they also exhibit linear convergence on broad classes of functions [15]. Consider a function with global minimizer $x^\star$. For a stochastic algorithm generating a sequence of incumbents $(X_t)_{t\in\mathbb{N}}$, linear convergence can be formalized by the existence of a constant $\mathrm{CR} > 0$ such that

$$\lim_{t\to\infty} \frac{1}{t} \ln \frac{\|X_t - x^\star\|}{\|X_0 - x^\star\|} = -\,\mathrm{CR}\,. \tag{12}$$

This definition is consistent with empirical observations from single runs: after a transient phase, that can be short if the parameters of the algorithm are well adapted, the log-distance to the optimum decreases approximately linearly with slope $-\,\mathrm{CR}$ when plotted on a logarithmic scale. Linear convergence is illustrated in Figure 3 where single runs are displayed using a log scale. Runs associated to a negative slope converge linearly and $-\,\mathrm{CR}$ is the limit of this slope for t to infinity.

Linear convergence of step-size adaptive evolution strategies has been proven for several step-size adaptive algorithms [1, 2, 7, 22]. In particular, this has been proven for the (1+1)-ES with success rule adaptation on classes that include μ-strongly convex objectives as well as some non-convex functions [2].

While the above studies consider concrete step-size mechanisms, a specific (artificial) step-size update plays a central role in the theory of ES, namely an

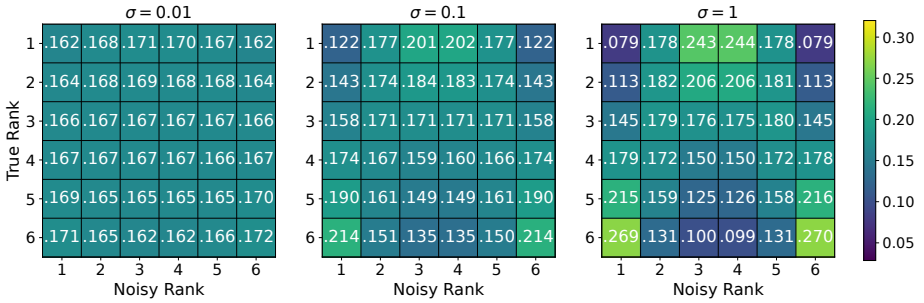

Fig. 2. Probability $p_{i,j}^{\sigma_\varepsilon}$ for six candidate solutions drawn from $\|e_1 + \sigma N\|^2 \eta$, where $N \sim \mathcal{N}(0, I_d)$, for $\sigma \in \{0.01, 0.1, 1\}$. The noise η is modelled according to the first-order noise model. The ranking is averaged through Monte Carlo simulations with 10^6 samples and parameters: $\sigma_\varepsilon = 10^6$ and $d = 10$.

update where the step-size is proportional to the distance to the optimum

$$\sigma_t = \sigma \, \|X_t - x^\star\|, \tag{13}$$

with $\sigma > 0$. It enforces scale invariance by normalizing the step size with the distance to the optimum. Although this rule uses information that is only available in a theoretical setting, it captures the scale-invariant nature of the algorithm. In particular, many practical step-size adaptation mechanisms aim to maintain a roughly constant ratio between the step size and the distance to the optimum [8, 15]. This property is the underlying idea of approaches that analyze the convergence of Evolution Strategies by investigating the stability of underlying Markov chains [10]. The above rule idealizes this behavior by keeping the normalized step size $\sigma_t / \|X_t - x^\star\| = \sigma$ constant. Proofs of the linear convergence of ES with step-size proportional to the optimum on spherical functions are simpler to establish and the associated convergence rate is a function of the proportionality constant σ [19, 21]. In the case of the noisy sphere with multiplicative noise, the proof works in the same way as without noise. We sketch the proof idea and refer to previous work for more precise mathematical arguments. We assume without loss of generality $x^\star = 0$. First, we express the log-distance to the optimum divided by t as an empirical sum. For σ_t following Eq. (13):

$$\frac{1}{t} \ln \frac{\|X_t\|}{\|X_0\|} = \frac{1}{t} \sum_{k=0}^{t-1} \ln \frac{\|X_{k+1}\|}{\|X_k\|} = \frac{1}{t} \sum_{k=0}^{t-1} \ln \frac{\|X_k + \sigma \|X_k\| \sum_{i=1}^{\mu} w_i U_{k+1}^{s(i)}\|}{\|X_k\|}. \tag{14}$$

We denote $e_1 = (1, 0, \ldots, 0)$ as the first unit vector. Using the rotational invariance of the Gaussian distribution and of the sphere function, and the scale invariance of the sphere function, the distribution of the random variable $\ln(\|X_k + \sigma \|X_k\| \sum_{i=1}^{\mu} w_i U_{k+1}^{s(i)}\| / \|X_k\|)$ is the same as the distribution of $\ln \|e_1 + \sigma \sum_{i=1}^{\mu} w_i U_{k+1}^{s_{e_1}(i)}\|$. Here, the selection permutation s_{e_1} is determined by ranking the candidate solutions $\{e_1 + \sigma U_{k+1}^i, i = 1, \ldots, \lambda\}$ sampled from e_1, on the noisy sphere

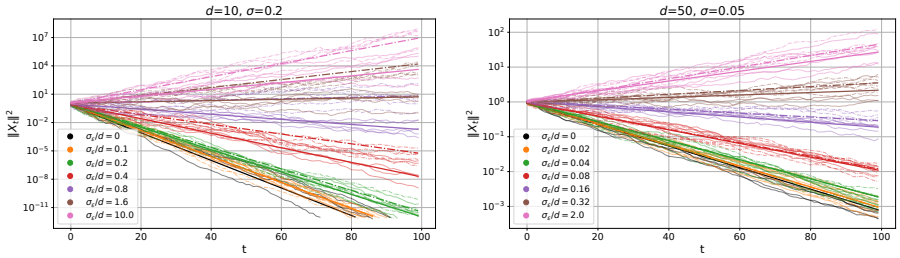

Fig. 3. Single runs displaying the linear convergence (or divergence) of a $(\mu/\mu_w, \lambda)$-ES, with $\mu = 5$ and $\lambda = 10$, on the noisy sphere with step-size proportional to the distance to the optimum. The log-normal model is displayed with solid lines while the first-order model is displayed with dotted lines. Different noise strength $\sigma_\varepsilon \in \{0, 1, 2, 4, 8, 16, 100\}$ are considered and the algorithm is initialized with logarithmic weights $w_i = \ln((\lambda + 1)/2) - \ln(i)$ and $(w_1, \ldots, w_\mu)$ [17]. On the y-axis, in log-scale, the noiseless objective function of the recombined individual is presented while the x-axis displays the number of iterations. The convergence rate corresponds to the slopes of the single runs shown in the background. The slope of the thick lines corresponds to the estimated convergence rate $G(\sigma)$ for both noise models computed using Monte Carlo integration with 10^6 samples.

$f(x, \eta) = \|x\|^2\eta$, i.e. the permutation s_{e_1} is determined by

$$\|e_1 + \sigma U_{k+1}^{s_{e_1}(1)}\|^2 \eta_{k+1}^{s_{e_1}(1)} \leq \ldots \leq \|e_1 + \sigma U_{k+1}^{s_{e_1}(\lambda)}\|^2 \eta_{k+1}^{s_{e_1}(\lambda)}.$$

We can thus rewrite (14) as

$$\frac{1}{t} \ln \frac{\|X_t\|}{\|X_0\|} \overset{d}{=} \frac{1}{t} \sum_{k=1}^{t} \ln \left\| e_1 + \sigma \sum_{i=1}^{\mu} w_i U_{k+1}^{s_{e_1}(i)} \right\|. \tag{15}$$

The random variables $\{Y_k = \ln \|e_1 + \sigma \sum_{i=1}^{\mu} w_i U_{k+1}^{s_{e_1}(i)}\|, k \in \mathbb{N}\}$ are i.i.d. [19]. Hence, when t goes to infinity, the RHS of (15) converges by the Law of Large Numbers to the expected value of Y_0. Overall we obtain the asymptotic limit

$$\begin{aligned}
\lim_{t\to\infty} \frac{1}{t} \ln \frac{\|X_t\|}{\|X_0\|} &= \mathbb{E}(Y_0) \\
&= \lambda! \mathbb{E}\left(\ln \|e_1 + \sigma \sum_{i=1}^{\mu} w_i U^i\| 1_{\|e_1+\sigma U^1\|^2\eta^1 \leq \ldots \leq \|e_1+\sigma U^\lambda\|^2\eta^\lambda} \right) \\
&= \underbrace{\int\int \ln \|e_1 + \sigma \sum_{i=1}^{\mu} w_i u^i\| \, p_{e_1,\sigma}^{\text{sel,noise}}(u, \nu) \, du \, d\nu}_{:= G(\sigma)}
\end{aligned} \tag{16}$$

where the density of the selected steps $p_{e_1,\sigma}^{\text{sel,noise}}$ is given by (8). In the previous equation, we identify the right-hand-side as the asymptotic convergence rate

– CR from (12). This convergence rate depends on σ and on the noise model and we denote it $G(\sigma)$. Only when $G(\sigma) < 0$ the algorithm converges linearly while it diverges when $G(\sigma) > 0$. We introduce two notations to denote the convergence rate as a function of σ for the log-normal noise model and the first-order model with normal noise:

- when $\eta \sim \exp\left(\frac{\sigma_\varepsilon}{d}\mathcal{N}\right)$, the function in the RHS of (16) is denoted $\sigma \mapsto G^{\mathrm{LN}}(\sigma)$
- when $\eta \sim \left(1 + \frac{\sigma_\varepsilon}{d}\mathcal{N}\right)$, the function in the RHS of (16) is denoted $\sigma \mapsto G^{\mathrm{FO}}(\sigma)$.

The difference between both functions and thus between both models is illustrated and discussed in Section 5.

In (16), we express the asymptotic linear convergence typically almost surely. Yet $G(\sigma)$ also corresponds to the log-progress of the ES with step-size proportional to the distance to the optimum. Assuming at each iteration $\sigma_t = \sigma\|X_t\|$ the log-progress of the $(\mu/\mu_w, \lambda)$-ES equals

$$G(\sigma) = \mathbb{E}(\ln \|X_{t+1}\|) - \mathbb{E}(\ln \|X_t\|),$$

which connects the analysis of a step-size proportional to the optimum with the progress rate theory [9, 11, 18].

4.1 Asymptotic Estimate of the Convergence Rate: both Models converge to the same Limit

The scale-invariant step-size also enables an asymptotic analysis of the function G in the limit for the dimension d to infinity: when scaling the noise strength σ_ε/d and step-size σ/d inversely proportional to d, the following theorem shows that $dG(\sigma/d)$ converges to the same limit for the two noise models.

Theorem 1. *Let $\sigma, \sigma_\varepsilon > 0$ and $d \in \mathbb{Z}_{>0}$ denote the scale-invariant step-size, noise strength and dimension, respectively. Let Z_d denote the random variable*

$$Z_d := \lambda! \ln \left\| e_1 + \sigma \sum_{i=1}^{\mu} w_i U^i \right\| 1_{\|e_1+\sigma U^1\|^2\eta^1 \leq \ldots \leq \|e_1+\sigma U^\lambda\|^2\eta^\lambda} \ ,$$

and suppose that the family of random variables $(Z_d)_d$ is uniformly integrable. Let G^{LN} and G^{FO} denote the convergence-rate functions corresponding to the log-normal and first-order noise models, respectively. Then their asymptotic limits coincide, i.e.,

$$\begin{aligned} \lim_{d\to\infty} d \cdot G^{\mathrm{LN}}\left(\frac{\sigma}{d}\right) &= \lim_{d\to\infty} d \cdot G^{\mathrm{FO}}\left(\frac{\sigma}{d}\right) \\ &= \frac{\sigma^2}{2}\sum_{i=1}^{\mu} w_i^2 + \frac{2\cdot\sigma}{\sqrt{4+\sigma_\varepsilon^2/\sigma^2}} \sum_{i=1}^{\mu} w_i \mathbb{E}(\mathcal{N}^{i:\lambda}). \end{aligned} \tag{17}$$

Here, $\mathcal{N}^{i:\lambda}$ denotes the i^{th} order statistic of λ independent standard normal distributions, that is, the i^{th} smallest value among $\mathcal{N}^1, \ldots, \mathcal{N}^\lambda$, with $\mathcal{N}^i \sim \mathcal{N}(0,1)$ independently.

Proof. Let $U := (U^i)_{i\in[1..\lambda]}$ be a vector of λ i.i.d. d-dimensional multivariate standard normal random vectors, and $\xi := (\xi^i)_{i\in[1..\lambda]}$ be λ i.i.d. standard normal random variables. Let X_d, $W^{\mathrm{LN}}_{U^i,\xi^i}$ and $W^{\mathrm{FO}}_{U^i,\xi^i}$ denote respectively the random variables $\|e_1 + \sigma/d \sum_{i=1}^{\mu} w_i U^i\|^2 - 1$, $\|e_1 + (\sigma/d)U^i\|^2 \exp(\sigma_\varepsilon \xi^i/d)$ and $\|e_1 + (\sigma/d)U^i\|^2(1 + \sigma_\varepsilon \xi^i/d)$. When we do not need to specify the underlying noise model, we use W_{U^i,ξ^i} for either $W^{\mathrm{LN}}_{U^i,\xi^i}$ or $W^{\mathrm{FO}}_{U^i,\xi^i}$. We are interested in the random variable $Z_d = \lambda! d \ln(\sqrt{1+X_d}) \prod_{i=1}^{\lambda-1} \mathbb{1}_{W_{U^i,\xi^i} < W_{U^{i+1},\xi^{i+1}}}$, whose expected value according to Eq. (16) is $dG(\sigma/d)$.

We derive the limits of the different terms of Z_d when $d \to \infty$. First,

$$X_d = 2\frac{\sigma}{d}\sum_{i=1}^{\mu} w_i U_1^i + \sum_{k=1}^{d}\left(\frac{\sigma}{d}\sum_{i=1}^{\mu} w_i U_k^i\right)^2$$

$$dX_d = 2\sigma\sum_{i=1}^{\mu} w_i U_1^i + \frac{\sigma^2}{d}\sum_{k=1}^{d}\left(\sum_{i=1}^{\mu} w_i^2 (U^i)^2 + 2\sum_{i=1}^{\mu}\sum_{j=i+1}^{\mu} w_i w_j U_k^i U_k^j\right) \ .$$

As d goes to infinity, by the law of large numbers $\sum_{k=1}^{d}(U_k^i)^2/d$ and $\sum_{k=1}^{d} U_k^i U_k^j/d$ for $i \neq j$ converges almost surely to respectively 1 and 0, so dX_d converges almost surely to $X_\infty := 2\sigma\sum_{i=1}^{\mu} w_i U_1^i + \sigma^2 \sum_{i=1}^{\mu} w_i^2$. By Taylor expansion of ln, $d\ln(\sqrt{1+X_d})$ converges almost surely to $X_\infty/2$. Similarly,

$$W^{\mathrm{LN}}_{U^i,\xi^i} = \left(1 + 2\frac{\sigma}{d}U_1 + \sum_{i=1}^{d}\frac{\sigma^2}{d^2}U_i^2\right)\left(1 + \frac{\sigma_\varepsilon}{d}\xi + O\left(\frac{1}{d^2}\right)\right)$$

$$d\left(W^{\mathrm{LN}}_{U^i,\xi^i} - 1\right) = 2\sigma U_1 + \sigma_\varepsilon \xi + \frac{\sigma^2}{d}\sum_{i=1}^{d} U_i^2 + O\left(\frac{1}{d}\right) \xrightarrow[d\to\infty]{a.s.} 2\sigma U_1 + \sigma_\varepsilon \xi + \sigma^2 \ .$$

The same limit holds for $d(W^{\mathrm{FO}}_{U^i,\xi^i} - 1)$. For $i \neq j$, the event that $W_{U^i,\xi^i} = W_{U^j,\xi^j}$ is negligible, so the indicator $\mathbb{1}_{W_{U^i,\xi^i} < W_{U^j,\xi^j}}$ converges almost surely to $\mathbb{1}_{Y^i < Y^j}$, where Y^i denotes $2\sigma U_1^i + \sigma_\varepsilon \xi^i$.

It follows that Z_d converges almost surely to $\lambda! X_\infty/2 \prod_{i=1}^{\lambda-1} \mathbb{1}_{Y^i < Y^{i+1}}$. Assuming that the family $(Z_d)_d$ is uniformly integrable, then

$$\lim_{d\to\infty} G\left(\frac{\sigma}{d}\right) = \lambda!\mathbb{E}\left(\left(\sigma\sum_{i=1}^{\mu} w_i U_1^i + \frac{\sigma^2}{2}\sum_{i=1}^{\mu} w_i^2\right)\prod_{i=1}^{\lambda-1}\mathbb{1}_{Y^i<Y^{i+1}}\right) \ .$$

Since U^i and ξ^i are i.i.d. and normally distributed, U^i and Y^i are a bivariate normal, and thus $\mathbb{E}(U^i|Y^i) = cY^i$ with $c = \mathrm{Cov}(U^i, Y^i)/\mathrm{Var}(Y^i)$. The covariance $\mathrm{Cov}(U^i, Y^i)$ is 2σ and the variance $\mathrm{Var}(Y^i)$ equals $4\sigma^2 + \sigma_\varepsilon^2$, so

$$\begin{aligned}\mathbb{E}\left(U_1^i \prod_{j=1}^{\lambda-1}\mathbb{1}_{Y^j<Y^{j+1}}\right) &= \mathbb{E}\left(\mathbb{E}(U_1^i|Y^i)\prod_{j=1}^{\lambda-1}\mathbb{1}_{Y^j<Y^{j+1}}\right)\\ &= \mathbb{E}\left(cY^i\prod_{j=1}^{\lambda-1}\mathbb{1}_{Y^j<Y^{j+1}}\right) = \frac{2\sigma}{\lambda!(4\sigma^2+\sigma_\varepsilon^2)}\mathbb{E}(Y^{i:\lambda}) \ ,\end{aligned}$$

where $Y^{i:\lambda}$ is the i^{th}-order statistic of the λ i.i.d. normal variables $(Y^i)_i$. Using that $Y^i \sim \sqrt{4\sigma^2 + \sigma_\varepsilon^2}\mathcal{N}(0,1)$, $\mathbb{E}(Y^{i:\lambda}) = \sqrt{4\sigma^2 + \sigma_\varepsilon^2}\mathbb{E}(\mathcal{N}^{i:\lambda})$, which concludes the proof. □

We have assumed the uniform integrability of the family of random variables $(Z_d)_d$ as it is typically technical to prove this property. Yet the formal proof should follow the same steps as [19, Proposition 4].

The theorem connects both noise models through their common asymptotic limit. In particular, the result of Theorem 1 recovers and extends the result previously derived in [5]. Here, Arnold and Beyer combined the progress rate theory with the first-order multiplicative noise model to obtain the asymptotic limit. The equivalence of the two follows from the fact that, as d tends to infinity $\exp(\sigma_\varepsilon/d \cdot \mathcal{N})$ admits the first-order Taylor approximation $(1+\sigma_\varepsilon/d \cdot \mathcal{N} + O(1/d^2))$ so that both models coincide to first-order in $1/d$.

Theorem 1 further shows that, under a scale-invariant step-size σ, the convergence rate decays as $1/d$ when the dimension d tends to infinity. Moreover, for large but finite dimensions, we derive a quantitative bound whose accuracy will be assessed through simulations in the next section. In the limit for d to infinity, this bound converges to a function that is quadratic in σ for fixed noise strength σ_ε. Since the quadratic term is non-negative, any potential negativity of the limit expression stems from the linear term involving the expected order statistics $\mathbb{E}(\mathcal{N}^{i:\lambda})$. The effect of the noise strength appears through the factor $2\sigma/\sqrt{4 + \sigma_\varepsilon^2/\sigma^2}$, which reduces the influence of the linear term and therefore also the convergence rate.

5 Experimental Results

This section complements the theoretical findings of the previous section by illustrating them through Monte Carlo integration to approximate the expected value $G(\sigma)$. For all subsequent experiments, we use logarithmic weights with $w_i = \ln((\lambda+1)/2) - \ln(i)$ and $(w_1, \ldots, w_\mu)$ [17], and fix the selection parameters to $\mu = 5$ and $\lambda = 10$.

Two parameters are of primary interest. First, for fixed dimension d, we study the effect of increasing noise strength on the convergence rate of the algorithm. This aspect is directly related to (16), since linear convergence requires the rate $G(\sigma)$ to be strictly negative. Secondly, in light of Theorem 1, we fix the noise strength σ_ε and investigate how the normalized convergence rates $dG(\sigma/d)$ of both noise models approach their common asymptotic limit as the dimension d increases. While Proposition 1 and Theorem 1 describe the differences and similarities of both models in an asymptotic setting, this section also aims to show when these effects start to occur.

5.1 Convergence Rate with Increasing Noise Strength

We present our experimental findings in Figure 4. Starting with the left plot, the convergence rates $G^{\text{LN}}(\sigma)$ and $G^{\text{FO}}(\sigma)$ are shown for different constant adaptation parameters σ and increasing noise strength σ_ε. Particular attention is given

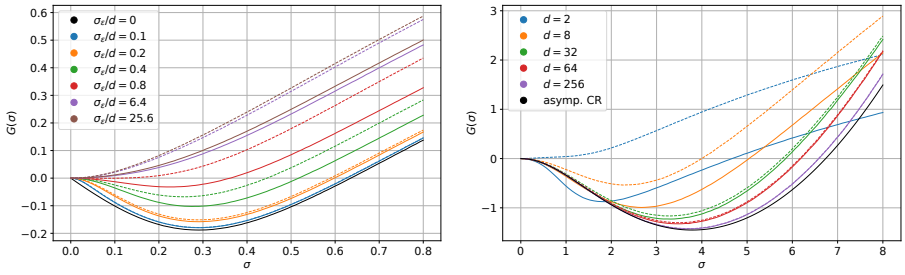

Fig. 4. Convergence rates of a $(\mu/\mu_w, \lambda)$-ES, with $\mu = 5$ and $\lambda = 10$, using a step-size proportional to the distance to the optimum. The log-normal model is shown with solid lines and the first-order model with dashed lines. Monte Carlo integration is used to obtain the plots with 10^6 samples for each data point. **Left**: $G^{\mathrm{LN}}(\sigma)$ and $G^{\mathrm{FO}}(\sigma)$ for $d = 10$ and $\sigma_\varepsilon \in \{0, 1, 2, 4, 8, 64, 256\}$. **Right**: Normalized convergence rates $dG^{\mathrm{LN}}(\sigma/d)$ and $dG^{\mathrm{FO}}(\sigma/d)$ for $\sigma_\varepsilon = 4$ and $d \in \{2, 8, 32, 64, 256\}$. The black curve shows the asymptotic limit as $d \to \infty$ (Eq. (17)).

to the convergence window, i.e., the region where G is strictly negative. As a first observation, the convergence window of the log-normal model is consistently larger than that of the first-order model, and the difference between the two increases with growing noise strength. Both convergence rates eventually saturate, approaching their respective limit curves, for increasing noise strength, close to $\sigma_\varepsilon/d = 6.4$. Here, the experiments indicate that the algorithm diverges for both noise models, even for very small values of σ.

The differences between both convergence rates can be attributed to the skewed selection mechanism that arises as the noise strength to dimension ratio σ_ε/d increases, for which the first-order approximation of $\exp((\sigma_\varepsilon/d)\mathcal{N})$ becomes increasingly inaccurate and the selection converges towards its limit distribution. However, for $\sigma_\varepsilon/d = 0.8$, which can be seen as a reasonable noise strength to dimension ratio, a significant difference in the convergence window is observed, where the first-order model exhibits almost no convergence, while the log-normal model converges over a wide range of σ values. This further highlights that, even at lower noise levels, the differences in ranking influence the convergence capabilities of the algorithm considerably. An additional distinction is observed between the noiseless and noisy cases. While the shapes are largely similar, especially for small σ, the noiseless curve shows a convex shape and requires a considerably smaller σ to achieve linear convergence.

5.2 Convergence Rate with Increasing Dimension

The right plot of Figure 4 illustrates the result from Theorem 1. The noise level is kept constant while the dimension increases. For $d = 2$ and $d = 8$ the behavior differs significantly from the asymptotic limit. This can be attributed to the choice of selection variables, μ and λ, and the effect of the weighted

recombination. As the noise strength to dimension ratio σ_ε/d decays, the difference in convergence rates diminishes, reflecting that the corresponding noise terms become increasingly similar, as highlighted in the proof of Theorem 1. For $d = 32$, the first-order model already provides a good approximation of the log-normal model. Moreover, both models approach the asymptotic convergence rate, presented in (17), and indicated by the black curve. A detailed analysis of the asymptotic convergence rate as a function of noise strength and selection parameters can be found in [5].

6 Conclusion

In this work, we analyzed the impact of multiplicative noise on the selection mechanism and convergence behavior of the $(\mu/\mu_w, \lambda)$-ES on the noisy sphere function. Two noise models were considered: the log-ξ noise model and its first-order approximation. The main results of this paper, Proposition 1 and Theorem 1, capture the differences and similarities of both models in terms of their induced noisy ranking of candidate solutions and their shared asymptotic limit of the convergence rate as the dimension tends to infinity.

While Proposition 1 investigates the selection for noise strength to infinity, the results of Section 3 are by no means restricted to an asymptotic noise strength. For finite dimension and various noise strenghts, Section 5 and Figure 4 highlighted the differences in convergence rate that arise for both models. While increasing noise strength reduces convergence, the first-order model exhibitis earlier divergence and slower convergence compared to the log-normal model. Although the first-order model provides a suitable approximation in higher dimensions, our results indicate that the log-ξ model is recommended for modeling noisy environments in a multiplicative setting.

This work leaves open several directions for future research. In particular, numerical simulations of the function G for large enough σ_ε do not allow to observe a convergence window. A theoretical proof of whether (or not) there exists an optimal step-size σ such that $G(\sigma) < 0$ for any finite noise strength σ_ε would help clarify the situation. Furthermore, this work relies on an idealized algorithm design through the use of a scale-invariant step-size. Extending the analysis to practically relevant step-size adaptation mechanisms, especially within the well-established theoretical framework using Markov Chains [13], remains an important direction for future work.

Acknowledgements

This article is based upon work from COST Action Randomised Optimisation Algorithms Research Network (ROAR-NET), CA22137, supported by COST (European Cooperation in Science and Technology). The authors also thank Tobias Glasmachers for his valuable input.

References

1. Akimoto, Y., Auger, A., Glasmachers, T.: Drift theory in continuous search spaces: expected hitting time of the (1+1)-ES with 1/5 success rule. In: Genetic and Evolutionary Computation Conference. p. 801–808. GECCO '18, ACM (2018)
2. Akimoto, Y., Auger, A., Glasmachers, T., Morinaga, D.: Global linear convergence of evolution strategies on more than smooth strongly convex functions. SIAM Journal on Optimization **32**(2), 1402–1429 (2022)
3. Arnold, D., Beyer, H.G.: A general noise model and its effects on evolution strategy performance. Evolutionary Computation, IEEE Transactions on **10**, 380 – 391 (09 2006). https://doi.org/10.1109/TEVC.2005.859467
4. Arnold, D.V.: Noisy optimization with evolution strategies, Genetic Algorithms and Evolutionary Computation, vol. 8. Springer (2002). https://doi.org/10.1007/978-1-4615-1105-2
5. Arnold, D.V., Beyer, H.G.: Local performance of the $(\mu/\mu_i, \lambda)$-ES in a noisy environment. In: Martin, W.N., Spears, W.M. (eds.) Foundations of Genetic Algorithms 6, pp. 127–141. Morgan Kaufmann, San Francisco, CA (2001). https://doi.org/10.1016/B978-155860734-7/50090-1
6. Arnold, D.V., Beyer, H.G.: Noisy local optimization with evolution strategies. Kluwer Academic Publishers, USA (2002)
7. Auger, A.: Convergence results for the (1, λ)-SA-ES using the theory of ϕ-irreducible markov chains. Theoretical Computer Science **334**(1-3), 35–69 (2005)
8. Auger, A.: Thèse d'habilitation à diriger des recherches" Analysis of Comparison-based Stochastic Continuous Black-Box Optimization Algorithms". Ph.D. thesis, University Paris Sud (2016), https://inria.hal.science/tel-01468781/document
9. Auger, A., Hansen, N.: Reconsidering the progress rate theory for evolution strategies in finite dimensions. In: Proceedings of the 8th annual conference on Genetic and evolutionary computation. pp. 445–452 (2006)
10. Auger, A., Hansen, N.: Linear convergence of comparison-based step-size adaptive randomized search on scaling-invariant functions. SIAM Journal on Optimization **26**(3), 1589–1624 (2016). https://doi.org/10.1137/15M102634X
11. Beyer, H.G., Arnold, D.V.: Theory of evolution strategies—a tutorial. Theoretical aspects of evolutionary computing pp. 109–133 (2001)
12. Boyd, S., Vandenberghe, L.: Convex optimization. Cambridge university press (2004)
13. Gissler, A.: Linear convergence of evolution strategies with covariance matrix adaptation. Theses, Institut Polytechnique de Paris (Dec 2024), https://hal.science/tel-04935192
14. Hansen, N., Arnold, D.V., Auger, A.: Evolution strategies. In: Kacprzyk, J., Pedrycz, W. (eds.) Handbook of Computational Intelligence. No. 871-898, Springer (2015), https://inria.hal.science/hal-01155533
15. Hansen, N., Arnold, D.V., Auger, A.: Evolution strategies. In: Springer handbook of computational intelligence, pp. 871–898. Springer (2015)
16. Hansen, N., Niederberger, A.S.P., Guzzella, L., Koumoutsakos, P.: A method for handling uncertainty in evolutionary optimization with an application to feedback control of combustion. IEEE Transactions on Evolutionary Computation **13**(1), 180–197 (2009). https://doi.org/10.1109/TEVC.2008.924423
17. Hansen, N., Ostermeier, A.: Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation **9**(2), 159–195 (2001). https://doi.org/10.1162/106365601750190398

18. Hellwig, M., Beyer, H.G.: On the steady state analysis of covariance matrix self-adaptation evolution strategies on the noisy ellipsoid model. Theoretical Computer Science **832**, 98–122 (2020). https://doi.org/https://doi.org/10.1016/j.tcs.2018.05.016, theory of Evolutionary Computation
19. Jebalia, M., Auger, A.: Log-linear convergence of the scale-invariant (μ/μ-w, λ)-ES and optimal μ for intermediate recombination for large population sizes. In: International Conference on Parallel Problem Solving from Nature. pp. 52–62. Springer (2010)
20. Jebalia, M., Auger, A., Hansen, N.: Log-linear convergence and divergence of the scale-invariant (1+1)-ES in noisy environments. Algorithmica **59**(3), 425–460 (2011). https://doi.org/10.1007/s00453-010-9403-3, https://doi.org/10.1007/s00453-010-9403-3
21. Jebalia, M., Auger, A., Liardet, P.: Log-linear convergence and optimal bounds for the (1+ 1)-ES. In: International Conference on Artificial Evolution (Evolution Artificielle). pp. 207–218. Springer (2007)
22. Toure, C., Auger, A., Hansen, N.: Global linear convergence of evolution strategies with recombination on scaling-invariant functions. Journal of Global Optimization **86**(1), 163–203 (2023)