

Linear Convergence on Positively Homogeneous Functions of a Comparison Based Step-Size Adaptive Randomized Search: the (1+1) ES with Generalized One-fifth Success Rule

Anne Auger, Nikolaus Hansen

▶ **To cite this version:**

Anne Auger, Nikolaus Hansen. Linear Convergence on Positively Homogeneous Functions of a Comparison Based Step-Size Adaptive Randomized Search: the (1+1) ES with Generalized One-fifth Success Rule. 2013. hal-00877161v1

HAL Id: hal-00877161

https://inria.hal.science/hal-00877161v1

Preprint submitted on 27 Oct 2013 (v1), last revised 30 Oct 2013 (v2)

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.

LINEAR CONVERGENCE ON POSITIVELY HOMOGENEOUS FUNCTIONS OF A COMPARISON BASED STEP-SIZE ADAPTIVE RANDOMIZED SEARCH: THE (1+1) ES WITH GENERALIZED ONE-FIFTH SUCCESS RULE

ANNE AUGER* AND NIKOLAUS HANSEN†

Abstract. In the context of unconstraint numerical optimization, this paper investigates the global linear convergence of a simple probabilistic derivative-free optimization algorithm (DFO). The algorithm samples a candidate solution from a standard multivariate normal distribution scaled by a step-size and centered in the current solution. This solution is accepted if it has a better objective function value than the current one. Crucial to the algorithm is the adaptation of the step-size that is done in order to maintain a certain probability of success. The algorithm, already proposed in the 60's, is a generalization of the well-known Rechenberg's $(1 + 1)$ Evolution Strategy (ES) with one-fifth success rule which was also proposed by Devroye under the name compound random search or by Schumer and Steiglitz under the name step-size adaptive random search.

In addition to be derivative-free, the algorithm is function-value-free: it exploits the objective function only through comparisons. It belongs to the class of comparison-based step-size adaptive randomized search (CB-SARS). For the convergence analysis, we follow the methodology developed in a companion paper for investigating linear convergence of CB-SARS: by exploiting invariance properties of the algorithm, we turn the study of global linear convergence on scaling-invariant functions into the study of the stability of an underlying normalized Markov chain (MC).

We hence prove global linear convergence by studying the stability (irreducibility, recurrence, positivity, geometric ergodicity) of the normalized MC associated to the $(1 + 1)$-ES. More precisely, we prove that starting from any initial solution and any step-size, linear convergence with probability one and in expectation occurs. Our proof holds on unimodal functions that are the composite of strictly increasing functions by positively homogeneous functions with degree α (assumed also to be continuously differentiable). This function class includes composite of norm functions but also non-quasi convex functions. Because of the composition by a strictly increasing function, it includes non continuous functions. We find that a sufficient condition for global linear convergence is the step-size increase on linear functions, a condition typically satisfied for standard parameter choices.

While introduced more than 40 years ago, we provide here the first proof of global linear convergence for the (1+1)-ES with generalized one-fifth success rule and the first proof of linear convergence for a CB-SARS on such a class of functions that includes non-quasi convex and non-continuous functions. Our proof also holds on functions where linear convergence of some CB-SARS was previously proven, namely convex-quadratic functions (including the well-know sphere function).

Key words. linear convergence, derivative-free-optimization, comparison-based algorithm, function-value-free optimization , evolution strategies, step-size adaptive randomized search

*INRIA Saclay-Ile-de-France (anne.auger_AT_inria.fr).

†INRIA Saclay-Ile-de-France (nikolaus.hansen_AT_inria.fr)

Contents.

1 Introduction **2**

2 Normalized Markov Chain and Objective Function Assumptions **5**
2.1 The (1+1)-ES as a Comparison-Based Step-Size Adaptive Randomized Search . . . 5
2.2 Invariances . . . 7
2.3 Normalized Markov Chain on Scaling-Invariant Functions . . . 8
2.4 Objective Function Assumptions . . . 9
2.5 State Space for the Normalized Markov Chain . . . 10

3 Study of the Normalized Chain **11**
3.1 Transition Probability Kernel . . . 11
3.2 Irreducibility, Small Sets and Aperiodicity . . . 12
3.2.1 φ-irreducibility . . . 12
3.2.2 Small Sets and Aperiodicity . . . 13
3.3 Geometric Ergodicity . . . 15
3.3.1 Geometric Drift Condition for Positively Homogenous Functions 16
3.4 Harris Recurrence and Positivity . . . 24

4 Linear Convergence of the $(1+1)$-ES with the Generalized One-fifth Success Rule **25**
4.1 Integrability w.r.t. the Stationary Measure . . . 25
4.2 Asymptotic Probability of Success . . . 26
4.3 Almost Sure Linear Convergence . . . 27
4.4 Linear convergence in expectation . . . 29
4.5 Consequences of Geometric Ergodicity: Adaptivity at a Geometric Rate 30

5 Discussion **32**

1. Introduction. Derivative-free optimization (DFO) algorithms have the advantage to handle numerical optimization problems where the function $f : \mathbb{R}^n \mapsto \mathbb{R}$ to be WLG minimized can be seen as a black-box that is only able to return an objective function value $f(\mathbf{x})$ for a given input vector $\mathbf{x}$. This context is particularly useful when dealing with many numerical optimization problems. Indeed, first, the function that needs to be optimized can result from a computer simulation where the source code might be too complex to exploit or might not be available to the person who has to do the optimization (this is typical in industry, where often only executables of the code are provided). Hence automatic differentiation to compute the gradient is not conceivable. Second, gradients can be non-exploitable because the function can be "rugged" that is noisy, very irregular, ...

Among DFO, we distinguish *function-value-free* (FVF) algorithms that do not exploit the exact objective function value but only comparisons between candidate solutions. The Nelder-Mead simplex algorithm is one of the oldest deterministic FVF algorithm [13]. While the distinction between DFO and FVF algorithms is rarely made, it has some importance both in theory and practice because FVF algorithms are invariant to composing the objective function by a strictly increasing function and hence can be seen as more robust.

We here focus on a particular class of probabilistic or randomized comparison-based (or FVF) algorithms that adapt a mean vector (thought as favorite solution)

and step-size. A general framework for those methods has been formalized under the name *comparison based step-size adaptive randomized search* (CB-SARS) [3]. Those methods find their roots among the first papers published on randomized FVF algorithms in the 60's [11, 16, 4, 15]. They were, later on, further developed in the Evolution Strategies (ES) community. The nowadays state-of-the-art Covariance Matrix Evolution Strategy (CMA-ES) where in addition to the step-size, a full covariance matrix is adapted (allowing to solve efficiently ill-conditioned problems) ensued from the developments on CB-SARS [6]. Note that contrary to some common preconception, randomized FVF (in particular CMA-ES) are competitive also for "local" optimization and can show superior performance compared to the standard BFGS or the NEWUOA [14] algorithm on *unimodal* functions provided they are significantly non-separable and non-convex [2].

We investigate the convergence of one of the earliest CB-SARS, introduced independently by Rechenberg under the name $(1+1)$-ES with one-fifth success rule [15], by Devroye as the compound random search [4] and by Schumer and Steiglitz as step-size adaptive random search [16]. Formally, let $\mathbf{X}_t \in \mathbb{R}^n$ be the mean of a multivariate normal distribution representing the favorite solution at the current iteration t. A new solution centered in $\mathbf{X}_t$ and following a multivariate normal distribution with standard deviation σ_t (corresponding also to the step-size) is sampled:

$$\mathbf{X}_t^1 = \mathbf{X}_t + \sigma_t \mathbf{U}_t^1 \tag{1.1}$$

where $\mathbf{U}_t^1$ follows a standard multivariate normal distribution, i.e., $\mathbf{U}_t^1 \sim \mathcal{N}(0, I_n)$. The new solution is evaluated on the objective function f and compared to $\mathbf{X}_t$. If it is better than $\mathbf{X}_t$, in this case we talk about success, it becomes $\mathbf{X}_{t+1}$, otherwise it is rejected:

$$\mathbf{X}_{t+1} = \mathbf{X}_t + \sigma_t \mathbf{U}_t^1 1_{\{f(\mathbf{X}_t^1) \leq f(\mathbf{X}_t)\}} \ . \tag{1.2}$$

As for the step-size, it is increased in case of success and decreased otherwise [16, 4, 15]. We denote $\gamma > 1$ the increasing factor and introduce a parameter $q \in \mathbb{R}_>^+$ such that the factor for decrease equals $\gamma^{-1/q}$. Overall the step-size update reads

$$\sigma_{t+1} = \sigma_t \gamma 1_{\{f(\mathbf{X}_t^1) \leq f(\mathbf{X}_t)\}} + \sigma_t \gamma^{-1/q} 1_{\{f(\mathbf{X}_t^1) > f(\mathbf{X}_t)\}} \ . \tag{1.3}$$

The idea to maintain a probability of success around $1/5$ was proposed in [16, 4, 15]. The constant $1/5$ is a trade-off between the asymptotic (in n) optimal success probability on the sphere function $f(\mathbf{x}) = \|\mathbf{x}\|^2$ where it is approximately 0.27 [16, 15] and the corridor function[1] [15]. One implementation of the update of the step-size with target probability of success of $1/5$ is to set $-1/q = -1/4$[2]. We call in the sequel the algorithm following equations (1.1), (1.2) and (1.3), a $(1+1)$-ES with generalized one-fifth success rule and sometimes in short $(1+1)$-ES as there is no ambiguity for this paper that the step-size mechanism adopted is the generalized one-fifth success rule.

CB-SARS algorithms are observed to typically converge linearly towards local optima on a wide class of functions. Linear convergence of single runs is illustrated in Figure 5.1 for the $(1+1)$-ES on the simple sphere function $f(\mathbf{x}) = \|\mathbf{x}\|^2$. We observe

[1] The corridor function is defined as $f(\mathbf{x}) = \mathbf{x}_1$ for $-b < \mathbf{x}_2 < b, \ldots - b < \mathbf{x}_n < b$ (where $b > 0$) otherwise $+\infty$.

[2] Assuming indeed a probability of success of $1/5$ and having set $q = 4$ we find that $E[\ln \sigma_{t+1} | \sigma_t] = \ln \sigma_t + \frac{1}{5} \ln \gamma + \frac{4}{5} \ln \gamma^{-1/4} = \ln \sigma_t$, i.e., the step-size is stationary.

that both the distance to the optimum $\|\mathbf{X}_t\|$ and the step-size σ_t converge linearly at the same rate, in the sense that the logarithm of $\|\mathbf{X}_t\|$ (or σ_t) divided by t converges to $-\mathrm{CR}$ (where $-\mathrm{CR}$ corresponds to the slope of the line observed in the second stage of the convergence).

Despite overwhelming empirical evidence of the linear convergence of CB-SARS and the fact that the methods are relatively old, few formal proofs of their linear convergence actually exist. A variant of the (1+1)-ES presented here was however studied by Jägersküpper[3] who proved on the sphere function and some convex-quadratic functions lower and upper bounds (on the time to reduce the error by a given fraction) that imply linear convergence [10, 9, 7, 8]. The linear convergence of another CB-SARS using so-called self-adaptation as step-size adaptation mechanism was also proven on the sphere function [1].

We study in this paper the global linear convergence of the (1+1)-ES on a class of unimodal functions. More precisely convergence is investigated on functions h that are the composition of a strictly increasing transformation g by a positively homogeneous function with degree α, f, i.e., satisfying $f(\rho(\mathbf{x}-\mathbf{x}^\star)) = \rho^\alpha f(\mathbf{x}-\mathbf{x}^\star)$ for any $\rho > 0$, $\alpha > 0$ and $\mathbf{x}^\star$ that is the global optimum of the function (we assume that f is strictly positive except in $\mathbf{x}^\star$ where it can be zero). This class of function is a subset of scaling-invariant functions [3].

Under the assumptions that f is continuously differentiable plus mild assumptions, we prove global linear convergence of the (1+1)-ES optimizing h provided $\gamma > 1$ and the condition $\frac{1}{2}\left(\frac{1}{\gamma^\alpha} + \gamma^{\alpha/q}\right) < 1$ is satisfied. (This latter condition translates that the step-size increases on a linear function in the sense that one over expected change to the α on a linear function is smaller 1.) More formally, under the conditions sketched above, assuming w.l.g. that $\mathbf{x}^\star$ is zero, we prove the existence of $\mathrm{CR} > 0$ such that from any initial condition $(\mathbf{X}_0, \sigma_0)$, almost surely

$$\frac{1}{t}\ln\frac{\|\mathbf{X}_t\|}{\|\mathbf{X}_0\|} \xrightarrow[t\to\infty]{} -\mathrm{CR} \text{ and } \frac{1}{t}\ln\frac{\sigma_t}{\sigma_0} \xrightarrow[t\to\infty]{} -\mathrm{CR}$$

hold. We provide a comprehensive expression for the convergence rate as

$$\mathrm{CR} = -\ln\gamma\left(\frac{q+1}{q}\mathrm{PS} - \frac{1}{q}\right)$$

where PS is the asymptotic probability of success. We also prove that in expectation from any initial condition $(\mathbf{X}_0, \sigma_0) = (\mathbf{x}, \sigma)$

$$E_{\substack{\mathbf{x}\\ \sigma}} \ln\frac{\|\mathbf{X}_{t+1}\|}{\|\mathbf{X}_t\|} \xrightarrow[t\to\infty]{} -\mathrm{CR} \text{ and } E_{\substack{\mathbf{x}\\ \sigma}} \ln\frac{\sigma_{t+1}}{\sigma_t} \xrightarrow[t\to\infty]{} -\mathrm{CR}\ .$$

We finally precise the speed of convergence for the step-size of the two previous equations. We prove a Central Limit Theorem associated to the first equation and then prove that $E_{\substack{\mathbf{x}\\ \sigma}} \ln\frac{\sigma_{t+1}}{\sigma_t}$ converges geometrically fast towards $-\mathrm{CR}$.

Our proof technique follows a methodology developed in [3] exploiting the fact that the $(1+1)$-ES is a scale-invariant CB-SARS and that thus linear convergence on scaling-invariant functions can be turned into the stability study of the homogeneous Markov chain $\mathbf{Z}_t = \mathbf{X}_t/\sigma_t$. More precisely we study the ψ-irreducibility, Harris-reccurence, positivity and geometric ergodicity of $(\mathbf{Z}_t)_{t\in\mathbb{N}}$. We use for this, standard

[3] In this variant, the step-size is kept constant for a period of several iterations before to increase or decrease depending on the observed probability of success during the period.

tools for the analysis of Monte Carlo Markov chains algorithms and in particular Foster-Lyapunov drift conditions [12].

This paper is organized as follows. In Section 2 we summarize the results from the companion paper [3] setting the framework for the theoretical analysis, i.e., allowing us to define the normalized Markov chain that needs to be studied for proving the convergence. In addition, we define the objective functions under study and set some first assumptions. In Section 3 we study the normalized chain $(\mathbf{Z}_t)_{t\in\mathbb{N}}$ namely its φ-irreducibility, aperiodicity, investigate small sets and prove its geometric ergodicity that constitutes the core part of the study. Using those results, we finally prove in Section 4 the global linear convergence of the $(1+1)$-ES with generalized one-fifth success rule almost surely and in expectation. We provide a comprehensive expression for the convergence rate. Last we discuss our findings in Section 5.

Notations. We denote $\mathcal{N}(0, I_n)$ a standard multivariate normal distribution, i.e., with mean vector 0 and covariance matrix identity. Its density is denoted $p_{\mathcal{N}}$. Given a set C we denote its complementary C^c. We denote $\mathbb{R}^n_{\neq}$ the set $\mathbb{R}^n$ minus the null vector and $\mathbb{R}^+_>$ denotes the set of strictly positive real numbers. The set of strictly increasing functions g from $\mathbb{R}$ to $\mathbb{R}$ or a subset of $\mathbb{R}$ to $\mathbb{R}$ is denoted $\mathcal{M}$. Given an objective function f we denote by $\mathcal{L}_c$ the level set $\{\mathbf{x}, f(\mathbf{x}) = c\}$ and by $\bar{\mathcal{L}}_c$ the corresponding sublevel set, i.e., $\bar{\mathcal{L}}_c = \{\mathbf{x}, f(\mathbf{x}) \leq c\}$. We denote by $\mathbb{N}$ the set of natural numbers including 0, i.e., $\mathbb{N} = \{0, 1, \ldots\}$ and $\mathbb{N}_>$ the set $\{1, 2, \ldots\}$. The euclidian norm of a vector $\mathbf{x}$ is denoted $\|\mathbf{x}\|$. A ball of center $\mathbf{x}$ and radius r is denoted $\mathbf{B}(\mathbf{x}, r)$.

2. Normalized Markov Chain and Objective Function Assumptions.

In this section we summarize the main results from [3] allowing to define on the class of scaling-invariant functions the normalized Markov chain $\mathbf{X}_t/\sigma_t$. The study of the stability of this latter chain will imply the global linear convergence of the $(1+1)$-ES with generalized one-fifth success rule.

2.1. The (1+1)-ES as a Comparison-Based Step-Size Adaptive Randomized Search.

We remind in this section that the $(1+1)$-ES is a CB-SARS after recalling the general definition of a step-size adaptive randomized search (SARS) and a CB-SARS.

A SARS algorithm is identified to a sequence of random vectors $(\mathbf{X}_t, \sigma_t)_{t\in\mathbb{N}}$ where $\mathbf{X}_t \in \mathbb{R}^n$ and $\sigma_t \in \mathbb{R}^+_>$. The vector $(\mathbf{X}_t, \sigma_t)$ is the state of the algorithm at iteration t and $\Omega = \mathbb{R}^n \times \mathbb{R}^+_>$ is its state space. Let $\mathbb{U}$ be a subset of $\mathbb{R}^m$ that is called sampling space and $\mathbb{U}^p = \mathbb{U} \times \ldots \times \mathbb{U}$ for $p \in \mathbb{N}_>$. Given $(\mathbf{X}_0, \sigma_0) \in \Omega$, the sequence $(\mathbf{X}_t, \sigma_t)$ is inductively defined via

$$(\mathbf{X}_{t+1}, \sigma_{t+1}) = \mathcal{F}^f((\mathbf{X}_t, \sigma_t), \mathbf{U}_t)$$

where $\mathcal{F}$ is a measurable function and $(\mathbf{U}_t)_{t\in\mathbb{N}}$ is an independent and identically distributed (i.i.d.) sequence of random vectors of $\mathbb{U}^p$. The objective function f is also an input argument to the update function $\mathcal{F}$, however fixed over time, hence it is denoted as upper-script of $\mathcal{F}$. A comparison-based SARS is a particular case of SARS where candidate solutions are (i) sampled from $(\mathbf{X}_t, \sigma_t)$ using a solution function $\mathcal{S}ol$, (ii) evaluated on the objective function and ordered. The order of the candidate solutions is then *solely* used for updating the state $(\mathbf{X}_t, \sigma_t)$ of the algorithm. Formally, let us define first the solution function and ordering function.

DEFINITION 2.1 ($\mathcal{S}ol$ function). *A $\mathcal{S}ol$ function used to create candidate solutions is a measurable function mapping $\Omega \times \mathbb{U}$ into $\mathbb{R}^n$, i.e.,*

$$\mathcal{S}ol : \Omega \times \mathbb{U} \mapsto \mathbb{R}^n \ .$$

DEFINITION 2.2 ($\mathcal{O}rd$ function). *The ordering function $\mathcal{O}rd$ maps $\mathbb{R}^p$ to $\mathfrak{S}(p)$, the set of permutations with p elements and returns for any set of real values $(f_1, \ldots, f_p)$ the permutation of ordered indexes. That is $\mathcal{S} = \mathcal{O}rd(f_1, \ldots, f_p) \in \mathfrak{S}(p)$ where*

$$f_{\mathcal{S}(1)} \leq \ldots \leq f_{\mathcal{S}(p)} \ .$$

When more convenient we might denote $\mathcal{O}rd((f_i)_{i=1,\ldots,p})$ instead of $\mathcal{O}rd(f_1, \ldots, f_p)$. When needed for the sake of clarity, we might use the notations $\mathcal{O}rd^f$ or $\mathcal{S}^f$ to emphasize the dependency in f.

Given a permutation $\mathcal{S} \in \mathfrak{S}(p)$, the star operator $*$ defines the action of $\mathcal{S}$ on the coordinates of a vector $\mathbf{U} = (\mathbf{U}^1, \ldots, \mathbf{U}^p)$ belonging to $\mathbb{U}^p$ as

$$\begin{aligned} \mathfrak{S}(p) \times \mathbb{U}^p \to & \mathbb{U}^p \\ (\mathcal{S}, \mathbf{U}) \mapsto & \mathcal{S} * \mathbf{U} = \left(\mathbf{U}^{\mathcal{S}(1)}, \ldots, \mathbf{U}^{\mathcal{S}(p)}\right) \ . \end{aligned} \tag{2.1}$$

A CB-SARS can now be defined using a solution function $\mathcal{S}ol$, the ordering function and the star operator.

DEFINITION 2.3 (CB-SARS minimizing $f : \mathbb{R}^n \to \mathbb{R}$). *Let $p \in \mathbb{N}_>$ and $\mathbb{U}^p = \mathbb{U} \times \ldots \times \mathbb{U}$ where $\mathbb{U}$ is a subset of $\mathbb{R}^m$. Let $p_{\mathbf{U}}$ be a probability distribution defined on $\mathbb{U}^p$ where each $\mathbf{U}$ distributed according to $p_{\mathbf{U}}$ has a representation $(\mathbf{U}^1, \ldots, \mathbf{U}^p)$ (each $\mathbf{U}^i \in \mathbb{U}$). Let $\mathcal{S}ol$ be a solution function as in Definition 2.1. Let $\mathcal{G}_1 : \Omega \times \mathbb{U}^p \mapsto \mathbb{R}^n$ and $\mathcal{G}_2 : \mathbb{R}^+_> \times \mathbb{U}^p \mapsto \mathbb{R}^+$ be two mesurable mappings and denote $\mathcal{G} = (\mathcal{G}_1, \mathcal{G}_2)$.*

A CB-SARS is determined by the quadruplet $(\mathcal{S}ol, \mathcal{G}, \mathbb{U}^p, p_{\mathbf{U}})$ from which the recursive sequence $(\mathbf{X}_t, \sigma_t) \in \mathbb{R}^n \times \mathbb{R}^+_>$ is defined via $(\mathbf{X}_0, \sigma_0) \in \mathbb{R}^n \times \mathbb{R}^+_>$ and for all t:

$$\mathbf{X}_t^i = \mathcal{S}ol((\mathbf{X}_t, \sigma_t), \mathbf{U}_t^i), i = 1, \ldots, p \tag{2.2}$$

$$\mathcal{S} = \mathcal{O}rd(f(\mathbf{X}_t^1), \ldots, f(\mathbf{X}_t^p)) \in \mathfrak{S}(p) \tag{2.3}$$

$$\mathbf{X}_{t+1} = \mathcal{G}_1\left((\mathbf{X}_t, \sigma_t), \mathcal{S} * \mathbf{U}_t\right) \tag{2.4}$$

$$\sigma_{t+1} = \mathcal{G}_2\left(\sigma_t, \mathcal{S} * \mathbf{U}_t\right) \tag{2.5}$$

where $(\mathbf{U}_t)_{t \in \mathbb{N}}$ is an i.i.d. sequence of random vectors on $\mathbb{U}^p$ distributed according to $p_{\mathbf{U}}$, $\mathcal{O}rd$ is the ordering function as in Definition 2.2.

In the next lemma we state that the $(1+1)$-ES with generalized one-fifth success rule is a CB-SARS and define its different components. The proof is immediate and hence omitted.

LEMMA 2.4. *The $(1+1)$-ES with generalized one-fifth success rule satisfies Definition 2.3 with $p = 2$, $\mathbb{U} = \mathbb{R}^n$, $\mathbb{U}^p = \mathbb{R}^n \times \mathbb{R}^n$. Its solution function equals to*

$$\mathcal{S}ol : ((\mathbf{x}, \sigma), \mathbf{u}) \in (\mathbb{R}^n \times \mathbb{R}^+_>) \times \mathbb{U}) \mapsto \mathbf{x} + \sigma \mathbf{u} \ .$$

The sampling distribution is $p_{\mathbf{U}}(\mathbf{u}_1, \mathbf{u}_2) = p_{\mathcal{N}}(\mathbf{u}_1)\delta_0(\mathbf{u}_2)$ where $p_{\mathcal{N}}$ is the density of a standard multivariate normal distribution and δ_0 is the Dirac-delta function. The update function $\mathcal{G} = (\mathcal{G}_1, \mathcal{G}_2)$ equals

$$\mathcal{G}((\mathbf{x}, \sigma), \mathbf{y}) = \begin{pmatrix} \mathcal{G}_1((\mathbf{x}, \sigma), \mathbf{y}) \\ \mathcal{G}_2(\sigma, \mathbf{y}) \end{pmatrix} = \begin{pmatrix} \mathbf{x} + \sigma \mathbf{y}^1 \\ \sigma\left((\gamma - \gamma^{-1/q})1_{\{\mathbf{y}^1 \neq 0\}} + \gamma^{-1/q}\right) \end{pmatrix} \ . \tag{2.6}$$

The solution and update functions associated to the $(1+1)$-ES have a specific structure that is useful for proving invariance properties of the algorithm. We state those properties in the following lemma and omit the proof which is also immediate.

LEMMA 2.5. *Let $(\mathcal{S}ol, \mathcal{G}, \mathbb{U}^p, p_{\mathbf{U}})$ be the quadruplet associated to the CB-SARS $(1+1)$-ES with generalized one-fifth success rule. Then the following properties are satisfied:*
For all $\mathbf{x}, \mathbf{x}_0 \in \mathbb{R}^n$ for all $\sigma > 0$, for all $\mathbf{u} \in \mathbb{U}$, $\mathbf{y} \in \mathbb{U}^p$

$$\mathcal{S}ol((\mathbf{x}+\mathbf{x}_0, \sigma), \mathbf{u}) = \mathcal{S}ol((\mathbf{x}, \sigma), \mathbf{u}) + \mathbf{x}_0 \tag{2.7}$$

$$\mathcal{G}_1((\mathbf{x}+\mathbf{x}_0, \sigma), \mathbf{y}) = \mathcal{G}_1((\mathbf{x}, \sigma), \mathbf{y}) + \mathbf{x}_0 \ . \tag{2.8}$$

For all $\alpha > 0$, $(\mathbf{x}, \sigma) \in \Omega$, $\mathbf{u}^i \in \mathbb{U}$, $\mathbf{y} \in \mathbb{U}^p$

$$\mathcal{S}ol((\mathbf{x}, \sigma), \mathbf{u}^i) = \alpha \mathcal{S}ol\left(\left(\frac{\mathbf{x}}{\alpha}, \frac{\sigma}{\alpha}\right), \mathbf{u}^i\right) \tag{2.9}$$

$$\mathcal{G}_1((\mathbf{x}, \sigma), \mathbf{y}) = \alpha \mathcal{G}_1\left(\left(\frac{\mathbf{x}}{\alpha}, \frac{\sigma}{\alpha}\right), \mathbf{y}\right) \tag{2.10}$$

$$\mathcal{G}_2(\sigma, \mathbf{y}) = \alpha \mathcal{G}_2\left(\frac{\sigma}{\alpha}, \mathbf{y}\right) \ . \tag{2.11}$$

2.2. Invariances. As a direct consequence of the fact that the $(1+1)$-ES is comparison based, it is invariant to monotonically increasing transformations of the objective function. That is, for any $g \in \mathcal{M}$, the sequence $(\mathbf{X}_t, \sigma_t)_{t\in\mathbb{N}}$ optimizing $g \circ f$ or optimizing f are almost surely equal (see Proposition 2.4 in [3]). In addition the $(1+1)$-ES is translation and scale-invariant as detailed below.

Translation invariance implies identical behavior on a function $h(\mathbf{x})$ or any of its translated version $\mathbf{x} \mapsto h(\mathbf{x} - \mathbf{x}_0)$. It is formally defined for a SARS using a group homomorphism from the group $(\mathbb{R}^n, +)$ to the group $(\mathcal{A}(\Omega), \circ)$, set of invertible mappings from the state space Ω to itself endowed with the function composition $\circ$. More precisely, a SARS is translation invariant if there exists a group homomorphism $\Phi \in \mathrm{Homo}((\mathbb{R}^n, +), (\mathcal{A}(\Omega), \circ))$ such that for any objective function f, for any $\mathbf{x}_0 \in \mathbb{R}^n$, for any $(\mathbf{x}, \sigma) \in \Omega$ and for any $\mathbf{u} \in \mathbb{U}^p$

$$\mathcal{F}^{f(\mathbf{x})}((\mathbf{x}, \sigma), \mathbf{u}) = \underbrace{[\Phi(\mathbf{x}_0)]^{-1}}_{\Phi(-\mathbf{x}_0)} \left(\mathcal{F}^{f(\mathbf{x}-\mathbf{x}_0)}(\Phi(\mathbf{x}_0)(\mathbf{x}, \sigma), \mathbf{u})\right) \ . \tag{2.12}$$

The (1+1)-ES is translation invariant and the group homomorphism associated equals $\Phi(\mathbf{x}_0)(\mathbf{x}, \sigma) = (\mathbf{x} + \mathbf{x}_0, \sigma)$. This property is a consequence of (2.7) and (2.8) (see Proposition 2.7 in [3]). Similarly, scale-invariance that translates that an algorithm has no intrinsic notion of scale is defined via homomorphisms from the group $(\mathbb{R}^+_>, .)$ (where . denotes the multiplication in $\mathbb{R}$) to the group $(\mathcal{A}(\Omega), \circ)$. More precisely a SARS is scale-invariant if there exists an homomorphism $\Phi \in \mathrm{Homo}((\mathbb{R}^+_>, .), (\mathcal{A}(\Omega), \circ))$ such that for any f, for any $\alpha > 0$, for any $(\mathbf{x}, \sigma) \in \Omega$ and for any $\mathbf{u} \in \mathbb{U}^p$

$$\mathcal{F}^{f(\mathbf{x})}((\mathbf{x}, \sigma), \mathbf{u}) = \underbrace{[\Phi(\alpha)]^{-1}}_{\Phi(1/\alpha)} \left(\mathcal{F}^{f(\alpha\mathbf{x})}(\Phi(\alpha)(\mathbf{x}, \sigma), \mathbf{u})\right) \ . \tag{2.13}$$

The $(1+1)$-ES is scale-invariant and the group homomorphism associated equals $\Phi(\alpha)(\mathbf{x}, \sigma) = (\mathbf{x}/\alpha, \sigma/\alpha)$. This property is a consequence of the properties (2.10) and (2.11) (see Proposition 2.9 in [3]).

2.3. Normalized Markov Chain on Scaling-Invariant Functions. A class of functions that plays a specific role for CB-SARS are scaling invariant functions defined as: for all $\rho > 0$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$f(\rho(\mathbf{x} - \mathbf{x}^\star)) \leq f(\rho(\mathbf{y} - \mathbf{x}^\star)) \Leftrightarrow f(\mathbf{x} - \mathbf{x}^\star) \leq f(\mathbf{y} - \mathbf{x}^\star) \ , \tag{2.14}$$

where $\mathbf{x}^\star \in \mathbb{R}^n$. The latter function is said scaling-invariant w.r.t. $\mathbf{x}^\star$. A linear function or any $g \circ f$ where f is a norm and $g \in \mathcal{M}$ are scaling-invariant. Also some non quasi-convex functions are scaling-invariant. Scaling-invariant functions are essentially unimodal, formally they do not admit any strict local extrema (see Proposition 3.2 in [3]).

We assume given a scaling-invariant function w.r.t. $\mathbf{x}^\star = 0$ (w.l.g.). Then, for a translation and scale-invariant CB-SARS defined by the quadruplet $(\mathcal{S}ol, \mathcal{G}, \mathbb{U}^p, p_{\mathbf{U}})$ where scale-invariance is a consequence of the properties (2.9), (2.10) and (2.11), the normalized sequence $(\mathbf{X}_t/\sigma_t)_{t\in\mathbb{N}}$ is an homogeneous Markov chain (Proposition 4.1 in [3]). This Markov chain can be defined independently of $(\mathbf{X}_t, \sigma_t)$ provided $\mathbf{Z}_0 = \frac{\mathbf{X}_0}{\sigma_0}$ via

$$\mathbf{Z}_t^i = \mathcal{S}ol((\mathbf{Z}_t, 1), \mathbf{U}_t^i), i = 1, \ldots, p \tag{2.15}$$

$$\mathcal{S} = \mathcal{O}rd(f(\mathbf{Z}_t^1), \ldots, f(\mathbf{Z}_t^p)) \tag{2.16}$$

$$\mathbf{Z}_{t+1} = G(\mathbf{Z}_t, \mathcal{S} * \mathbf{U}_t) \tag{2.17}$$

where the transition function G equals for all $\mathbf{z} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{U}^p$

$$G(\mathbf{z}, \mathbf{y}) = \frac{\mathcal{G}_1((\mathbf{z}, 1), \mathbf{y})}{\mathcal{G}_2(1, \mathbf{y})} \ . \tag{2.18}$$

According to the previous equation, the transition function G for the normalized chain $(\mathbf{Z}_t = \frac{\mathbf{X}_t}{\sigma_t})_{t\in\mathbb{N}}$ associated to the $(1+1)$-ES on scaling-invariant functions is given by

$$G(\mathbf{z}, \mathbf{y}) = \frac{\mathbf{z} + \mathbf{y}^1}{(\gamma - \gamma^{-1/q})1_{\{\mathbf{y}^1 \neq 0\}} + \gamma^{-1/q}}$$

where the selected step $\mathbf{y} = (\mathbf{y}^1, \mathbf{y}^2)$ is according to the f-ranking of the solutions $\mathbf{z}+\mathbf{u}^1$ and $\mathbf{z}+\mathbf{u}^2$, i.e., $f(\mathbf{z}+\mathbf{y}^1) \leq f(\mathbf{z}+\mathbf{y}^2)$. However, since $\mathbf{u}^2 = 0$ (because $p_{\mathbf{U}}(\mathbf{u}^1, \mathbf{u}^2) = p_{\mathcal{N}}(\mathbf{u}^1)\delta_0(\mathbf{u}^2)$), $\mathbf{y}^1 = \mathbf{u}^1 1_{\{f(\mathbf{z}+\mathbf{u}^1) \leq f(\mathbf{z})\}}$. In addition since $\mathbf{U}_t^1 \sim \mathcal{N}(0, I_n)$, the event $\{\mathbf{Y}_t^1 \neq 0\}$ is almost surely equal to the event $\{\mathbf{Y}_t^1 = \mathbf{U}_t^1\}$ and hence almost surely equal to the event $\{f(\mathbf{Z}_t + \mathbf{U}_t^1) \leq f(\mathbf{Z}_t)\}$. Overall the Markov chain $(\mathbf{Z}_t)_{t\in\mathbb{N}}$ satisfies $\mathbf{Z}_0 = \frac{\mathbf{X}_0}{\sigma_0}$ and given $(\mathbf{U}_t^1)_{t\in\mathbb{N}}$ i.i.d with $\mathbf{U}_t^1 \sim \mathcal{N}(0, I_n)$

$$\boxed{\mathbf{Z}_{t+1} = \frac{\mathbf{Z}_t + \mathbf{U}_t^1 1_{\{f(\mathbf{Z}_t + \mathbf{U}_t^1) \leq f(\mathbf{Z}_t)\}}}{(\gamma - \gamma^{-1/q})1_{\{f(\mathbf{Z}_t + \mathbf{U}_t^1) \leq f(\mathbf{Z}_t)\}} + \gamma^{-1/q}} \ .} \tag{2.19}$$

Following [3], we introduce the notation $\eta^\star$ for the step-size change, i.e.,

$$\eta^\star = (\gamma - \gamma^{-1/q})1_{\{f(\mathbf{X}_t^1) \leq f(\mathbf{X}_t)\}} + \gamma^{-1/q} \tag{2.20}$$

and remind that on scaling-invariant functions, the step-size change starting from $(\mathbf{X}_t, \sigma_t)$ is the same as the step-size change starting from $(\mathbf{Z}_t, 1)$ (see Eq. (4.7) in [3]) such that

$$\mathbf{Z}_{t+1} = \frac{\mathbf{Z}_t + \mathbf{U}_t^1 1_{\{f(\mathbf{Z}_t + \mathbf{U}_t^1) \leq f(\mathbf{Z}_t)\}}}{\eta^\star} \ . \tag{2.21}$$

2.4. Objective Function Assumptions. We consider scaling-invariant functions formally defined by (2.14) where in addition we assume that $\mathbf{x}^\star = 0$. This can be done w.l.g. because the $(1+1)$-ES is translation invariant. This assumption is sufficient to build the normalized Markov chain $(\mathbf{Z}_t)_{t\in\mathbb{N}}$ (see Section 2.3). However for studying its stability, we will make further hypothesis on f.

We will consider a particular class of scaling-invariant functions, namely positively homogeneous functions. Formally a positively homogeneous function with degree α satisfies the following definition.

DEFINITION 2.6. *[Positively homogeneous functions] A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is said positively homogeneous with degree α if for all $\rho > 0$ and for all $\mathbf{x} \in \mathbb{R}^n$, $f(\rho\mathbf{x}) = \rho^\alpha f(\mathbf{x})$.*

Remark that positive homogeneity is not always preserved if f is composed by a non-decreasing transformation. We will in addition make the following assumptions on the objective function:

ASSUMPTION 1. *The function $f : \mathbb{R}^n \to \mathbb{R}$ is homogeneous with degree α and $f(\mathbf{x}) > 0$ for all $\mathbf{x} \neq 0$.*

This assumption implies that the function f has a *unique optimum* located w.l.g. in 0 (if the optimum $\mathbf{x}^\star$ is not in 0, consider $\tilde{f} = f(\mathbf{x} - \mathbf{x}^\star)$) as seen in the next lemma point (i). Remark that with this assumption, we exclude linear functions.

In the next lemma, we state some properties of positive homogeneous functions satisfying Assumptions 1. We denote for $c \geq 0$, $\bar{\mathcal{L}}_c = \{\mathbf{x}, f(\mathbf{x}) \leq c\}$ the sublevel set of f associated to c and $\mathcal{L}_c = \{\mathbf{x}, f(\mathbf{x}) = c\}$ its level set. The hypersphere surface of radius r is denoted $\mathbb{S}_r$, that is $\mathbb{S}_r = \{\mathbf{x}, \|\mathbf{x}\| = r\}$.

LEMMA 2.7. *Let f be an homogeneous function with degree $\alpha > 0$ and $f(\mathbf{x}) > 0$ for all $\mathbf{x} \neq 0$ and $f(\mathbf{x})$ finite for every $\mathbf{x} \in \mathbb{R}^n$. Then the following holds:*

(i) $\lim_{t\to 0} f(t\mathbf{x}) = 0$ and assuming that $f(\mathbf{0}) = 0$, for all $\mathbf{s} \neq 0$, the function $f_\mathbf{s} : t \in [0, +\infty[\mapsto f(t\mathbf{s})$ is continuous, strictly increasing and converges to $+\infty$ when t goes to $+\infty$.

(ii) If f is lower semi-continuous, then $\bar{\mathcal{L}}_c$ is compact.

Proof. (i) Since $f(t\mathbf{x}) = t^\alpha f(\mathbf{x})$, fixing $\mathbf{x}$ and taking the limit for t to zero we have that $\lim_{t\to 0} f(t\mathbf{x}) = 0$. For any $\mathbf{s}$, the function $f_\mathbf{s}$ satisfies $f_\mathbf{s}(t) = t^\alpha f(\mathbf{s})$. It is thus continuous on $[0, +\infty[$, strictly increasing and converges to infinity when t goes to infinity.

(ii) Since f is lower semi continuous, the inverse image of sets of the form $(-\infty, r]$ are closed sets. Hence $\bar{\mathcal{L}}_c = f^{-1}((-\infty, c])$ is closed. Let us consider the surface $\mathbb{S}_r$ for $r > 0$. Since f is lower semi-continuous, there exists $\mathbf{x}_0 \in \mathbb{S}_r$ such that $\inf_{\mathbf{x}\in\mathbb{S}_r} f(\mathbf{x}) = f(\mathbf{x}_0)$. Since $f(\mathbf{x}) > 0$ for $\mathbf{x} \neq 0$, $f(\mathbf{x}_0) := m \neq 0$. Hence we have $\bar{\mathcal{L}}_m \subset B_r$. Because f is homogeneous with degree α, we have thus $\bar{\mathcal{L}}_{m\sigma^\alpha} \subset B_{r\sigma}$ for all $\sigma > 0$. Hence, for any c we can include $\bar{\mathcal{L}}_c$ is a ball which proves that it is bounded and hence compact. ∎

A positively homogeneous function satisfies for all $\mathbf{x} \neq 0$

$$f(\mathbf{x}) = \|\mathbf{x}\|^\alpha f\left(\mathbf{x}/\|\mathbf{x}\|\right) \quad . \tag{2.22}$$

From this latter relation it follows that f is continuous on $\mathbb{S}_1 = \{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\| = 1\}$ if and only if f is continuous on $\mathbb{R}^n_{\neq}$. Assuming continuity on $\mathbb{R}^n_{\neq}$, we denote in the sequel m the minimum of f on $\mathbb{S}_1$ and M its maximum, that is

$$m = \min_{\mathbf{z}\in\mathbb{S}_1} f(\mathbf{z}) \tag{2.23}$$

$$M = \max_{\mathbf{z}\in\mathbb{S}_1} f(\mathbf{z}) \quad . \tag{2.24}$$

The following lemma will be used several times when investigating the stability of the normalized Markov chain $\mathbf{Z} = (\mathbf{Z}_t)_{t\in\mathbb{N}}$.

LEMMA 2.8. *Let f satisfy Assumptions 1 and f be continuous on $\mathbb{S}_1$. Then for all $\mathbf{z} \neq 0$*

$$\|\mathbf{z}\| m^{1/\alpha} \leq f(\mathbf{z})^{1/\alpha} \leq \|\mathbf{z}\| M^{1/\alpha} \ , \tag{2.25}$$

where m and M are defined in (2.23) and (2.24). Hence, $f(\mathbf{z}) \to 0$ when $\|\mathbf{z}\| \to 0$, $f(\mathbf{z}) \to \infty$ when $\|\mathbf{z}\|$ goes to ∞ and $|\ln \|\mathbf{z}\||f(\mathbf{z})^{1/\alpha} \to 0$ when $\|\mathbf{z}\| \to 0$.

Proof. By homogeneity, for all $\mathbf{z} \neq 0$, we have $f(\mathbf{z}) = f\left(\|\mathbf{z}\| \frac{\mathbf{z}}{\|\mathbf{z}\|}\right) = \|\mathbf{z}\|^\alpha f\left(\frac{\mathbf{z}}{\|\mathbf{z}\|}\right)$. Since f is continuous on the compact $\mathbb{S}_1$, $m = \min_{\mathbf{z}\in\mathbb{S}_1} f(\mathbf{z}) > 0$ and $M = \max_{\mathbf{z}\in\mathbb{S}_1} f(\mathbf{z}) > 0$ and $M < \infty$. We hence have

$$\|\mathbf{z}\|^\alpha \underbrace{\min_{\mathbf{z}\in\mathbb{S}_1} f(\mathbf{z})}_{m} \leq f(\mathbf{z}) \leq \|\mathbf{z}\|^\alpha \underbrace{\max_{\mathbf{z}\in\mathbb{S}_1} f(\mathbf{z})}_{M}$$

and thus $\|\mathbf{z}\| m^{1/\alpha} \leq f(\mathbf{z})^{1/\alpha} \leq \|\mathbf{z}\| M^{1/\alpha}$. Since $\|\mathbf{z}\||\ln\|\mathbf{z}\|| \to 0$ when $\|\mathbf{z}\| \to 0$, we hence obtain that $|\ln\|\mathbf{z}\||f(\mathbf{z})^{1/\alpha} \to 0$. □

The following lemma is a consequence of the previous one and will be useful in the sequel.

LEMMA 2.9. *Let f satisfy Assumptions 1 and f be continuous on $\mathbb{S}_1$, for all $\rho > 0$, the ball centered in 0 and of radius ρ is included in the sublevel set of degree $\rho^\alpha M$, i.e.,*

$$\mathbf{B}(0,\rho) \subset \bar{\mathcal{L}}_{\rho^\alpha M} \ . \tag{2.26}$$

For all $K > 0$, the sublevel set of degree K is included into the ball centered in 0 and of radius $(K/m)^\alpha$, i.e.,

$$\bar{\mathcal{L}}_K \subset \mathbf{B}(0,(K/m)^\alpha) \ . \tag{2.27}$$

Proof. From Lemma 2.8 we have that for all $\mathbf{z}$, $m\|\mathbf{z}\|^\alpha \leq f(\mathbf{z}) \leq M\|\mathbf{z}\|^\alpha$. Let $\mathbf{z} \in \mathbf{B}(0,\rho)$, then $f(\mathbf{z}) \leq M\rho^\alpha$, i.e., $\mathbf{z} \in \bar{\mathcal{L}}_{\rho^\alpha M}$. Let $\mathbf{z} \in \bar{\mathcal{L}}_K$, then $f(\mathbf{z}) \leq K$ and hence $\|\mathbf{z}\| \leq (K/m)^\alpha$. □

Last, we remind the Euler's homogeneous function theorem.

THEOREM 2.10 (Euler's homogeneous function theorem). *Suppose that the function $f : \mathbb{R}^n\backslash\{0\} \mapsto \mathbb{R}$ is continuously differentiable. Then f is positive homogeneous of degree α if and only if*

$$\mathbf{x} \cdot \nabla f(\mathbf{x}) = \alpha f(\mathbf{x}) \ .$$

This theorem implies that if f is positively homogeneous and continuously differentiable, if $f(\mathbf{x}) > 0$ for $\mathbf{x} \neq 0$ (i.e., Assumption 1), then

$$\nabla f(\mathbf{x}) \neq 0 \text{ for all } \mathbf{x} \neq 0 \ . \tag{2.28}$$

2.5. State Space for the Normalized Markov Chain. The state-space for the normalized Markov chain $\mathbf{Z} = (\mathbf{Z}_t)_{t\in\mathbb{N}}$ is *a priori* $\mathbb{R}^n$. However if we start from $\mathbf{Z}_0 = 0$, we will stay in 0 forever, i.e., $\mathbf{Z}_t = 0$ for all t. This is due to the fact that the $(1+1)$-ES cannot accept worse solutions and 0 is the global optimum of f. This would then preclude the chain $\mathbf{Z}$ to be irreducible w.r.t. to a non-singular measure. We therefore exclude 0 from the state space that is now equal to $\mathcal{Z} = \mathbb{R}^n_{\neq}$.

3. Study of the Normalized Chain. We study in this section different properties of the homogeneous Markov chain $\mathbf{Z} = (\mathbf{Z}_t)_{t\in\mathbb{N}}$ defined in Section 2.3. Those properties will imply linear convergence of the $(1+1)$-ES as we will see in Section 4. We start in the next section by expressing the transition kernel of the Markov chain.

3.1. Transition Probability Kernel. We follow standard notations and terminology for a time homogeneous Markov chain $(\mathbf{Z}_t)_{t\in\mathbb{N}}$ on a topological space $\mathcal{Z}$. The Borel sets of $\mathcal{Z}$ are denoted $\mathcal{B}(\mathcal{Z})$. A kernel T is any function on $\mathcal{Z}\times\mathcal{B}(\mathcal{Z})$ such that $T(.,A)$ is measurable for all $A\in\mathcal{B}(\mathcal{Z})$ and $T(\mathbf{z},.)$ is a measure for all $\mathbf{z}\in\mathcal{Z}$. The transition probability kernel for $(\mathbf{Z}_t)_{t\in\mathbb{N}}$ is a kernel P such that $P(.,A)$ is a non-negative measurable function for all $A\in\mathcal{B}(\mathcal{Z})$ and the measure $P(\mathbf{z},.)$ for all $\mathbf{z}$ is a probability measure. It is defined as

$$P(\mathbf{z},A) = P_{\mathbf{z}}(\mathbf{Z}_1\in A)\ ,$$

where $P_{\mathbf{z}}$ denotes the probability law of the chain under the initial condition $\mathbf{Z}_0=\mathbf{z}$. Similarly $E_{\mathbf{z}}$ denotes the expectation of the chain under the initial condition $\mathbf{Z}_0=\mathbf{z}$. If a probability μ on $(\mathcal{Z},\mathcal{B}(\mathcal{Z}))$ is the initial distribution, the probability law and expectation under μ are denoted P_μ and E_μ. The *n-step transition probability law* is defined iteratively by setting $P^0(\mathbf{z},A)=\delta_{\mathbf{z}}(A)$ and for $t\geq 1$, inductively by

$$P^t(\mathbf{z},A) = \int P(\mathbf{z},d\mathbf{y})P^{t-1}(\mathbf{y},A)\ .$$

The relation $P^t(\mathbf{z},A) = P_{\mathbf{z}}(\mathbf{Z}_t\in A)$ holds. With an abuse of notations similar to [12, p 56], we will also for instance denote $\Pr(\mathcal{N}\in A)$ or $\Pr(f(\mathbf{z}+\mathcal{N})>f(\mathbf{z}))$ for the probability of the events $\{\mathcal{N}\in A\}$, $\{f(\mathbf{z}+\mathcal{N})>f(\mathbf{z})\}$ (where $\mathcal{N}$ will typically be a standard normal multivariate distribution) without specifically defining the space where $\mathcal{N}$ exists which could be the space where $\mathbf{Z}$ is defined or another space. Similarly $E[1_{\{f(\mathbf{z}+\mathcal{N})>f(\mathbf{z})\}}]$ will be used for the expectation of the random variable $1_{\{f(\mathbf{z}+\mathcal{N})>f(\mathbf{z})\}}$.

We derive in the next proposition an expression for the transition kernel of $(\mathbf{Z}_t)_{t\in\mathbb{N}}$ when f is a scaling-invariant function.

PROPOSITION 3.1. *Let $f:\mathbb{R}^n\mapsto\mathbb{R}$ be a scaling-invariant function and let $\mathbf{Z}$ be the Markov chain defined in (2.19). Its transition probability kernel is given for all $\mathbf{z}\in\mathcal{Z}=\mathbb{R}^n_{\neq}$ and $A\in\mathcal{B}(\mathcal{Z})$ by*

$$P(\mathbf{z},A) = \int 1_A(\mathbf{u})q(\mathbf{z},\mathbf{u})d\mathbf{u} + 1_A(\mathbf{z}\gamma^{1/q})\Pr(f(\mathbf{z}+\mathcal{N})>f(\mathbf{z})) \tag{3.1}$$

where $q(\mathbf{z},\mathbf{u}) = 1_{\{f(\mathbf{u})\leq f(\mathbf{z}/\gamma)\}}(\mathbf{u})p_{\mathcal{N}}(\gamma\mathbf{u}-\mathbf{z})\gamma$ with $p_{\mathcal{N}}$ the density of a standard multivariate normal distribution and $\mathcal{N}\sim\mathcal{N}(0,I_n)$.

Proof. Given $\mathbf{Z}_0=\mathbf{z}$, and $\mathbf{U}_0^1=\mathcal{N}$ where $\mathcal{N}\sim\mathcal{N}(0,I_n)$, $\mathbf{Z}_1$ satisfies $\mathbf{Z}_1 = \frac{\mathbf{z}+\mathcal{N}1_{\{f(\mathbf{z}+\mathcal{N})\leq f(\mathbf{z})\}}}{(\gamma-\gamma^{-1/q})1_{\{f(\mathbf{z}+\mathcal{N})\leq f(\mathbf{z})\}}+\gamma^{-1/q}}$. Hence, the transition probability kernel satisfies $P(\mathbf{z},A) =$

$$\Pr\left(\frac{\mathbf{z}+\mathcal{N}}{\gamma}\in A\ \cap\{f(\mathbf{z}+\mathcal{N})\leq f(\mathbf{z})\}\right) + \Pr\left(\frac{\mathbf{z}}{\gamma^{-1/q}}\in A\cap\{f(\mathbf{z}+\mathcal{N})>f(\mathbf{z})\}\right)\ ,$$

and thus satisfies

$$\begin{aligned}P(\mathbf{z},A) &= \int 1_A\left(\frac{\mathbf{z}+\mathbf{u}}{\gamma}\right)1_{\{f(\mathbf{z}+\mathbf{u})\leq f(\mathbf{z})\}}(\mathbf{u})p_{\mathcal{N}}(\mathbf{u})d\mathbf{u} + 1_A(\mathbf{z}\gamma^{1/q})\Pr(f(\mathbf{z}+\mathcal{N})>f(\mathbf{z}))\\ &= \int 1_A(\bar{\mathbf{u}})\underbrace{1_{\{f(\gamma\bar{\mathbf{u}})\leq f(\mathbf{z})\}}(\bar{\mathbf{u}})p_{\mathcal{N}}(\gamma\bar{\mathbf{u}}-\mathbf{z})\gamma}_{q(\mathbf{z},\bar{\mathbf{u}})}d\bar{\mathbf{u}} + 1_A(\mathbf{z}\gamma^{\frac{1}{q}})\Pr(f(\mathbf{z}+\mathcal{N})>f(\mathbf{z})).\square\end{aligned}$$

3.2. Irreducibility, Small Sets and Aperiodicity. A Markov Chain $\mathbf{Z} = (\mathbf{Z}_t)_{t\in\mathbb{N}}$ on a state space $\mathcal{Z}$ is said φ-irreducible if there exists a measure φ on $\mathcal{Z}$ such that for all $A \in \mathcal{B}(\mathcal{Z})$, $\varphi(A) > 0$ implies that $P_{\mathbf{z}}(\tau_A < \infty) > 0$ for all $\mathbf{z} \in \mathcal{Z}$ where $\tau_A = \min\{t > 0 : \mathbf{Z}_t \in A\}$ is the first return time to A. Another equivalent definition is for all $\mathbf{z} \in \mathcal{Z}$ and for all $A \in \mathcal{B}(\mathcal{Z})$

$$\varphi(A) > 0 \Rightarrow \exists\ t_{\mathbf{z},A} \in \mathbb{N} \text{ such that } P_{\mathbf{z}}(\mathbf{Z}_{t_{\mathbf{z},A}} \in A) > 0\ .$$

Given that a chain $\mathbf{Z}$ is φ-irreducible, there exists a maximal irreducibility measure ψ and all maximal irreducibility measure are equivalent (see [12, Proposition 4.4.2]). The set of positive ψ-measure is denoted

$$\mathcal{B}^+(\mathcal{Z}) := \{A \in \mathcal{B}(\mathcal{Z}) : \psi(A) > 0\}\ .$$

In the sequel we continue to denote ψ the maximal irreducibility measure and hence if $\mathbf{Z}$ is ψ-irreducible it means that it is φ-irreducible for some φ and that ψ is a maximal irreducibility measure. A set A is *full* if $\psi(A^c) = 0$ and *absorbing* if $P(\mathbf{z}, A) = 1$ for $\mathbf{z} \in A$. In addition, a set C is a *small set* if there exists $t \in \mathbb{N}$ and a non-trivial measure ν_t on $\mathcal{B}(\mathcal{Z})$ such that for all $\mathbf{z} \in C$

$$P^t(\mathbf{z}, A) \geq \nu_t(A)\,, A \in \mathcal{B}(\mathcal{Z})\ . \tag{3.2}$$

The small set is then called a ν_t-small set. Consider a small set C satisfying the previous equation with $\nu_t(C) > 0$ and denote $\nu_t = \nu$. The chain is called aperiodic if the g.c.d. of the set

$$E_C = \{k \geq 1 : C \text{ is a } \nu_k\text{-small set with } \nu_k = \alpha_k\nu \text{ for some } \alpha_k > 0\}$$

is one for some (and then for every) small set C.

We establish now the φ-irreducibility, identify some small sets and show the aperiodicity of the normalized chain associated to the $(1 + 1)$-ES.

3.2.1. φ-irreducibility. We denote μ_{Leb} the Lebesgue measure on $\mathcal{Z} = \mathbb{R}^n_{\neq}$. We prove in the next proposition that the normalized MC associated to the $(1 + 1)$-ES is irreducible with respect to the Lebesgue measure.

PROPOSITION 3.2. *Assume that f satisfies Assumptions 1 and is continuous on $\mathbb{R}^n_{\neq}$. Assume that $\gamma > 1$. Then, the Markov chain $\mathbf{Z}$ associated to the $(1 + 1)$-ES is irreducible w.r.t. the Lebesgue measure μ_{Leb}.*

Proof. Let $\mathbf{z} \in \mathcal{Z}$ and $\bar{\mathcal{L}}^f_{f(\mathbf{z}/\gamma)}$ be the sublevel set $\{\mathbf{u} \in \mathcal{Z}, f(\mathbf{u}) \leq f(\mathbf{z}/\gamma)\} = \{\mathbf{u} \in \mathcal{Z}, f(\gamma\mathbf{u}) \leq f(\mathbf{z})\}$. And let $A \in \mathcal{B}(\mathcal{Z})$ such that $\mu_{\text{Leb}}(A) > 0$. By the regularity of the Lebesgue measure we can include a compact K in A such that $\mu(K) > 0$. Since $K \subset A$, for all $\mathbf{z}$, $P(\mathbf{z}, A) \geq P(\mathbf{z}, K)$.

If (i) $K \subset \bar{\mathcal{L}}^f_{f(\mathbf{z}/\gamma)}$ then $P(\mathbf{z}, K) = \int_K \gamma p_{\mathcal{N}}(\gamma\mathbf{u} - \mathbf{z})d\mathbf{u} > 0$ as $\mathbf{u} \in \mathbb{R}^n \mapsto p_{\mathcal{N}}(\gamma\mathbf{u} - \mathbf{z}) > 0$.

If (ii) K is not included in $\bar{\mathcal{L}}^f_{f(\mathbf{z}/\gamma)}$, then by sampling $\mathcal{N}$ such that $f(\mathbf{z} + \mathcal{N}) > f(\mathbf{z})$ (which happens with strictly positive probability since by Lemma 2.7, $\bar{\mathcal{L}}^f_{f(\mathbf{z})}$ is bounded, hence sampling such that $f(\mathbf{z} + \mathcal{N}) > f(\mathbf{z})$ can be achieved by sampling outside a ball), $\mathbf{Z}_1 = \mathbf{z}\gamma^{1/q}$ which is at a larger distance from 0 (as we assumed that $\gamma > 1$). By repeating this, we build a sequence $\mathbf{Z}_t = \mathbf{z}\gamma^{t/q}$ and $f(\mathbf{Z}_t) = (\gamma^{t/q})^\alpha f(\mathbf{z})$ hence $f(\mathbf{Z}_t)$ and $f(\mathbf{Z}_t/\gamma)$ go to ∞. The set K being compact we can find a ball $\mathbf{B}(0, \rho)$ such that $K \subset \mathbf{B}(0, \rho)$. In addition from Lemma 2.9, we know that for all

ρ, $\mathbf{B}(0,\rho) \subset \bar{\mathcal{L}}_{\rho^\alpha M}$, hence choosing t large enough such that $\bar{\mathcal{L}}^f_{\rho^\alpha M} \subset \bar{\mathcal{L}}^f_{f(\mathbf{Z}_t/\gamma)}$, we have that $K \subset \bar{\mathcal{L}}^f_{f(\mathbf{Z}_t/\gamma)}$ and by (i) $P(\mathbf{Z}_t = \mathbf{z}\gamma^{t/q}, K) > 0$. Thus $P^{t+1}(\mathbf{z}, A) \geq P^{t+1}(\mathbf{z}, K) \geq \Pr(f(\mathbf{z}+\mathcal{N}) > f(\mathbf{z})) \dots \Pr\left(f(\mathbf{z}\gamma^{\frac{t-1}{q}} + \mathcal{N}) > f(\mathbf{z}\gamma^{\frac{t-1}{q}})\right) P(\mathbf{z}\gamma^{t/q}, K) \geq \theta^t P(\mathbf{z}\gamma^{t/q}, K) > 0$ where $\theta > 0$ is a lower bound on $\bar{\mathbf{z}} \to \Pr(f(\bar{\mathbf{z}} + \mathcal{N}) > f(\bar{\mathbf{z}}))$ on a ball that includes $\bar{\mathcal{L}}^f_{f(\mathbf{Z}_t)}$. Indeed, since $f(\mathbf{Z}_k)$ increases, for all $k \leq t$, $f(\mathbf{Z}_k) \in \bar{\mathcal{L}}^f_{f(\mathbf{Z}_t)}$. However according to Lemma 2.9, there exists R such that $\bar{\mathcal{L}}^f_{f(\mathbf{Z}_t)} \subset B(0,R)$. Then $\{\mathcal{N} \in B(0,2R)^c\} \subset \{f(\bar{\mathbf{z}} + \mathcal{N}) > f(\bar{\mathbf{z}})\}$ for all $\bar{\mathbf{z}}$ in $\bar{\mathcal{L}}^f_{f(\mathbf{Z}_t)}$. We can take $\theta = \Pr(\mathcal{N} \in B(0,2R)^c)$. □

3.2.2. Small Sets and Aperiodicity. We investigate small sets for the (1+1)-ES assuming that f is positively homogeneous with degree α with $f(\mathbf{x}) > 0$ for $\mathbf{x} \neq 0$ and f is continuous on $\mathbb{R}^n_{\neq}$. Consider sets $D_{[l_1,l_2]}$ with $0 < l_1 < l_2$ defined as

$$D_{[l_1,l_2]} := \{\mathbf{z} \in \mathcal{Z}, l_1 \leq f(\mathbf{z}) \leq l_2\} \ . \tag{3.3}$$

Because f is continuous, the sets $D_{[l_1,l_2]} = f^{-1}([l_1,l_2])$ are closed and by Lemma 2.9 they are also bounded such that the sets $D_{[l_1,l_2]}$ are compact sets. We prove in this section that the sets $D_{[l_1,l_2]}$ are small sets for the Markov chain $\mathbf{Z}$.

LEMMA 3.3. *Assume that f is positively homogeneous with degree α, $f(\mathbf{x}) > 0$ for $\mathbf{x} \neq 0$ and f is continuous on $\mathbb{R}^n_{\neq}$. Assume that $\gamma > 1$. Let $D_{[l_1,l_2]}$ be a set of the type (3.3) with $0 < l_1 < l_2$. Let $t_0 \geq 1$ and let $R > 0$ such that $\bar{\mathcal{L}}_{\gamma^{\frac{\alpha t_0}{q}} l_2} \subset B(0,R)$ (see Lemma 2.9). Then for all $\mathbf{z}$ in $D_{[l_1,l_2]}$ and for all $t \leq t_0$*

$$\Pr\left(f(\mathbf{z}\gamma^{\frac{t}{q}} + \mathcal{N}) > f\left(\mathbf{z}\gamma^{\frac{t}{q}}\right)\right) \geq \Pr\left(\mathcal{N} \in B(0,2R)^c\right) =: \theta_{3.3} \tag{3.4}$$

where $\mathcal{N} \sim \mathcal{N}(0, I_n)$. For all $\mathbf{z} \in D_{[l_1,l_2]}$ and $t \leq t_0$, the following minorization holds, for $A \in \mathcal{B}(\mathcal{Z})$

$$P^{t+1}(\mathbf{z}, A) \geq \theta^t_{3.3} \gamma \delta_t \int 1_{A \cap D_{[l_1,l_2]}}(\mathbf{u}) 1_{\{f(\gamma\mathbf{u}) \leq \gamma^{t\alpha/q} l_1\}}(\mathbf{u}) d\mathbf{u} =: \nu_{t+1}(A) \ , \tag{3.5}$$

where $\delta_t = \min_{(\mathbf{z},\mathbf{u}) \in D^2_{[l_1,l_2]}} p_{\mathcal{N}}(\gamma\mathbf{u} - \mathbf{z}\gamma^{t/q}) > 0$. In addition ν_{t+1} is a non-trivial measure if $t > q$ and hence $D_{[l_1,l_2]}$ is a ν_{t+1}-small set provided $t > q$.

Proof. Note first that for all $t \leq t_0$, for all $\mathbf{z} \in D_{[l_1,l_2]}$, $\bar{\mathcal{L}}_{f(\mathbf{z}\gamma^{t/q})} \subset \bar{\mathcal{L}}_{\gamma^{\alpha t_0/q} l_2}$ (we use here that $\gamma > 1$). We now claim that if $\mathcal{N} \in B(0,2R)^c$, then $f(\mathbf{z}\gamma^{t/q} + \mathcal{N}) > f\left(\mathbf{z}\gamma^{t/q}\right)$. Indeed, if $\mathcal{N} \in B(0,2R)^c$, then for all $\mathbf{z}$ in $D_{[l_1,l_2]}$, $\mathbf{z}\gamma^{\frac{t}{q}}$ is inside $\bar{\mathcal{L}}_{f(\mathbf{z}\gamma^{t/q})} \subset \bar{\mathcal{L}}_{\gamma^{\alpha t_0/q} l_2} \subset B(0,R)$ (by definition of R) and hence $\mathbf{z}\gamma^{\frac{t}{q}} + \mathcal{N}$ will be outside $B(0,R)$, i.e., outside $\bar{\mathcal{L}}_{f(\mathbf{z}\gamma^{t/q})}$, i.e., $f(\mathbf{z}\gamma^{t/q} + \mathcal{N}) > f\left(\mathbf{z}\gamma^{t/q}\right)$. Hence (3.4). We will now prove (3.5). We lower bound the probability $P^{t+1}(\mathbf{z}, A) = P_{\mathbf{z}}(\mathbf{Z}_{t+1} \in A)$ by the probability to reach A in $t+1$ steps starting from $\mathbf{z}$ by having no success for the first $t-1$ iterations:

$$P^{t+1}(\mathbf{z}, A) \geq \underbrace{P_{\mathbf{z}}\left(\{\mathbf{Z}_{t+1} \in A\} \cap \{f(\mathbf{Z}_t + \mathbf{U}^1_t) \leq f(\mathbf{Z}_t)\} \bigcap_{k=0}^{t-1} \{f(\mathbf{Z}_k + \mathbf{U}^1_k) > f(\mathbf{Z}_k)\}\right)}_{A_1} \ .$$

However if $\{f(\mathbf{Z}_k + \mathbf{U}_k^1) > f(\mathbf{Z}_k)\}$ then $\mathbf{Z}_{k+1} = \mathbf{Z}_k \gamma^{1/q}$ such that given that $\mathbf{Z}_0 = \mathbf{z}$, the following equalities between events holds

$$\{\mathbf{Z}_{t+1} \in A\} \cap \{f(\mathbf{Z}_t + \mathbf{U}_t^1) \leq f(\mathbf{Z}_t)\} \bigcap_{k=0}^{t-1} \{f(\mathbf{Z}_k + \mathbf{U}_k^1) > f(\mathbf{Z}_k)\} =$$

$$\left\{ \frac{\mathbf{z}\gamma^{t/q} + \mathbf{U}_t^1}{\gamma} \in A \right\} \cap \left\{ f(\mathbf{z}\gamma^{t/q} + \mathbf{U}_t^1) \leq f(\mathbf{z}\gamma^{t/q}) \right\} \bigcap_{k=0}^{t-1} \{f(\mathbf{z}\gamma^{k/q} + \mathbf{U}_k^1) > f(\mathbf{z}\gamma^{k/q})\} \Bigg\} \ .$$

Hence by independence of the $(\mathbf{U}_k)_{0 \leq k \leq t}$, $A_1 =$

$$P_{\mathbf{z}} \left(\frac{\mathbf{z}\gamma^{t/q} + \mathbf{U}_t^1}{\gamma} \in A \cap \{f(\mathbf{z}\gamma^{\frac{t}{q}} + \mathbf{U}_t^1) \leq f(\mathbf{z}\gamma^{\frac{t}{q}})\} \right) \prod_{k=0}^{t-1} P_{\mathbf{z}} \left(f(\mathbf{z}\gamma^{\frac{k}{q}} + \mathbf{U}_k^1) > f(\mathbf{z}\gamma^{\frac{k}{q}}) \right)$$

using now (3.4)

$$\begin{aligned}
&\geq P_{\mathbf{z}} \left(\frac{\mathbf{z}\gamma^{t/q} + \mathbf{U}_t^1}{\gamma} \in A \cap \{f(\mathbf{z}\gamma^{t/q} + \mathbf{U}_t^1) \leq f(\mathbf{z}\gamma^{t/q})\} \right) \theta_{3.3}^t \\
&= \theta_{3.3}^t \int 1_A \left(\frac{\mathbf{z}\gamma^{t/q} + \mathbf{u}}{\gamma} \right) 1_{\{f(\mathbf{z}\gamma^{t/q} + \mathbf{u}) \leq f(\mathbf{z}\gamma^{t/q})\}}(\mathbf{u}) p_{\mathcal{N}}(\mathbf{u}) d\mathbf{u} \\
&= \theta_{3.3}^t \int 1_A(\bar{\mathbf{u}})\, 1_{\{f(\gamma\bar{\mathbf{u}}) \leq f(\mathbf{z}\gamma^{t/q})\}}(\bar{\mathbf{u}}) p_{\mathcal{N}}(\gamma\bar{\mathbf{u}} - \mathbf{z}\gamma^{t/q}) \gamma d\bar{\mathbf{u}} \\
&\geq \theta_{3.3}^t \int 1_{A \cap D_{[l_1,l_2]}}(\bar{\mathbf{u}})\, 1_{\{f(\gamma\bar{\mathbf{u}}) \leq f(\mathbf{z}\gamma^{t/q})\}}(\bar{\mathbf{u}}) p_{\mathcal{N}}(\gamma\bar{\mathbf{u}} - \mathbf{z}\gamma^{t/q}) \gamma d\bar{\mathbf{u}} \ .
\end{aligned}$$

For all $\mathbf{z} \in D_{[l_1,l_2]}$, $\bar{\mathbf{u}}$, $1_{\{f(\gamma\bar{\mathbf{u}}) \leq f(\mathbf{z}\gamma^{t/q})\}}(\bar{\mathbf{u}}) = 1_{\{f(\gamma\bar{\mathbf{u}}) \leq \gamma^{t\alpha q} f(\mathbf{z})\}}(\bar{\mathbf{u}}) \geq 1_{\{f(\gamma\bar{\mathbf{u}}) \leq \gamma^{t\alpha q} l_1\}}(\bar{\mathbf{u}})$ and for all $(\mathbf{z}, \bar{\mathbf{u}}) \in D_{[l_1,l_2]}$, $p_{\mathcal{N}}(\gamma\bar{\mathbf{u}} - \mathbf{z}\gamma^{t/q}) \geq \min_{(\mathbf{z},\bar{\mathbf{u}}) \in D^2_{[l_1,l_2]}} p_{\mathcal{N}}(\gamma\bar{\mathbf{u}} - \mathbf{z}\gamma^{t/q})$. Since $D_{[l_1,l_2]}$ is compact, $\{\gamma\bar{\mathbf{u}} - \mathbf{z}\gamma^{t/q}, (\mathbf{z}, \bar{\mathbf{u}}) \in D^2_{[l_1,l_2]}\}$ is also compact and thus there exists δ_t such that $\min_{(\mathbf{z},\mathbf{u}) \in D^2_{[l_1,l_2]}} p_{\mathcal{N}}(\gamma\bar{\mathbf{u}} - \mathbf{z}\gamma^{t/q}) \geq \delta_t > 0$. Hence

$$P^{t+1}(\mathbf{z}, A) \geq A_1 \geq \theta_{3.3}^t \delta_t \gamma \int 1_{A \cap D_{[l_1,l_2]}}(\mathbf{u}) 1_{\left\{f(\gamma\mathbf{u}) \leq \gamma^{t\alpha/q} l_1\right\}}(\mathbf{u}) d\mathbf{u}$$

which is a non-trivial measure if $t > q$. □

Remark that the constant $\theta_{3.3}$ defined in (3.4) and used in (3.5) depends on t_0 as the radius R of the ball where the sublevel set $\bar{\mathcal{L}}_{\gamma^{\alpha t_0/q} l_2}$ is included depends on t_0. To prove the aperiodicity of the chain we construct a joint minorization measure ν working for two consecutive integers having hence 1 as greatest common divisor and that satisfies $\nu(D_{[l_1,l_2]}) > 0$. More precisely we prove the following proposition.

PROPOSITION 3.4. *Assume that f is positively homogeneous with degree α, $f(\mathbf{x}) > 0$ for $\mathbf{x} \neq 0$ and f is continuous on $\mathbb{R}^n_{\neq}$. Assume that $\gamma > 1$. Let $D_{[l_1,l_2]}$ be a set of the type (3.3). Let $\bar{q} = \lfloor q \rfloor + 1$. Then for all $\mathbf{z} \in D_{[l_1,l_2]}$, $A \in \mathcal{B}(\mathcal{Z})$*

$$P^{\bar{q}+1}(\mathbf{z}, A) \geq \zeta^{\bar{q}} \nu(A) \tag{3.6}$$

$$P^{\bar{q}+2}(\mathbf{z}, A) \geq \zeta^{\bar{q}+1} \nu(A) \tag{3.7}$$

where ν is the measure defined by $\nu(A) = \delta\gamma \int 1_{A \cap D_{[l_1,l_2]}}(\mathbf{u}) 1_{\left\{f(\gamma\mathbf{u}) \leq \gamma^{\bar{q}\alpha/q} l_1\right\}}(\mathbf{u}) d\mathbf{u}$ with $\delta = \min\{\delta_{\bar{q}}, \delta_{\bar{q}+1}\}$ and ζ is the constant $\theta_{3.3}$ in (3.4) for $t_0 = \bar{q} + 2$. In addition $\nu(D_{[l_1,l_2]}) > 0$ which implies that the chain $\mathbf{Z}$ is aperiodic.

Proof. From Lemma 3.3

$$P^{\bar{q}+1}(\mathbf{z}, A) \geq \zeta^{\bar{q}} \delta_{\bar{q}} \gamma \int 1_{A \cap D_{[l_1,l_2]}}(\mathbf{u}) 1_{\{f(\gamma \mathbf{u}) \leq \gamma^{\bar{q}\alpha/q} l_1\}}(\mathbf{u}) d\mathbf{u} \ , \tag{3.8}$$

$$P^{\bar{q}+2}(\mathbf{z}, A) \geq \zeta^{\bar{q}+1} \gamma \delta_{\bar{q}+1} \int 1_{A \cap D_{[l_1,l_2]}}(\mathbf{u}) 1_{\{f(\gamma \mathbf{u}) \leq \gamma^{(\bar{q}+1)\alpha/q} l_1\}}(\mathbf{u}) d\mathbf{u} \ . \tag{3.9}$$

Since $\delta \leq \delta_{\bar{q}}$ we find using (3.8) that

$$P^{\bar{q}+1}(\mathbf{z}, A) \geq \zeta^{\bar{q}} \delta \gamma \int 1_{A \cap D_{[l_1,l_2]}}(\mathbf{u}) 1_{\{f(\gamma \mathbf{u}) \leq \gamma^{\bar{q}\alpha/q} l_1\}}(\mathbf{u}) d\mathbf{u},$$

which is exactly (3.6). Using in addition the fact that $1_{\{f(\gamma \mathbf{u}) \leq \gamma^{(\bar{q}+1)\alpha/q} l_1\}}(\mathbf{u}) \geq 1_{\{f(\gamma \mathbf{u}) \leq \gamma^{\bar{q}\alpha/q} l_1\}}(\mathbf{u})$ that we inject in (3.9), we find (3.7). Since $\bar{q}/q > 1$, $\nu(D_{[l_1,l_2]}) > 0$. As two consecutive integers have 1 as g.c.d., the g.c.d. of $\bar{q}+1$ and $\bar{q}+2$ is one and hence the chain $\mathbf{Z}$ is aperiodic. □

3.3. Geometric Ergodicity. In this section we derive the geometric ergodicity of the chain $\mathbf{Z}$. Geometric ergodicity will imply the other stability properties needed, namely positivity and Harris recurrence whose definitions are reminded below. First, let us recall that a σ-finite measure π on $\mathcal{B}(\mathcal{Z})$ with the property

$$\pi(A) = \int_{\mathcal{Z}} \pi(d\mathbf{z}) P(\mathbf{z}, A), \ A \in \mathcal{B}(\mathcal{Z})$$

is called invariant. A φ-irreducible chain admitting an invariant probability measure is called a *positive* chain. Harris recurrence is a concept ensuring that a chain visits the state space sufficiently often. It is defined for a ψ-irreducible chain as: A ψ-irreducible Markov chain is *Harris-recurrent* if for all $A \subset \mathcal{Z}$ with $\psi(A) > 0$, and for all $\mathbf{z} \in \mathcal{Z}$, the chain will eventually reach A with probability 1 starting from $\mathbf{z}$, formally if $P_{\mathbf{z}}(\eta_A = \infty) = 1$ where η_A be the *occupation time* of A, i.e., $\eta_A = \sum_{t=1}^{\infty} 1_{\mathbf{Z}_t \in A}$. An (Harris-)recurrent chain admits an unique (up to a constant multiples) invariant measure [12, Theorem 10.0.4].

For a function $V \geq 1$, the V-norm for a signed measure ν is defined as

$$\|\nu\|_V = \sup_{k:|k| \leq V} |\nu(k)| = \sup_{k:|k| \leq V} \left| \int k(\mathbf{y}) \nu(d\mathbf{y}) \right| \ .$$

Geometric ergodicity translates the fact that convergence to the invariant measure takes place at a geometric rate. Different notions of geometric ergodicity do exist (see [12]) and we will consider the form that appears in the following theorem. For any V, PV is defined as $PV(\mathbf{z}) := \int P(\mathbf{z}, d\mathbf{y}) V(\mathbf{y})$.

THEOREM 3.5. *(Geometric Ergodic Theorem [12, Theorem 15.0.1]) Suppose that the chain $\mathbf{Z}$ is ψ-irreducible and aperiodic. Then the following three conditions are equivalent: (i) The chain $\mathbf{Z}$ is positive recurrent with invariant probability measure π, and there exists some petite set $C \in \mathcal{B}^+(\mathcal{Z})$, $\rho_C < 1$ and $M_C < \infty$ and $P^{\infty}(C) > 0$ such that for all $\mathbf{z} \in C$*

$$|P^t(\mathbf{z}, C) - P^{\infty}(C)| \leq M_C \rho_C^t.$$

(ii) There exists some petite set C and $\kappa > 1$ such that

$$\sup_{\mathbf{z} \in C} E_{\mathbf{z}}[\kappa^{\tau_C}] < \infty \ .$$

(iii) There exists a petite set $C \in \mathcal{B}(\mathcal{Z})$, constants $b < \infty$, $\vartheta < 1$ and a function $V \geq 1$ finite at some one $\mathbf{z}_0 \in \mathcal{Z}$ satisfying

$$PV(\mathbf{z}) \leq \vartheta V(\mathbf{z}) + b1_C(\mathbf{z}), \mathbf{z} \in \mathcal{Z}. \tag{3.10}$$

Any of these three conditions imply that the following two statements hold. The set $S_V = \{\mathbf{z} : V(\mathbf{z}) < \infty\}$ is absorbing and full, where V is any solution to (3.10). Furthermore, there exist constants $r > 1$, $R < \infty$ such that for any $\mathbf{z} \in S_V$

$$\sum_t r^t \|P^t(\mathbf{z}, .) - \pi\|_V \leq RV(\mathbf{z}) \ . \tag{3.11}$$

The drift operator is defined as $\Delta V(\mathbf{z}) = PV(\mathbf{z}) - V(\mathbf{z})$. The inequality (3.10) is called a drift condition that can be re-written as

$$\Delta V(\mathbf{z}) \leq \underbrace{(\vartheta - 1)}_{<0} V(\mathbf{z}) + b1_C(\mathbf{z}) \ .$$

P is then said to admit a drift towards the set C. The previous theorem is using the notion of petite sets but small sets are actually also petite sets (see Section 5.5.2 [12]). We will in the sequel prove a geometric drift towards a small set C that will hence imply a geometric drift towards petite set. It will subsequently imply the existence of a probability invariant measure and Harris recurrence [12].

3.3.1. Geometric Drift Condition for Positively Homogenous Functions. In this section we investigate drift conditions for functions that are a monotonically increasing transformation of a positively homogeneous function, i.e., $h = g \circ f$ for f a positively homogeneous function with degree α and $g \in \mathcal{M}$.

We have shown that the sets $D_{[l_1,l_2]}$ are some small sets for $\mathbf{Z}$ (under the assumptions of Lemma 3.3). Hence proving negativity of the drift function outside a small set requires to prove negativity for $f(\mathbf{z})$ "large" as well as for $f(\mathbf{z})$ close to 0. We are going to prove that under some regularity assumptions on f, the function

$$V(\mathbf{z}) = f(\mathbf{z})1_{\{f(\mathbf{z})\geq 1\}} + \frac{1}{f(\mathbf{z})}1_{\{f(\mathbf{z})<1\}}$$

satisfies a geometric drift condition for the $(1 + 1)$-ES algorithm provided $\gamma > 1$ and the expected inverse of the step-size change to the α on *linear functions* is strictly smaller one that directly translates into:

$$\frac{1}{2}\left(\frac{1}{\gamma^\alpha} + \gamma^{\alpha/q}\right) < 1 \ .$$

Given the shape of the small sets proven in Section 3.2.2, to establish a geometric drift condition, it is enough to prove that the limit of PV/V is strictly smaller 1 when $\mathbf{z}$ goes to 0 and to ∞:

LEMMA 3.6. *Let f be positively homogeneous function with degree α and $f(\mathbf{x}) > 0$ for $\mathbf{x} \neq 0$ and f continuous on $\mathbb{R}^n_{\neq}$. Assume that $\gamma > 1$. Let V be a function finite at some one $\mathbf{z}_0 \in \mathcal{Z}$ and such that $V \geq 1$ that satisfies*

$$\lim_{\|\mathbf{z}\|\to\infty} \frac{PV(\mathbf{z})}{V(\mathbf{z})} < 1 \text{ and } \lim_{\|\mathbf{z}\|\to 0} \frac{PV(\mathbf{z})}{V(\mathbf{z})} < 1 \ . \tag{3.12}$$

Then V is a geometric drift in the sense of (3.10) for the $(1+1)$-ES.

Proof. According to Lemma 3.3, the sets $D_{[l_1,l_2]} = \{\mathbf{z} \in \mathcal{Z}, l_1 \leq f(\mathbf{z}) \leq l_2\}$ with $0 < l_1 < l_2$ are small sets for $\mathbf{Z}$. The limit (3.12) (left) gives that for all $\epsilon > 0$ small enough, there exists a_2 such that for $\|\mathbf{z}\| \geq a_2$, $PV(\mathbf{z}) \leq (1-\epsilon)V(\mathbf{z})$. According to (2.25) it implies the existence of l_2 such that for all $\mathbf{z}$ with $f(\mathbf{z}) \geq l_2$, $PV(\mathbf{z}) \leq (1-\epsilon)V(\mathbf{z})$. Similarly, the limit (3.12) (right) gives that for all $\epsilon > 0$ small enough, there exists $a_1 > 0$ such that for $\|\mathbf{z}\| \leq a_1$, $PV(\mathbf{z}) \leq (1-\epsilon)V(\mathbf{z})$. According to (2.25) it implies the existence of l_1 such that for all $\mathbf{z}$ with $f(\mathbf{z}) \leq l_1$, $PV(\mathbf{z}) \leq (1-\epsilon)V(\mathbf{z})$. Hence taking $\vartheta = 1-\epsilon$ for epsilon small enough, we have that outside the small set $D_{[l_1,l_2]}$, $PV(\mathbf{z}) \leq \vartheta V(\mathbf{z})$. □

Technical Results. Before to establish the main proposition of this section, we derive a few technical results.

LEMMA 3.7. *Assume that f is continuous on $\mathbb{R}^n$ and for all $\mathbf{n} \in \mathbb{R}^n$, $f(\mathbf{n}) > f(0)$ then $\lim_{\|\mathbf{z}\|\to 0} \Pr(f(\mathbf{z}+\mathcal{N}) > f(\mathbf{z})) = 1$, where $\mathcal{N} \sim \mathcal{N}(0, I_n)$.*

Proof. We express the probability $\Pr(f(\mathbf{z}+\mathcal{N}) > f(\mathbf{z}))$ using the density of $\mathcal{N}$:

$$\Pr(f(\mathbf{z}+\mathcal{N}) > f(\mathbf{z})) = \int 1_{\{f(\mathbf{z}+\mathbf{n})-f(\mathbf{z})>0\}}(\mathbf{n}) p_{\mathcal{N}}(\mathbf{n}) d\mathbf{n} \ .$$

For all $\mathbf{n}$ except for $\mathbf{n} = 0$, $1_{\{f(\mathbf{z}+\mathbf{n})-f(\mathbf{z})>0\}}(\mathbf{n})$ converges to $1_{\{f(\mathbf{n})-f(0)>0\}}(\mathbf{n})$ when $\mathbf{z}$ goes to 0 (the function $t \mapsto 1_{\{t>0\}}(t)$ being discontinuous in $t = 0$, for $\mathbf{n} = 0$, for $\mathbf{z}$ to 0, we arrive at the discontinuity point of the indicator, hence we cannot conclude about the limit). Hence by the dominated convergence theorem, we find that $\Pr(f(\mathbf{z}+\mathcal{N}) > f(\mathbf{z}))$ converges to $\Pr(f(\mathcal{N}) > f(0)) = 1$. □

LEMMA 3.8. *Assume that f is a positively homogeneous function with degree α and $f(\mathbf{x}) > 0$ for all $\mathbf{x} \neq 0$ and assume that f is continuously differentiable. Then*

$$\lim_{\|\mathbf{z}\|\to\infty} \Pr(f(\mathbf{z}+\mathcal{N}) > f(\mathbf{z})) = \frac{1}{2} \tag{3.13}$$

$$\lim_{\|\mathbf{z}\|\to\infty} \Pr\left(\{f(\mathbf{z}+\mathcal{N}) \leq f(\mathbf{z})\} \cap \{f(\mathbf{z}+\mathcal{N}) \geq \gamma^{\alpha}\}\right) = \frac{1}{2} \tag{3.14}$$

where $\mathcal{N} \sim \mathcal{N}(0, I_n)$.

Proof. We investigate first the limit (3.13) and want to prove that

$$\forall \epsilon, \exists\, T > 0, \text{ such that for all } \mathbf{z} \text{ with } \|\mathbf{z}\| > T, |\Pr(f(\mathbf{z}+\mathcal{N}) > f(\mathbf{z})) - \frac{1}{2}| < \epsilon \ . \tag{3.15}$$

Let us fix one arbitrary ϵ for the rest of the proof and use the homogeneity property to write $\Pr(f(\mathbf{z}+\mathcal{N}) > f(\mathbf{z})) = \Pr(f(\frac{\mathbf{z}}{f(\mathbf{z})^{1/\alpha}} + \frac{\mathcal{N}}{f(\mathbf{z})^{1/\alpha}}) > 1)$. Since f is continuously differentiable, the mean value theorem gives us the existence for all $\mathbf{n} \in \mathbb{R}^n$ of $c_{\mathbf{n}} \in [0,1]$ such that

$$f\left(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}} + \frac{\mathbf{n}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\right) = \underbrace{f\left(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\right)}_{=1} + \nabla f\left(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}} + c_{\mathbf{n}}\frac{\mathbf{n}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\right).\frac{\mathbf{n}}{f(\mathbf{z})^{\frac{1}{\alpha}}} \ . \tag{3.16}$$

The event $\{f(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}} + \frac{\mathcal{N}}{f(\mathbf{z})^{\frac{1}{\alpha}}}) > 1\}$ thus equals $\{\nabla f\left(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}} + c_{\mathcal{N}}\frac{\mathcal{N}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\right).\frac{\mathcal{N}}{f(\mathbf{z})^{\frac{1}{\alpha}}} > 0\}$ also equal to the event $\{\nabla f\left(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}} + c_{\mathcal{N}}\frac{\mathcal{N}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\right).\mathcal{N} > 0\}$. Let us define the function $g_{3.8} : \mathcal{L}_1 \times [0,1]$ as follows

$$g_{3.8}(\mathbf{u}, v) = \Pr\left(\nabla f(\mathbf{u} + c_{\mathcal{N}} v\mathcal{N}).\mathcal{N} > 0\right) \tag{3.17}$$

such that $\Pr(f(\mathbf{z}+\mathcal{N}) > f(\mathbf{z})) = g_{3.8}\left(\frac{\mathbf{z}}{f(\mathbf{z})^{1/\alpha}}, \frac{1}{f(\mathbf{z})^{1/\alpha}}\right)$. (Given the definition domain of $g_{3.8}$ we have assumed that $\mathbf{z}$ is large enough such that $f(\mathbf{z}) \geq 1$.) We now prove the continuity of $g_{3.8}$ that we express with its integral form as

$$g_{3.8}(\mathbf{u}, v) = \int 1_{\{\nabla f(\mathbf{u}+c_{\mathbf{n}} v\mathbf{n}).\mathbf{n}>0\}}(\mathbf{n}) p_{\mathcal{N}}(\mathbf{n}) d\mathbf{n} \ .$$

Because we have assumed that the differential of f is continuous, for all $\mathbf{n}$, the function $(\mathbf{u}, v) \mapsto \nabla f(\mathbf{u} + c_{\mathbf{n}} v\mathbf{n}).\mathbf{n}$ is continuous. The indicator function has one discontinuity point that could be reached if $\nabla f(\mathbf{u}+c_{\mathbf{n}} v\mathbf{n}).\mathbf{n} = 0$. We thus exclude the point $\mathbf{n} = 0$. In addition, with Property (2.28), $\nabla f(\mathbf{u} + c_{\mathbf{n}} v\mathbf{n}) \neq 0$ if $\mathbf{u} + c_{\mathbf{n}} v\mathbf{n} \neq 0$. Hence, given $(\mathbf{u}_0, v_0)$ where we want to prove the continuity of $g_{3.8}$, let $\mathfrak{N} = \{\mathbf{n} | \mathbf{n} = \alpha \mathbf{u}_0, \alpha \in \mathbb{R}\}$ (this is a set of null measure provided $n \geq 2$) then for all $\mathbf{n} \in \mathbb{R}^n \backslash \mathfrak{N}$, the function $(\mathbf{u}, v) \mapsto 1_{\{\nabla f(\mathbf{u}+c_{\mathbf{n}} v\mathbf{n}).\mathbf{n}>0\}}(\mathbf{n}) p_{\mathcal{N}}(\mathbf{n})$ is continuous in $(\mathbf{u}_0, v_0)$ and by the dominated convergence theorem we deduce the continuity of $g_{3.8}(\mathbf{u}, v)$ on $\mathcal{L}_1 \times [0, 1]$. By symmetry of $p_{\mathcal{N}}(\mathbf{n})$, for all $\mathbf{u}$,

$$g_{3.8}(\mathbf{u}, 0) = \int 1_{\{\nabla f(\mathbf{u}).\mathbf{n}>0\}}(\mathbf{n}) p_{\mathcal{N}}(\mathbf{n}) d\mathbf{n} = 1/2 \ .$$

Since $g_{3.8}$ is continuous on a compact, it is uniformly continuous and hence there exists $\beta > 0$ such that for all $v \leq \beta$, and $\mathbf{u} \in \mathcal{L}_1$, $|g_{3.8}(\mathbf{u}, v) - g_{3.8}(\mathbf{u}, 0)| \leq \epsilon$. Taking $T' = \frac{1}{\beta}$, we then have that if $f(\mathbf{z})^{1/\alpha} \geq T'$, $|g_{3.8}(\frac{\mathbf{z}}{f(\mathbf{z})^{1/\alpha}}, \frac{1}{f(\mathbf{z})^{1/\alpha}}) - \frac{1}{2}| \leq \epsilon$. From Lemma 2.8, we find that if $\|\mathbf{z}\| \geq T := T'/M^{1/\alpha}$, then $f(\mathbf{z})^{1/\alpha} \geq T'$ and $|g_{3.8}(\frac{\mathbf{z}}{f(\mathbf{z})^{1/\alpha}}, \frac{1}{f(\mathbf{z})^{1/\alpha}}) - \frac{1}{2}| \leq \epsilon$. Hence we have proven (3.15) that proves (3.13) in the case $n \geq 2$. The case $n = 1$ is even simpler and boils down to look at the limit when v goes to 0 of $\int 1_{\{f(\mathbf{u}+v\mathbf{n})>1\}}(\mathbf{n}) p(\mathbf{n}) d\mathbf{n} = \int_{\mathbb{R}^+} 1_{\{f'(\mathbf{u})+o(1)>0\}}(\mathbf{n}) p(\mathbf{n}) d\mathbf{n} + \int_{\mathbb{R}^-} 1_{\{-f'(\mathbf{u})+o(1)>0\}}(\mathbf{n}) p(\mathbf{n}) d\mathbf{n}$. Using the dominated convergence theorem, this latter limit equals $1/2$.

In a similar manner we investigate the limit (3.14). Using (3.16), we define in a similar manner on $\mathcal{L}_1 \times [0, 1($ the function

$$h_{3.8}(\mathbf{u}, v) = \int 1_{\{\nabla f(\mathbf{u}+c_{\mathbf{n}} v\mathbf{n}).\mathbf{n}<0\}}(\mathbf{n}) 1_{\{f(\mathbf{u}/v+\mathbf{n})\geq\gamma^\alpha\}}(\mathbf{n}) p_{\mathcal{N}}(\mathbf{n}) d\mathbf{n} \ .$$

Similarly we prove the continuity of $h_{3.8}$ on $\mathcal{L}_1 \times]0, 1($ and prolong it by continuity for $v = 0$ using the fact that for all $\mathbf{n}$, $\lim_{v\to 0} 1_{\{f(\mathbf{u}/v+\mathbf{n})\geq\gamma^\alpha\}} = 1$. We find then that

$$h_{3.8}(\mathbf{u}, 0) = \int 1_{\{\nabla f(\mathbf{u}).\mathbf{n}<0\}}(\mathbf{n}) p_{\mathcal{N}}(\mathbf{n}) d\mathbf{n} = \frac{1}{2} \ .$$

Like in the previous case we prove that for all ϵ, there exists $T > 0$ such that

$$\text{for all } \|\mathbf{z}\| > T, |\Pr(\{f(\mathbf{z}+\mathcal{N}) \leq f(\mathbf{z})\} \cap \{f(\mathbf{z}+\mathcal{N}) \geq \gamma^\alpha\}) - \frac{1}{2}| < \epsilon$$

that proves (3.14). We omit the details as the proof follows the same lines as before. □

LEMMA 3.9. *Let f be a positively homogeneous function with degree α satisfying $f(\mathbf{x}) > 0$ for all $\mathbf{x} \neq 0$. Assume that f is continuous on $\mathbb{R}^n_{\neq}$. Let $\mathcal{N}$ denote a standard multivariate normal distribution. Then for all $\mathbf{z}$*

$$\text{If } \alpha \leq 1, E[f(\mathbf{z}+\mathcal{N})] \leq M(\|\mathbf{z}\| + E(\|\mathcal{N}\|))^\alpha < +\infty \ , \tag{3.18}$$

$$\text{If } \alpha \geq 1, E[f(\mathbf{z}+\mathcal{N})] \leq M(\|\mathbf{z}\| + E[\|\mathcal{N}\|^\alpha]^{1/\alpha})^\alpha < +\infty \ . \tag{3.19}$$

If in addition, $\alpha \leq n$, then there exists a constant $c_{3.9}$ such that for all $\mathbf{z}$

$$E\left[\frac{1}{f(\mathbf{z}+\mathcal{N})}\right] < c_{3.9} \ . \tag{3.20}$$

Consequently if $\alpha \leq n$, the function $V(\mathbf{z}) = f(\mathbf{z})1_{\{f(\mathbf{z})\geq 1\}} + \frac{1}{f(\mathbf{z})}1_{\{f(\mathbf{z})<1\}}$ satisfies for all $\mathbf{z} \in \mathbb{R}^n_{\neq}$

$$\int V(\mathbf{y})P(\mathbf{z}, d\mathbf{y}) < \infty \ . \tag{3.21}$$

Proof. We start by proving (3.18) and (3.19). Note first that $E[\|\mathcal{N}\|^\alpha] < \infty$ for all $\alpha > 0$. According to Lemma 2.8, $f(\mathbf{z}+\mathcal{N}) \leq M\|\mathbf{z}+\mathcal{N}\|^\alpha$. From the triangle inequality we obtain

$$\|\mathbf{z}+\mathcal{N}\|^\alpha \leq (\|\mathbf{z}\| + \|\mathcal{N}\|)^\alpha \ . \tag{3.22}$$

For $\alpha < 1$, $x \in [0, +\infty] \mapsto x^\alpha$ being concave, we obtain from Jensen inequality that $E[(\|\mathbf{z}\| + \|\mathcal{N}\|)^\alpha] \leq (E[\|\mathbf{z}\| + E(\|\mathcal{N}\|)])^\alpha = (\|\mathbf{z}\| + E(\|\mathcal{N}\|))^\alpha$ which achieves to prove (3.18) (the case for $\alpha = 1$ being the equality case for the last equations). For $\alpha \geq 1$, we can apply the Minkowski inequality stating that $E[(\|\mathbf{z}\| + \|\mathcal{N}\|)^\alpha]^{1/\alpha} \leq E[\|\mathbf{z}\|^\alpha]^{1/\alpha} + E[\|\mathcal{N}\|^\alpha]^{1/\alpha} = \|\mathbf{z}\| + E[\|\mathcal{N}\|^\alpha]^{1/\alpha}$. Hence $E((\|\mathbf{z}\| + \|\mathcal{N}\|)^\alpha) \leq (\|\mathbf{z}\| + E[\|\mathcal{N}\|^\alpha]^{1/\alpha})^\alpha$. Overall using the upper bound on $f(\mathbf{z}+\mathcal{N})$ and (3.22), we find (3.19). We prove now (3.20). We are writing in the sequel integrals of positive functions that are possibly infinite. We will prove actually that the functions are integrable (and the integrals finite) and prove that we have a bound for the integral independent of $\mathbf{z}$. Using Lemma 2.8

$$A_{3.9} = E\left[\frac{1}{f(\mathbf{z}+\mathcal{N})}\right] \leq E\left[\frac{1}{m\|\mathbf{z}+\mathcal{N}\|^\alpha}\right] = \frac{1}{m}\int_{\mathbb{R}^n} \frac{1}{\|\mathbf{z}+\mathbf{y}\|^\alpha} p_{\mathcal{N}}(\mathbf{y})d\mathbf{y} \tag{3.23}$$

Using a change of variables

$$\int_{\mathbb{R}^n} \frac{1}{\|\mathbf{z}+\mathbf{y}\|^\alpha} p_{\mathcal{N}}(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^n} \frac{1}{\|\tilde{\mathbf{y}}\|^\alpha} p_{\mathcal{N}}(\tilde{\mathbf{y}}-\mathbf{z})d\tilde{\mathbf{y}} \ . \tag{3.24}$$

The previous integral is possibly infinite because the function $\tilde{\mathbf{y}} \mapsto \frac{1}{\|\tilde{\mathbf{y}}\|^\alpha}$ has a singularity in zero. Let us study the integrability close to zero, i.e., investigate

$$\int_{B(0,1)} \frac{1}{\|\mathbf{y}\|^\alpha} p_{\mathcal{N}}(\mathbf{y}-\mathbf{z})d\mathbf{y} \leq K_{3.9}\int_{B(0,1)} \frac{1}{\|\mathbf{y}\|^\alpha} d\mathbf{y}$$

where $K_{3.9}$ is an upper bound on the density $p_{\mathcal{N}}$ (hence independent of $\mathbf{z}$). Using spherical coordinates for $n \geq 2$

$$\int_{B(0,1)} \frac{1}{\|\mathbf{y}\|^\alpha} d\mathbf{y} = \int_0^1 \frac{r^{n-1}}{r^\alpha} dr \prod_{i=1}^{n-2} \int_0^\pi \sin^{n-1-i}(\varphi_i)d\varphi_i \int_0^{2\pi} d\varphi_{n-1}$$
$$\leq 2(\pi)^{n-1} \int_0^1 r^{n-1-\alpha} dr \ .$$

The latter integral is finite for $\alpha + 1 - n \leq 1$, i.e., $\alpha \leq n$. For $n = 1$ we directly obtain $\int_{-1}^{1} \frac{1}{|\mathbf{y}|^\alpha} d\mathbf{y} < \infty$ if $\alpha \leq 1$. To prove that $A_{3.9}$ is bounded for all $\mathbf{z}$ by a constant independent of $\mathbf{z}$, we write

$$\begin{aligned}\int_{\mathbb{R}^n} \frac{1}{\|\mathbf{y}\|^\alpha} p_{\mathcal{N}}(\mathbf{y}-\mathbf{z})d\mathbf{y} &\leq K_{3.9} \int_{\mathbb{R}^n} \frac{1}{\|\mathbf{y}\|^\alpha} 1_{\{\|\mathbf{y}\|\leq 1\}} d\mathbf{y} + \int_{\mathbb{R}^n} \frac{1}{\|\mathbf{y}\|^\alpha} 1_{\{\|\mathbf{y}\|\geq 1\}} p_{\mathcal{N}}(\mathbf{y}-\mathbf{z})d\mathbf{y} \\ &\leq K_{3.9} \int_{B(0,1)} \frac{1}{\|\mathbf{y}\|^\alpha} d\mathbf{y} + \underbrace{\int_{\mathbb{R}^n} p_{\mathcal{N}}(\mathbf{y}-\mathbf{z})d\mathbf{y}}_{=1}\end{aligned}$$

Hence $\int_{\mathbb{R}^n} \frac{1}{\|\mathbf{y}\|^\alpha} p_{\mathcal{N}}(\mathbf{y}-\mathbf{z})d\mathbf{y}$ is bounded by a constant independent of $\mathbf{z}$ if $\alpha \leq n$. Using this with (3.23) and (3.24) proves that $A_{3.9}$ is bounded by a constant independent of $\mathbf{z}$.

Finally we prove (3.21): Using the expression of the Markov chain given in (2.19) denoting $\mathcal{N}$ a standard normal distribution we find that $\int V(\mathbf{y})P(\mathbf{z}, d\mathbf{y}) =$

$$\begin{aligned}&E\left[f\left(\frac{\mathbf{z}+\mathcal{N}}{\gamma}\right) 1_{\{f(\mathbf{z}+\mathcal{N})\leq f(\mathbf{z})\}} 1_{\{f(\frac{\mathbf{z}+\mathcal{N}}{\gamma})\geq 1\}}\right] + E\left[\frac{1_{\{f(\mathbf{z}+\mathcal{N})\leq f(\mathbf{z})\}} 1_{\{f((\mathbf{z}+\mathcal{N})/\gamma)<1\}}}{f\left(\frac{\mathbf{z}+\mathcal{N}}{\gamma}\right)}\right] + \\ &E\left[f\left(\frac{\mathbf{z}}{\gamma^{-1/q}}\right) 1_{\{f(\mathbf{z}+\mathcal{N})>f(\mathbf{z})\}} 1_{\{f(\mathbf{z}/\gamma^{-\frac{1}{q}})\geq 1\}}\right] + E\left[\frac{1_{\{f(\mathbf{z}+\mathcal{N})>f(\mathbf{z})\}} 1_{\{f(\mathbf{z}/\gamma^{-1/q})<1\}}}{f(\mathbf{z}/\gamma^{-1/q})}\right] \\ &\leq \frac{1}{\gamma^\alpha} E\left[f(\mathbf{z}+\mathcal{N})\right] + \frac{1}{\gamma^{-\frac{\alpha}{q}}} f(\mathbf{z}) + \gamma^\alpha E\left[\frac{1}{f(\mathbf{z}+\mathcal{N})}\right] + \gamma^{-\frac{\alpha}{q}} \frac{1}{f(\mathbf{z})} \quad .\end{aligned}$$

Using now (3.18), (3.19) and (3.20) in the previous inequality we find (3.21). □

Sufficient conditions for geometric ergodicity. We are now ready to establish the main result of this section, namely a sufficient condition for geometric ergodicity. We need to make some further assumptions on the objective function that we gather as Assumption 2.

ASSUMPTION 2. *The function $f : \mathbb{R}^n \to [0, +\infty[$ satisfies Assumptions 1, i.e., is a positively homogeneous function with degree α and $f(\mathbf{x}) > 0$ for all $\mathbf{x} \neq 0$.*
The function f is continuously differentiable and $\alpha \leq n$. There exists $k \in \mathbb{N}_>$, $c_0, \ldots, c_k$ in $\mathbb{R}$ such that for all $\tilde{\mathbf{z}} \in \mathcal{L}_1$, $\mathbf{y} \in \mathbb{R}^n$, $c_{\tilde{\mathbf{z}}}, c_{\mathbf{y}} \in [0, 1]$

$$\|\nabla f(\tilde{\mathbf{z}} + c_{\tilde{\mathbf{z}}} c_{\mathbf{y}} \mathbf{y})\|^2 \leq c_0 + \sum_{i=1}^{k} c_i \|\mathbf{y}\|^i \quad . \tag{3.25}$$

In the next lemma, we verify that convex-quadratic functions satisfy the previous assumptions if $n \geq 2$.

LEMMA 3.10. *Let $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T H \mathbf{x}$ with H symmetric positive definite. It satisfies Assumptions 2 if $n \geq 2$.*

Proof. The function f is positively homogeneous with degree 2. Hence to satisfy the assumption $\alpha \leq n$ we need $n \geq 2$. Moreover, it is continuously differentiable and homogeneous with degree 2 and satisfies $\nabla f(\mathbf{x}) = H\mathbf{x}$. Hence $\|\nabla H(\tilde{\mathbf{z}} + \gamma_0 c_{\mathbf{y}} \mathbf{y})\|^2 \leq \|\!|H|\!\| \|\tilde{\mathbf{z}} + c_{\tilde{\mathbf{z}}} c_{\mathbf{y}} \mathbf{y}\|^2 \leq K(\|\tilde{\mathbf{z}}\| + \|\mathbf{y}\|)^2 \leq K(K_1 + \|\mathbf{y}\|)^2 = KK_1 + 2KK_1\|\mathbf{y}\| + KK_1^2\|\mathbf{y}\|^2$ where $\|\!|.|\!\|$ is the induced matrix norm associated to the euclidian norm $\|.\|$ and K is a bound for $\|\!|H|\!\|$ and K_1 a bound for the elements of $\bar{\mathcal{L}}_1$. Hence (3.25) is satisfied with $k = 2$. □

We are now ready to state the main result of this section.

THEOREM 3.11. *Consider $(\mathbf{X}_t, \sigma_t)_{t\in\mathbb{N}}$, a (1+1)-ES with generalized one-fifth success rule as defined in (1.1), (1.2) and (1.3) optimizing $h = g \circ f$ where $g \in \mathcal{M}$ and $f : \mathbb{R}^n \to [0, +\infty[$ satisfies Assumptions 2. Let $\mathbf{Z} = (\mathbf{Z}_t = \mathbf{X}_t/\sigma_t)_{t\in\mathbb{N}}$ be the Markov chain associated to the (1+1)-ES optimizing h defined in (2.19). Then the function*

$$V(\mathbf{z}) = f(\mathbf{z})1_{\{f(\mathbf{z})\geq 1\}} + \frac{1}{f(\mathbf{z})}1_{\{f(\mathbf{z})<1\}} \tag{3.26}$$

satisfies a drift condition for geometric ergodicity (in the sense of (3.10)) for the Markov chain $\mathbf{Z}$ if $\gamma > 1$ and

$$\frac{1}{2}\left(\frac{1}{\gamma^\alpha} + \gamma^{\alpha/q}\right) < 1 \ . \tag{3.27}$$

The theorem calls for a few remarks. The LHS of (3.27) corresponds to the expectation of the step-size change to the α on a linear function (i.e., $f(\mathbf{x}) = \mathbf{a}.\mathbf{x} + b$ for $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$). Using the notation $\eta^\star$ introduced in (2.20) for the step-size change, the condition (3.27) requires that on linear function

$$E[1/(\eta^\star)^\alpha_{\text{linear}}] < 1 \ , \tag{3.28}$$

that translates a step-size increase on linear function. This condition is similar to the one found to prove geometric ergodicity for the $(1, \lambda)$ with self-adaptation [1]. For an algorithm without elitist selection, the condition (3.28) is the only one formulated on the step-size change to guarantee geometric ergodicity. It ensures that the limit of PV/V is smaller 1 for $\mathbf{z}$ to infinity (see proof). For the $(1 + 1)$-ES another condition appears due to the limit of PV/V in zero (see the details in the proof) that is reflected in the fact that $\gamma^{-\alpha/4} < 1$, i.e., the step-size should decrease in case of failure. This translates for the one-fifth success rule into $\gamma > 1$. Note that we also need this condition $\gamma > 1$ for the irreducibility, the small sets and the aperiodicity.

Proof. Using the definition of V we can write $PV(\mathbf{z}) = E[V(\mathbf{Z}_{t+1})|\mathbf{Z}_t = \mathbf{z}]$ as

$$PV(\mathbf{z}) = E\left[f(\mathbf{Z}_{t+1})1_{\{f(\mathbf{Z}_{t+1})\geq 1\}} + \frac{1}{f(\mathbf{Z}_{t+1})}1_{\{f(\mathbf{Z}_{t+1})<1\}}|\mathbf{Z}_t = \mathbf{z}\right] \ .$$

According to Lemma 3.6, we need to study the limits of $PV(\mathbf{z})/V(\mathbf{z})$ for $\mathbf{z}$ to 0 and to ∞.

Investigating the limit of PV/V for $\mathbf{z}$ to infinity. We first investigate the limit for $\|\mathbf{z}\|$ to infinity and consider $\mathbf{z}$ large enough, in particular we can assume that

$$f(\mathbf{z}) \geq \max\{1, \gamma^{\alpha/q}\} \tag{3.29}$$

and hence $V(\mathbf{z}) = f(\mathbf{z})$. Then

$$\frac{PV(\mathbf{z})}{V(\mathbf{z})} = \underbrace{E\left[\frac{f(\mathbf{Z}_{t+1})1_{\{f(\mathbf{Z}_{t+1})\geq 1\}}}{f(\mathbf{z})}|\mathbf{Z}_t = \mathbf{z}\right]}_{A(\mathbf{z})} + \underbrace{\frac{E\left[\frac{1}{f(\mathbf{Z}_{t+1})}1_{\{f(\mathbf{Z}_{t+1})\leq 1\}}|\mathbf{Z}_t = \mathbf{z}\right]}{f(\mathbf{z})}}_{B(\mathbf{z})} \tag{3.30}$$

Throughout this proof we will denote $\mathcal{N}$ the multivariate normal distribution used at iteration t to sample a new candidate solution. Namely the update for $\mathbf{Z}_t$ reads:

$$\boxed{\mathbf{Z}_{t+1} = \frac{\mathbf{Z}_t + \mathcal{N}}{\gamma}1_{\{f(\mathbf{Z}_t+\mathcal{N})\leq f(\mathbf{Z}_t)\}} + \frac{\mathbf{Z}_t}{\gamma^{-1/q}}1_{\{f(\mathbf{Z}_t+\mathcal{N})>f(\mathbf{Z}_t)\}}} \ .$$

Let us first investigate the term $A(\mathbf{z})$ introduced in (3.30). It is equal to

$$E\left[\frac{f(\frac{\mathbf{z}+\mathcal{N}}{\gamma})}{f(\mathbf{z})}1_{\{f(\mathbf{z}+\mathcal{N})\leq f(\mathbf{z})\}}1_{\{f(\frac{\mathbf{z}+\mathcal{N}}{\gamma})\geq 1\}}\right]+E\left[\frac{f(\frac{\mathbf{z}}{\gamma^{-\frac{1}{q}}})}{f(\mathbf{z})}1_{\{f(\mathbf{z}+\mathcal{N})>f(\mathbf{z})\}}\underbrace{1_{\{f(\frac{\mathbf{z}}{\gamma^{-1/q}})\geq 1\}}}_{=1\ (\text{see}(3.29))}\right]$$

$$=\frac{1}{\gamma^{\alpha}}\underbrace{E\left[f\left(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}}+\frac{\mathcal{N}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\right)1_{\{f(\mathbf{z}+\mathcal{N})\leq f(\mathbf{z})\}}1_{\{f(\frac{\mathbf{z}+\mathcal{N}}{\gamma})\geq 1\}}\right]}_{A_1}+\gamma^{\alpha/q}\underbrace{E\left[1_{\{f(\mathbf{z}+\mathcal{N})>f(\mathbf{z})\}}\right]}_{A_2}\ .$$

Using Lemma 3.8 we obtain that A_2 converges to $1/2$ when $\mathbf{z}$ goes to ∞. Let us now handle the term A_1. Using the mean value theorem we have the existence of $c_{\mathcal{N}}\in[0,1]$ such that

$$f\left(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}}+\frac{\mathcal{N}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\right)=\underbrace{f\left(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\right)}_{=1}+\nabla f\left(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}}+c_{\mathcal{N}}\frac{\mathcal{N}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\right).\frac{\mathcal{N}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\ . \tag{3.31}$$

Hence the term A_1 can be decomposed in two terms:

$$A_1=\underbrace{E\left[1_{\{f(\mathbf{z}+\mathcal{N})\leq f(\mathbf{z})\}}1_{\{f(\frac{\mathbf{z}+\mathcal{N}}{\gamma})\geq 1\}}\right]}_{A_{11}}+ \tag{3.32}$$

$$\frac{1}{f(\mathbf{z})^{\frac{1}{\alpha}}}\underbrace{E\left[\nabla f\left(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}}+c_{\mathcal{N}}\frac{\mathcal{N}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\right).\mathcal{N}1_{\{f(\mathbf{z}+\mathcal{N})\leq f(\mathbf{z})\}}1_{\{f(\frac{\mathbf{z}+\mathcal{N}}{\gamma})\geq 1\}}\right]}_{A_{12}}\ .$$

The term A_{11} equals

$$A_{11}=\int 1_{\{f(\mathbf{z}+\mathbf{n})\leq f(\mathbf{z})\}}(\mathbf{n})1_{\{f(\frac{\mathbf{z}+\mathbf{n}}{\gamma})\geq 1\}}(\mathbf{n})p_{\mathcal{N}}(\mathbf{n})d\mathbf{n}\ .$$

According to Lemma 3.8, the term A_{11} converges to $1/2$ when $\mathbf{z}$ goes to ∞. Note that for all $\mathbf{n}$, the indicator $1_{\{f(\frac{\mathbf{z}+\mathbf{n}}{\gamma})\geq 1\}}(\mathbf{n})$ converges to 1 for $\mathbf{z}$ to ∞. We now take care of the term A_{12} and prove that $|A_{12}|$ is bounded which will imply that $A_{12}\frac{1}{f(\mathbf{z})^{1/\alpha}}$ converges to zero when $\mathbf{z}$ goes to ∞.

$$|A_{12}|\leq E\left[|\nabla f\left(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}}+c_{\mathcal{N}}\frac{\mathcal{N}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\right).\mathcal{N}|\right]\leq E\left[\|\nabla f\left(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}}+c_{\mathcal{N}}\frac{\mathcal{N}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\right)\|\|\mathcal{N}\|\right]$$

$$\leq E\left[\|\nabla f\left(\frac{\mathbf{z}}{f(\mathbf{z})^{\frac{1}{\alpha}}}+c_{\mathcal{N}}\frac{\mathcal{N}}{f(\mathbf{z})^{\frac{1}{\alpha}}}\right)\|^2\right]^{\frac{1}{2}}\underbrace{E\left[\|\mathcal{N}\|^2\right]^{\frac{1}{2}}}_{=\sqrt{n}}\ .$$

For the last two inequalities we have applied Cauchy-Schwarz inequalities. We denote $\tilde{\mathbf{z}}=\frac{\mathbf{z}}{f(\mathbf{z})^{1/\alpha}}$ and $c_{\tilde{\mathbf{z}}}=\frac{1}{f(\mathbf{z})^{1/\alpha}}$ and apply (3.25). Hence we find

$$E\left[\|\nabla f\left(\tilde{\mathbf{z}}+c_{\mathcal{N}}c_{\tilde{\mathbf{z}}}\mathcal{N}\right)\|^2\right]\leq c_0+\sum_{i=1}^{k}c_iE[\|\mathcal{N}\|^i]=:M_{3.11}$$

and it follows that $|A_{12}| \leq \sqrt{M_{3.11} n}$. Hence

$$\lim_{\|\mathbf{z}\| \to \infty} A(\mathbf{z}) = \frac{1}{2}\left(\frac{1}{\gamma^\alpha} + \gamma^{\alpha/q}\right) \quad .$$

We investigate now the term $B(\mathbf{z})$ defined in (3.30).

$$B(\mathbf{z}) = \frac{1}{f(\mathbf{z})} \underbrace{E\left[\frac{1_{\{f(\mathbf{z}+\mathcal{N}) \leq f(\mathbf{z})\}} 1_{\{f(\frac{\mathbf{z}+\mathcal{N}}{\gamma}) \leq 1\}}}{f(\frac{\mathbf{z}+\mathcal{N}}{\gamma})}\right]}_{B_{11}} + \frac{1}{f(\mathbf{z})} \underbrace{E\left[\frac{1_{\{f(\mathbf{z}+\mathcal{N}) > f(\mathbf{z})\}} 1_{\{f(\mathbf{z}/\gamma^{-\frac{1}{q}}) \leq 1\}}}{f(\mathbf{z}/\gamma^{-\frac{1}{q}})}\right]}_{=0 \text{ as } f(\mathbf{z}/\gamma^{-\frac{1}{q}}) > 1}$$

Let us now take care of the term B_{11} which is upper bounded by

$$B_{11} \leq E\left[\frac{1}{f(\frac{\mathbf{z}+\mathcal{N}}{\gamma})}\right] = \gamma^\alpha E\left[\frac{1}{f(\mathbf{z}+\mathcal{N})}\right] < \gamma^\alpha c_{3.9} \quad ,$$

where for the latter term we have used Lemma 3.9. Overall we find that

$$0 \leq B(\mathbf{z}) < \frac{1}{f(\mathbf{z})} \gamma^\alpha c_{3.9} \xrightarrow[\|\mathbf{z}\| \to \infty]{} 0$$

where the latter limit comes from the fact that $\frac{1}{f(\mathbf{z})}$ converges to zero when $\mathbf{z}$ goes to infinity. Overall we have proven that

$$\lim_{\|\mathbf{z}\| \to \infty} \frac{PV(\mathbf{z})}{V(\mathbf{z})} = \frac{1}{2}\left(\frac{1}{\gamma^\alpha} + \gamma^{\alpha/q}\right) \quad . \tag{3.33}$$

Investigating the limit of PV/V for $\mathbf{z}$ to zero. We now investigate the limit for $\|\mathbf{z}\|$ to zero and consider thus $\mathbf{z}$ small enough, in particular we can assume

$$f(\mathbf{z}) < \min\{1, \gamma^{\alpha/q}\}$$

and hence $1/V(\mathbf{z}) = f(\mathbf{z})$. The quantity $PV(\mathbf{z})/V(\mathbf{z})$ writes

$$\frac{PV(\mathbf{z})}{V(\mathbf{z})} = \underbrace{f(\mathbf{z}) E\left[f(\mathbf{Z}_{t+1}) 1_{\{f(\mathbf{Z}_{t+1}) \geq 1\}} | \mathbf{Z}_t = \mathbf{z}\right]}_{C(\mathbf{z})} + \underbrace{E\left[\frac{f(\mathbf{z})}{f(\mathbf{Z}_{t+1})} 1_{\{f(\mathbf{Z}_{t+1}) < 1\}} | \mathbf{Z}_t = \mathbf{z}\right]}_{D(\mathbf{z})} \quad .$$

Let us investigate the term $C(\mathbf{z})$:

$$\begin{aligned} C(\mathbf{z}) = f(\mathbf{z}) E\left[f\left(\frac{\mathbf{z}+\mathcal{N}}{\gamma}\right) 1_{\{f(\mathbf{z}+\mathcal{N}) \leq f(\mathbf{z})\}} 1_{\{f((\mathbf{z}+\mathcal{N})/\gamma) \geq 1\}}\right] \\ + f(\mathbf{z}) E\left[f(\gamma^{\frac{1}{q}}\mathbf{z}) 1_{\{f(\mathbf{z}+\mathcal{N}) > f(\mathbf{z})\}} 1_{\{f(\gamma^{\frac{1}{q}}\mathbf{z}) \geq 1\}}\right] \end{aligned} \tag{3.34}$$

and hence

$$0 \leq C(\mathbf{z}) \leq f(\mathbf{z}) E\left[f\left(\frac{\mathbf{z}+\mathcal{N}}{\gamma}\right)\right] + f(\mathbf{z})^2 \gamma^{\alpha/q} \quad .$$

According to (3.18) and (3.19), for $\|\mathbf{z}\|$ small enough (hence staying in a bounded region), $E\left[f\left(\frac{\mathbf{z}+\mathcal{N}}{\gamma}\right)\right]$ is a bounded function of $\mathbf{z}$ and thus $C(\mathbf{z})$ converges to zero when $\mathbf{z}$ goes to 0:

$$\lim_{\mathbf{z}\to 0} C(\mathbf{z}) = 0 \ . \tag{3.35}$$

Let us investigate the term $D(\mathbf{z})$:

$$\begin{aligned} D(\mathbf{z}) &= f(\mathbf{z})E\left[\frac{1_{\{f(\mathbf{z}+\mathcal{N})\leq f(\mathbf{z})\}}1_{\{f((\mathbf{z}+\mathcal{N})/\gamma)<1\}}}{f((\mathbf{z}+\mathcal{N})/\gamma)}\right] + f(\mathbf{z})E\left[\frac{1_{\{f(\mathbf{z}+\mathcal{N})>f(\mathbf{z})\}}}{f(\mathbf{z}/\gamma^{-1/q})}\underbrace{1_{\{f(\frac{\mathbf{z}}{\gamma^{-\frac{1}{q}}})<1\}}}_{=1}\right] \\ &= f(\mathbf{z})E\left[\frac{1_{\{f(\mathbf{z}+\mathcal{N})\leq f(\mathbf{z})\}}1_{\{f((\mathbf{z}+\mathcal{N})/\gamma)<1\}}}{f((\mathbf{z}+\mathcal{N})/\gamma)}\right] + \gamma^{-\alpha/q}E\left[1_{\{f(\mathbf{z}+\mathcal{N})>f(\mathbf{z})\}}\right] \\ &= \underbrace{f(\mathbf{z})\gamma^{\alpha}E\left[\frac{1_{\{f(\mathbf{z}+\mathcal{N})\leq f(\mathbf{z})\}}1_{\{f((\mathbf{z}+\mathcal{N})/\gamma)<1\}}}{f(\mathbf{z}+\mathcal{N})}\right]}_{D_1(\mathbf{z})} + \gamma^{-\alpha/q}\Pr(f(\mathbf{z}+\mathcal{N}) > f(\mathbf{z})) \end{aligned}$$

The term $D_1(\mathbf{z})$ is upper bounded by $f(\mathbf{z})\gamma^{\alpha}E\left[\frac{1}{f(\mathbf{z}+\mathcal{N})}\right]$ and using Lemma 3.9 we find that for all $\mathbf{z}$, $E\left[\frac{1}{f(\mathbf{z}+\mathcal{N})}\right]$ is bounded by a constant. Hence since $f(\mathbf{z})$ converges to 0 when $\mathbf{z}$ goes to 0 (Lemma 2.8), so does $D_1(\mathbf{z})$. Overall since according to Lemma 3.7, $\Pr(f(\mathbf{z}+\mathcal{N}) > f(\mathbf{z}))$ converges to 1 when $\mathbf{z}$ goes to 0, we find that

$$\lim_{\mathbf{z}\to 0} D(\mathbf{z}) = \gamma^{-\alpha/q} \ . \tag{3.36}$$

Overall, we have proven that

$$\lim_{\mathbf{z}\to 0} PV(\mathbf{z})/V(\mathbf{z}) = \gamma^{-\alpha/q} \ . \tag{3.37}$$

According to Lemma 3.6, we obtain a drift condition for geometric ergodicity if the limits in (3.33) and (3.37) are strictly smaller 1, i.e., if

$$\frac{1}{2}\left(\frac{1}{\gamma^{\alpha}} + \gamma^{\alpha/q}\right) < 1$$

and $\gamma^{-\alpha/q} < 1$. This latter condition being equivalent to $\gamma > 1$. □

3.4. Harris Recurrence and Positivity. Harris recurrence and positivity of the chain $\mathbf{Z}$ follow from the geometric drift proven in Theorem 3.11. Indeed, remind the following drift result for Harris recurrence:

THEOREM 3.12 (Theorem 9.1.8. in [12]). *(Drift condition for Harris recurrence) Suppose $\mathbf{Z}$ is a ψ-irreducible chain. If there exists a petite set C and a function V which is unbounded off petite sets such that*

$$\Delta V(\mathbf{z}) \leq 0, \mathbf{z} \in C^c \tag{3.38}$$

holds, then $\mathbf{Z}$ is Harris recurrent.

In the previous theorem, a function $V : \mathcal{Z} \mapsto \mathbb{R}_+$ is unbounded off petite sets for $\mathbf{Z}$ if for any $c < \infty$, the sublevel sets $\mathcal{L}_c = \{\mathbf{y} : V(\mathbf{y}) \leq c\}$ is petite (see [12, Section

8.4.2]). From [12, Theorem 10.4.4] a recurrent chain admits an unique (up to constant multiples) invariant measure. The positivity is deduced from another drift condition as expressed in the following theorem.

THEOREM 3.13 (From Theorem 13.0.1 in [12]). *Suppose that* $\mathbf{Z}$ *is an aperiodic Harris recurrent chain with invariant measure* π*. The following are equivalent:*
The chain is positive Harris: that is, the unique invariant measure π *is finite.*
There exists some petite set C*, some* $b < \infty$ *and a non-negative function* V *finite at some* $\mathbf{z}_0 \in \mathcal{Z}$*, satisfying*

$$\Delta V(\mathbf{z}) \leq -1 + b1_C(\mathbf{z}), \mathbf{z} \in \mathcal{Z} \ . \tag{3.39}$$

Using those two theorems we deduce the corollary that under the conditions of Theorem 3.11 the chain $\mathbf{Z}$ is positive Harris recurrent:

COROLLARY 3.14. *Consider a (1+1)-ES with generalized one-fifth success rule as defined in* (1.1)*,* (1.2) *and* (1.3) *optimizing* $h = g \circ f$ *where* $g \in \mathcal{M}$*,* $f : \mathbb{R}^n \to [0, +\infty[$ *satisfies Assumptions 2. Let* $\mathbf{Z} = (\mathbf{Z}_t = \mathbf{X}_t/\sigma_t)_{t\in\mathbb{N}}$ *be the Markov chain associated to the (1+1)-ES optimizing* h *defined in* (2.19)*. If* $\gamma > 1$ *and* $\frac{1}{2}\left(\frac{1}{\gamma^\alpha} + \gamma^{\alpha/q}\right) < 1$*, then* $\mathbf{Z}$ *is positive Harris recurrent.*

Proof. The assumptions on f ensure that the chain is φ-irreducible and aperiodic (see Section 3.2) and the geometric drift function exhibited in Theorem 3.11 satisfies obviously (3.38) and (3.39). It is unbounded off petite sets as the sublevel sets are small sets for $\mathbf{Z}$. □

4. Linear Convergence of the $(1+1)$-ES with the Generalized One-fifth Success Rule. Using the properties derived on the normalized chain $(\mathbf{Z}_t)_{t\in\mathbb{N}}$ we can now prove the global linear convergence of the $(1+1)$-ES with generalized one-fifth success rule. Linear convergence is formulated almost surely and in expectation. In a last part we characterize how fast the stationary regime where linear convergence takes place is reached. Common to the linear convergence results is the integrability of $\mathbf{z} \mapsto \ln \|\mathbf{z}\|$ with respect to the invariant probability measure π of the chain $\mathbf{Z}$ that we investigate in the next section.

4.1. Integrability w.r.t. the Stationary Measure. To verify the integrability of $\mathbf{z} \mapsto \ln \|\mathbf{z}\|$ with respect to the invariant probability measure π, we use (3.11) which is a consequence of the existence of a geometric drift. More formally we derive the following general technical lemma.

LEMMA 4.1. *Let* V *be a geometric drift function for* $\mathbf{Z}$ *in the sense of* (3.10) *and* π *its invariant probability measure. Assume that there exists* $\mathbf{z} \in S_V$ *such that* $\int V(\mathbf{y})P(\mathbf{z}, d\mathbf{y}) < \infty$*, then* V *is integrable against* π*:* $\|\pi\|_V = \int V(\mathbf{z})\pi(d\mathbf{z}) < +\infty$.

Proof. From the inequality (3.11), there exists $R < \infty$ and $\rho < 1$ such that for any $\mathbf{z} \in S_V$

$$\|P(\mathbf{z}, .) - \pi\|_V \leq \rho R V(\mathbf{z}) \ . \tag{4.1}$$

Consider a sequence of simple positive functions V_k such that for each $\mathbf{z}$, $V_k(\mathbf{z})$ converges to $V(\mathbf{z})$ and $V_k(\mathbf{z})$ is increasing. Then we know that $\int V(\mathbf{z})\pi(d\mathbf{z}) = \lim_k \int V_k(\mathbf{z})\pi(d\mathbf{z})$ where the latter limit always exist but may be infinite. From the

triangular inequality we deduce that for all k

$$\int V_k(\mathbf{y})\pi(d\mathbf{y}) = \int V_k(\mathbf{y})\pi(d\mathbf{y}) - \int V_k(\mathbf{y})P(\mathbf{z},d\mathbf{y}) + \int V_k(\mathbf{y})P(\mathbf{z},d\mathbf{y}) \tag{4.2}$$

$$\leq |\int V_k(\mathbf{y})\pi(d\mathbf{y}) - \int V_k(\mathbf{y})P(\mathbf{z},d\mathbf{y})| + \int V_k(\mathbf{y})P(\mathbf{z},d\mathbf{y}) \tag{4.3}$$

$$\leq \|P(\mathbf{z},.) - \pi\|_V + \int V(\mathbf{y})P(\mathbf{z},d\mathbf{y}) \tag{4.4}$$

where for the last inequality we have used the fact that $0 \leq V_k \leq V$ and the definition of $\|P(\mathbf{z},.) - \pi\|_V$ namely $\|P(\mathbf{z},.) - \pi\|_V = \sup_{k,|k|\leq V} |\int k(\mathbf{y})(P(\mathbf{z},d\mathbf{y}) - \pi(d\mathbf{y}))|$. The fact that $\int V_k(\mathbf{y})P(\mathbf{z},d\mathbf{y}) \leq \int V(\mathbf{y})P(\mathbf{z},d\mathbf{y})$ is a consequence of $V_k \leq V$. In addition, $\int V(\mathbf{y})P(\mathbf{z},d\mathbf{y}) < \infty$ according to the assumptions. Using (4.1) we find that for all k

$$\int V_k(\mathbf{y})\pi(d\mathbf{y}) \leq RV(\mathbf{z}) + \int V(\mathbf{y})P(\mathbf{z},d\mathbf{y}) < \infty$$

And hence $\int V\pi = \lim_k \int V_k(\mathbf{y})\pi(d\mathbf{y}) < \infty$. □

We can now apply the previous lemma to the drift function $V(\mathbf{z}) = f(\mathbf{z})1_{\{f(\mathbf{z})\geq 1\}} + \frac{1}{f(\mathbf{z})}1_{\{f(\mathbf{z})<1\}}$ assuming that f satisfies Assumptions 2 and that the sufficient conditions for a geometric drift of Theorem 3.11 are satisfied (the fact that $PV(\mathbf{z})$ is finite for one $\mathbf{z}$ in S_V comes from Lemma 3.9). We now prove that $|\ln\|\mathbf{z}\||$ is upper bounded by a constant times $V(\mathbf{z})$ that implies together with the previous lemma the integrability of $\ln\|\mathbf{z}\|$ w.r.t. stationary measure π.

LEMMA 4.2. *Assume that f satisfies Assumptions 1 and is continuous on $\mathbb{R}^n_{\neq}$. Then there exists a constant $K_{4.2}$ such that*

$$|\ln\|\mathbf{z}\|| \leq K_{4.2}V(\mathbf{z}) \ . \tag{4.5}$$

Assume that f satisfies Assumptions 2 and that that $\gamma > 1$ and $\frac{1}{2}(1/\gamma^\alpha + \gamma^{\alpha/q}) < 1$. Then $\ln\|\mathbf{z}\|$ is integrable w.r.t. the stationary measure π of $\mathbf{Z}$.

Proof. For $\|\mathbf{z}\|$ close to 0, $f(\mathbf{z})$ is close to 0 and thus $V(\mathbf{z}) = \frac{1}{f(\mathbf{z})}$. Hence there exists c such that

$$\frac{|\ln\|\mathbf{z}\||}{V(\mathbf{z})} = \frac{|\ln\|\mathbf{z}\||}{f(\mathbf{z})} \leq c\frac{|\ln\|\mathbf{z}\||}{\|\mathbf{z}\|^\alpha}$$

where the latter term is bounded since $\|\mathbf{z}\|^\alpha|\ln\|\mathbf{z}\||$ goes to 0 when $\mathbf{z}$ goes to 0. Note that for the middle inequality we have used (2.25). For $\mathbf{z}$ large, $V(\mathbf{z}) = f(\mathbf{z})$ and using $|\ln\|\mathbf{z}\|| \leq \|\mathbf{z}\|^\alpha \leq f(\mathbf{z})$ (again we use (2.25)) we find that for $\mathbf{z}$ large $|\ln\|\mathbf{z}\|| \leq V(\mathbf{z})$. Since $|\ln\|\mathbf{z}\||$ is continuous and hence bounded on all $\|\mathbf{z}\|$ in an interval $[a,b]$ with $0 < a < b < +\infty$ then there exists $\tilde{c}$ such that $|\ln\|\mathbf{z}\||1_{\{a\leq\|\mathbf{z}\|\leq b\}} \leq \tilde{c} \leq \tilde{c}V(\mathbf{z})1_{\{a\leq\|\mathbf{z}\|\leq b\}}$. Overall we have proven that (4.5) holds.

Since according to Lemma 3.9 $\int V(\mathbf{y})P(\mathbf{z},d\mathbf{y}) < +\infty$ for $\mathbf{z} \in \mathbb{R}^n_{\neq}$, when the conditions of Theorem 3.11 are satisfied, we deduce from the previous lemma that V is integrable w.r.t. π and hence $|\ln\|\mathbf{z}\||$ is integrable with respect to π. □

4.2. Asymptotic Probability of Success. We investigate now the asymptotic probability of success that comes into play in the convergence rate of the algorithm. Success is defined as whether a candidate solution is better than the current solution $\mathbf{X}_t$, i.e., as $P_{\frac{\mathbf{x}}{\sigma}}\left(f(\mathbf{X}_t + \sigma_t\mathbf{U}_t^1) \leq f(\mathbf{X}_t)\right)$ and due to the scale-invariant property of f, this latter quantity can be expressed with the Markov chain $\mathbf{Z}$ as

$$P_{\mathbf{z}}\left(f(\mathbf{Z}_t + \mathbf{U}_t^1) \leq f(\mathbf{Z}_t)\right) = E_{\mathbf{z}}\left[1_{\{f(\mathbf{Z}_t+\mathbf{U}_t^1)\leq f(\mathbf{Z}_t)\}}\right] \ .$$

The convergence of the probability of success is a consequence of the positivity and aperiodicity and can be deduced from [12, Theorem 14.0.1].

PROPOSITION 4.3 (Asymptotic probability of success). *Let the $(1+1)$-ES with generalized one-fifth success rule optimize $h = g \circ f$ where $g \in \mathcal{M}$ and f satisfies Assumptions 2. Assume that $\gamma > 1$ and $\frac{1}{2}(1/\gamma^{\alpha} + \gamma^{\alpha/q}) < 1$. Let π be the invariant probability measure of the normalized Markov chain $\mathbf{Z}$. Then for any initial condition $(\mathbf{x}, \sigma) \in \mathbb{R}^n_{\neq} \times \mathbb{R}^+_{>}$, the following holds*

(4.6)
$$\mathrm{PS} := \lim_{t\to\infty} P_{\frac{\mathbf{x}}{\sigma}}\left(f(\mathbf{X}_t + \sigma_t \mathbf{U}_t^1) \leq f(\mathbf{X}_t)\right) = \int 1_{\{f(\mathbf{y}+\mathbf{n})\leq f(\mathbf{y})\}}(\mathbf{y},\mathbf{n})\pi(d\mathbf{y})p_{\mathcal{N}}(\mathbf{n})d\mathbf{n} \ .$$

Proof. From [12, Theorem 14.0.1], given a ψ-irreducible and aperiodic chain, given a function $k \geq 1$, if the chain is positive recurrent with invariant probability measure π and $\pi(k) = \int \pi(d\mathbf{z})k(\mathbf{z}) < \infty$, then for any $\mathbf{z} \in S_{\tilde{V}} = \{\mathbf{z} : \tilde{V}(\mathbf{z}) < \infty\}$ where the function $\tilde{V}$ is an extended-valued function satisfying

$$\Delta\tilde{V}(\mathbf{z}) \leq -k(\mathbf{z}) + b1_C(\mathbf{z}) \tag{4.7}$$

for some petite set C and $b \in \mathbb{R}$, $\|P^t(\mathbf{z},.) - \pi\|_k \to 0$ holds. We take here $k(\mathbf{z}) = 1$ and hence the geometric drift proven in Theorem 3.11 implies also (4.7). From Corollary 3.14, the chain is positive and hence the function $k(\mathbf{y}) = 1$ is integrable w.r.t. π. Remark that

$$P_{\mathbf{z}}\left(f(\mathbf{Z}_t + \mathbf{U}_t^1) \leq f(\mathbf{Z}_t)\right) = \int w(\mathbf{y})P^t(\mathbf{z},\mathbf{y})d\mathbf{y}$$

where $w(\mathbf{y}) = \int 1_{\{f(\mathbf{y}+\mathbf{n})\leq f(\mathbf{y})\}}(\mathbf{n})p_{\mathcal{N}}(\mathbf{n})d\mathbf{n}$, then $w(\mathbf{y}) \leq 1$ and hence from [12, Theorem 14.0.1] we deduce that ($\mathbf{z} = \mathbf{x}/\sigma$)

$$|P_{\mathbf{z}}\left(f(\mathbf{Z}_t + \mathbf{U}_t^1) \leq f(\mathbf{Z}_t)\right) - \int w(\mathbf{y})\pi(d\mathbf{y})| \leq \|P^t(\mathbf{z},.) - \pi\|_{\mathbf{y}\mapsto 1} \xrightarrow[t\to\infty]{} 0 \ .$$

□ We also derive a Law of Large Numbers for the asymptotic probability of success.

PROPOSITION 4.4. *Let the $(1+1)$-ES with generalized one-fifth success rule optimize $h = g \circ f$ where $g \in \mathcal{M}$ and f satisfies Assumptions 2. Assume that $\gamma > 1$ and $\frac{1}{2}(1/\gamma^{\alpha} + \gamma^{\alpha/q}) < 1$. Let $(\mathbf{Z}_t = \mathbf{X}_t/\sigma_t)_{t\in\mathbb{N}}$ be the normalized Markov chain associated to $(\mathbf{X}_t, \sigma_t)_{t\in\mathbb{N}}$. Then for all initial condition $(\mathbf{X}_0, \sigma_0)$*

$$\frac{1}{t}\sum_{k=0}^{t-1} 1_{\{f(\mathbf{X}_k+\sigma_k\mathbf{U}_k^1)\leq f(\mathbf{X}_k)\}} = \frac{1}{t}\sum_{k=0}^{t-1} 1_{\{f(\mathbf{Z}_k+\mathbf{U}_k^1)\leq f(\mathbf{Z}_k)\}} \xrightarrow[t\to\infty]{} \mathrm{PS} \tag{4.8}$$

where the asymptotic probability of success PS *is defined in* (4.6).

Proof. From Corollary 3.14, the chain is positive and Harris recurrent. The function $\mathbf{y} \mapsto w(\mathbf{y}) = \int 1_{\{f(\mathbf{y}+\mathbf{n})\leq f(\mathbf{y})\}}(\mathbf{n})p_{\mathcal{N}}(\mathbf{n})d\mathbf{n}$ being integrable w.r.t. the stationary measure π, we can thus apply the Law of Large Numbers ([12, Theorem 17.0.1]) that gives us (4.8). □

4.3. Almost Sure Linear Convergence. Almost sure linear convergence derives from the application of a Law of Large Number (LLN) (see Theorem 5.2 in [3]). Some assumptions to be able to apply a LLN to $(\mathbf{Z}_t)_{t\in\mathbb{N}}$ are positivity, Harris-recurrence and integrability of $\ln\|\mathbf{z}\|$. We are then now ready to prove the almost sure linear convergence of the $(1+1)$-ES with generalized one-fifth success rule.

THEOREM 4.5. *Let the $(1+1)$-ES with generalized one-fifth success rule optimize $h = g \circ f$ where $g \in \mathcal{M}$ and f satisfies Assumptions 2. Assume that $\gamma > 1$*

and $\frac{1}{2}(1/\gamma^{\alpha} + \gamma^{\alpha/q}) < 1$. *Then for all initial condition* $(\mathbf{X}_0, \sigma_0)$ *almost sure linear convergence for the mean vector and for the step-size holds, i.e.,*

$$\frac{1}{t} \ln \frac{\|\mathbf{X}_t\|}{\|\mathbf{X}_0\|} \xrightarrow[t\to\infty]{} \ln \gamma \left(\frac{q+1}{q} \mathrm{PS} - \frac{1}{q} \right) \tag{4.9}$$

$$\frac{1}{t} \ln \frac{\sigma_t}{\sigma_0} \xrightarrow[t\to\infty]{} \ln \gamma \left(\frac{q+1}{q} \mathrm{PS} - \frac{1}{q} \right) \quad . \tag{4.10}$$

where PS *is the asymptotic probability of success defined in* (4.6).

Proof. The proof follows the same line as the proof of Theorem 5.2 in [3]. We start by re-writing the log-progress:

$$\frac{1}{t} \ln \frac{\|\mathbf{X}_t\|}{\|\mathbf{X}_0\|} = \frac{1}{t} \sum_{k=0}^{t-1} \ln \frac{\|\mathbf{X}_{k+1}\|}{\|\mathbf{X}_k\|} = \frac{1}{t} \sum_{k=0}^{t-1} \ln \frac{\sigma_{k+1} \|\mathbf{Z}_{k+1}\|}{\sigma_k \|\mathbf{Z}_k\|} = \frac{1}{t} \sum_{k=0}^{t-1} \ln \frac{\mathcal{G}_2(1, \mathbf{Y}_k) \|\mathbf{Z}_{k+1}\|}{\|\mathbf{Z}_k\|}$$

$$= \underbrace{\frac{1}{t} \sum_{k=0}^{t-1} \ln \|\mathbf{Z}_{k+1}\|}_{A} - \underbrace{\frac{1}{t} \sum_{k=0}^{t-1} \ln \|\mathbf{Z}_k\|}_{B} + \underbrace{\frac{1}{t} \sum_{k=0}^{t-1} \ln \mathcal{G}_2(1, \mathbf{Y}_k)}_{C}$$

where we have used the scale-invariant property (2.11) and the fact that $\mathbf{Y}_k = \mathcal{S}_{(\mathbf{X}_t, \sigma_t)}(\mathbf{U}_k) * \mathbf{U}_t = \mathcal{S}_{(\mathbf{Z}_k, 1)}(\mathbf{U}_k)$. Since $\ln \|\mathbf{z}\|$ is integrable w.r.t. the stationary measure π we can apply the LLN to the terms A and B and we find that they both converge towards $\int \ln \|\mathbf{z}\| \pi(d\mathbf{z})$ such that A minus B converges to zero. Let us investigate the term C, since $\mathcal{G}_2(1, \mathbf{Y}_t) = (\gamma - \gamma^{-1/q}) 1_{\{f(\mathbf{Z}_t + \mathbf{U}_t^1) \leq f(\mathbf{Z}_t)\}} + \gamma^{-1/q}$ we find that $\ln \mathcal{G}_2(1, \mathbf{Y}_t) = \ln \gamma ((1 + 1/q) 1_{\{f(\mathbf{Z}_t + \mathbf{U}_t^1) \leq f(\mathbf{Z}_t)\}} - \frac{1}{q})$. Therefore

$$C = \frac{1}{t} \sum_{k=0}^{t-1} \ln \mathcal{G}_2(1, \mathbf{Y}_k) = \ln \gamma \left(1 + \frac{1}{q}\right) \frac{1}{t} \sum_{k=0}^{t-1} 1_{\{f(\mathbf{Z}_k + \mathbf{U}_k^1) \leq f(\mathbf{Z}_k)\}} - \frac{1}{q} \ln \gamma \to \tag{4.11}$$

$$\ln \gamma \left(1 + \frac{1}{q}\right) \int 1_{\{f(\mathbf{z}+\mathbf{u}) \leq f(\mathbf{z})\}}(\mathbf{z}, \mathbf{u}) \pi(d\mathbf{z}) p_{\mathcal{N}}(\mathbf{u}) d\mathbf{u} - \frac{1}{q} \ln \gamma = \ln \gamma \left[\left(1 + \frac{1}{q}\right) \mathrm{PS} - \frac{1}{q} \right]$$

where we have used Proposition 4.4 for the latter limit. The limit (4.10) follows from the fact that

$$\frac{1}{t} \ln \frac{\sigma_t}{\sigma_0} = \frac{1}{t} \sum_{k=0}^{t-1} \ln \frac{\sigma_{k+1}}{\sigma_k} = \frac{1}{t} \sum_{k=0}^{t-1} \ln \mathcal{G}_2(1, \mathbf{Y}_k)$$

with as above $\mathbf{Y}_k = \mathcal{S}_{(\mathbf{X}_t, \sigma_t)}(\mathbf{U}_k) * \mathbf{U}_t = \mathcal{S}_{(\mathbf{Z}_k, 1)}(\mathbf{U}_k)$. Using (4.11) we obtain (4.10). ∎

We define the convergence rate as minus the almost sure limit of the logarithm of $\|\mathbf{X}_t\|$ or of σ_t that corresponds to minus expectation of the logarithm of the step-size change w.r.t. the stationary distribution, i.e.,

$$\boxed{\mathrm{CR} := -E_{\pi} \left[\ln \eta^{\star} \right] = -\ln \gamma \left(\frac{q+1}{q} \mathrm{PS} - \frac{1}{q} \right)} \quad . \tag{4.12}$$

Figure 5.1 presents some convergence graphs of the $(1+1)$-ES with generalized 1/5 success rule. The slope of the linear decrease observed in log scale (after a small adaptation period on the left graph) corresponds to $-\mathrm{CR}$.

Sign of the convergence rate. Convergence will take place if CR > 0. We prove in the next proposition an alternative expression for the convergence rate that allows us to conclude that CR > 0.

PROPOSITION 4.6. *Let the* $(1+1)$*-ES with generalized one-fifth success rule optimize* $h = g \circ f$ *where* $g \in \mathcal{M}$ *and* f *satisfies Assumptions 2. Assume that* $\gamma > 1$ *and* $\frac{1}{2}(1/\gamma^{\alpha} + \gamma^{\alpha/q}) < 1$. *Let* CR *be the convergence rate of the algorithm given in* (4.12). *Then*

$$\text{CR} = -\frac{1}{\alpha} E_{\pi}\left(\ln \frac{f(\mathbf{Z}_t + \mathbf{U}_t^1 1_{\{f(\mathbf{Z}_t+\mathbf{U}_t^1)\leq f(\mathbf{Z}_t)\}})}{f(\mathbf{Z}_t)}\right) \tag{4.13}$$

$$= -\frac{1}{\alpha}\int \ln \frac{f(\mathbf{z}+\mathbf{n}1_{f(\mathbf{z}+\mathbf{n})\leq f(\mathbf{z})})}{f(\mathbf{z})} p_{\mathcal{N}}(\mathbf{n}) p_{\pi}(\mathbf{z}) d\mathbf{z} d\mathbf{n} \tag{4.14}$$

where p_{π} *is the density of the invariant probability measure* π *with respect to the Lebesgue measure. Consequently* CR > 0*, i.e., linear* convergence *indeed takes place.*

Proof. Because π is the invariant probability measure of $\mathbf{Z}$, if $\mathbf{Z}_0 \sim \pi$ then for all t, $\mathbf{Z}_t \sim \pi$ such that $E_{\pi}\left[\ln \frac{f(\mathbf{Z}_{t+1})}{f(\mathbf{Z}_t)}\right] = 0$. On the other hand, since $\mathbf{Z}_{t+1} = (\mathbf{Z}_t + \mathbf{U}_t^1 1_{\{f(\mathbf{Z}_t+\mathbf{U}_t^1)\leq f(\mathbf{Z}_t)\}})/\eta^{\star}$ we deduce that

$$E_{\pi}\left[\ln \frac{f(\mathbf{Z}_{t+1})}{f(\mathbf{Z}_t)}\right] = E_{\pi}\left[-\alpha \ln \eta^{\star} + \ln \frac{f(\mathbf{Z}_t + \mathbf{U}_t^1 1_{\{f(\mathbf{Z}_t+\mathbf{U}_t^1)\leq f(\mathbf{Z}_t)\}})}{f(\mathbf{Z}_t)}\right] = 0$$

Thus

$$\text{CR} = -E_{\pi}\left[\ln \eta^{\star}\right] = -\frac{1}{\alpha} E_{\pi}\left(\ln \frac{f(\mathbf{Z}_t + \mathbf{U}_t^1 1_{\{f(\mathbf{Z}_t+\mathbf{U}_t^1)\leq f(\mathbf{Z}_t)\}})}{f(\mathbf{Z}_t)}\right) \tag{4.15}$$

$$= -\frac{1}{\alpha}\int \ln \frac{f(\mathbf{z}+\mathbf{n}1_{f(\mathbf{z}+\mathbf{n})\leq f(\mathbf{z})})}{f(\mathbf{z})} p_{\mathcal{N}}(\mathbf{n}) p_{\pi}(\mathbf{z}) d\mathbf{z} d\mathbf{n} \tag{4.16}$$

where in the previous equation we have used the fact that according to [12, Theorem 10.4.9], π and the maximal irreducibility measure for $\mathbf{Z}$ are equivalent. Hence π is equivalent to the Lebesgue measure. We denoted p_{π} its density. Since $\frac{f(\mathbf{z}+\mathbf{n}1_{f(\mathbf{z}+\mathbf{n})\leq f(\mathbf{z})})}{f(\mathbf{z})} \leq 1$ we see that CR ≥ 0. However CR $= 0$ is impossible as it would imply that $\ln \frac{f(\mathbf{z}+\mathbf{n}1_{f(\mathbf{z}+\mathbf{n})\leq f(\mathbf{z})})}{f(\mathbf{z})} p_{\mathcal{N}}(\mathbf{n}) p_{\pi}(\mathbf{z}) = 0$ almost everywhere. □

The fact that the convergence rate CR is strictly positive is equivalent to having the asymptotic probability of success satisfying PS $< 1/(q+1)$. In the case where the target probability of success is $1/5$ (as proposed in [16, 15]), this implies PS $< 1/5$. Hence we find that when convergence occurs the asymptotic probability of success is strictly smaller than $1/5$.

In the case of a non elitist algorithm, it is not easy to obtain the sign of the convergence rate and one needs to resort to numerical simulation (see [1]).

4.4. Linear convergence in expectation. Linear convergence in expectation is formulated in the next theorem. The proof follows the lines of Theorem 5.3 in [3].

THEOREM 4.7. *Let the* $(1+1)$*-ES with generalized one-fifth success rule optimize* $h = g \circ f$ *where* $g \in \mathcal{M}$ *and* f *satisfies Assumptions 2. Assume that* $\gamma > 1$ *and* $\frac{1}{2}(1/\gamma^{\alpha} + \gamma^{\alpha/q}) < 1$. *Then for all initial condition* $(\mathbf{X}_0, \sigma_0) = (\mathbf{x}_0, \sigma_0) \in \mathbb{R}^n_{\neq} \times \mathbb{R}^+_{>}$

$$\lim_{t\to\infty} E_{\frac{\mathbf{x}_0}{\sigma_0}}\left[\ln \frac{\|\mathbf{X}_{t+1}\|}{\|\mathbf{X}_t\|}\right] = -\text{CR} \tag{4.17}$$

and

$$\lim_{t\to\infty} E_{\frac{\mathbf{x}_0}{\sigma_0}}\left[\ln\frac{\sigma_{t+1}}{\sigma_t}\right] = -\mathrm{CR} \ . \tag{4.18}$$

Proof. The conditions of Theorem 5.3 of [3] are satisfied. Hence we can conclude to the linear convergence in expectation for all initial condition $(\mathbf{x}_0, \sigma_0)$ such that $V(\mathbf{x}_0/\sigma_0) < \infty$ where V is a function satisfying

$$\Delta V(\mathbf{z}) \leq -(|\ln\|\mathbf{z}\|| + 1) + b1_C(\mathbf{z}) \ . \tag{4.19}$$

Let us show that the previous condition is satisfied for a function proportional to the geometric drift function of Theorem 3.11. This will hence imply that the initial condition $(\mathbf{x}_0, \sigma_0)$ can be taken in $\mathbb{R}^n_{\neq} \times \mathbb{R}^+_>$.

Indeed, consider the function $\tilde{V}$ given in (3.26) (to avoid ambiguity we denote $\tilde{V}$ the function originally denoted V). We have proven that it satisfies a drift condition for geometric ergodicity, that is, there exists some petite set C, some constants $b < \infty$ and $0 < \vartheta < 1$ such that

$$\Delta\tilde{V} \leq (\vartheta - 1)\tilde{V}(\mathbf{z}) + b1_C(\mathbf{z}) \ . \tag{4.20}$$

Since in Lemma 4.2 we have proven that $|\ln\|\mathbf{z}\|| \leq K_{4.2}\tilde{V}(\mathbf{z})$ and $\tilde{V} \geq 1$, the following inequality holds: $|\ln\|\mathbf{z}\|| + 1 \leq (K_{4.2}+1)\tilde{V}(\mathbf{z})$. We deduce that $(\vartheta - 1)(K_{4.2}+1)\tilde{V}(\mathbf{z}) \leq (\vartheta - 1)(|\ln\|\mathbf{z}\|| + 1)$ and hence

$$(\vartheta - 1)(K_{4.2} + 1)\tilde{V}(\mathbf{z})/(1 - \vartheta) \leq -(|\ln\|\mathbf{z}\|| + 1) \tag{4.21}$$

Let us take $V = (K_{4.2} + 1)/(1 - \vartheta)\tilde{V}$, (4.20) implies that

$$\Delta V \leq (\vartheta - 1)V(\mathbf{z}) + b\frac{K_{4.2} + 1}{1 - \vartheta}1_C(\mathbf{z}) \leq -(|\ln\|\mathbf{z}\|| + 1) + b\frac{K_{4.2} + 1}{1 - \vartheta}1_C(\mathbf{z})$$

where we have used (4.21) for the latter inequality. Since $V(\mathbf{z}) < \infty$ whenever $f(\mathbf{z}) < \infty$ and $1/f(\mathbf{z}) < \infty$ we deduce from Theorem 5.3 in [3] that the limits (4.17) and (4.18) hold for all $\mathbf{x}_0/\sigma_0$ such that $f(\mathbf{x}_0/\sigma_0) < \infty$ and $1/f(\mathbf{x}_0/\sigma_0) < \infty$, i.e., for all $(\mathbf{x}_0, \sigma_0) \in \mathbb{R}^n_{\neq} \times \mathbb{R}^+_>$. □

4.5. Consequences of Geometric Ergodicity: Adaptivity at a Geometric Rate. The geometric ergodicity translates that the invariant probability distribution is reached geometrically fast. It implies that from any starting point in $\mathbb{R}^n_{\neq} \times \mathbb{R}^+_>$, the expected (step-size) log progress $E_{\frac{\mathbf{x}_0}{\sigma_0}}[\ln\frac{\sigma_{t+1}}{\sigma_t}]$ approaches the convergence rate CR geometrically fast. More precisely we have the following theorem:

THEOREM 4.8. *Let the* $(1+1)$*-ES with generalized one-fifth success rule optimize* $h = g \circ f$ *where* $g \in \mathcal{M}$ *and* f *satisfies Assumptions 2. Assume that* $\gamma > 1$ *and* $\frac{1}{2}(1/\gamma^\alpha + \gamma^{\alpha/q}) < 1$. *Then there exists* $r > 1$ *and* $R < \infty$ *such that for all initial condition* $(\mathbf{X}_0, \sigma_0) = (\mathbf{x}_0, \sigma_0) \in \mathbb{R}^n_{\neq} \times \mathbb{R}^+_>$

$$\sum_t r^t |E_{\frac{\mathbf{x}_0}{\sigma_0}} \ln\frac{\sigma_{t+1}}{\sigma_t} - \mathrm{CR}| \leq RV(\mathbf{x}_0/\sigma_0) \ . \tag{4.22}$$

This equation implies in particular that for any initial condition $(\mathbf{x}_0, \sigma_0) \in \mathbb{R}^n_{\neq} \times \mathbb{R}^+_>$

$$|E_{\frac{\mathbf{x}_0}{\sigma_0}} \ln\frac{\sigma_{t+1}}{\sigma_t} - \mathrm{CR}|r^t \xrightarrow[t\to\infty]{} 0 \tag{4.23}$$

where r is independent of the starting point.

Proof. Under the assumptions of the theorem we have proven in Theorem 3.11 that V defined in (3.26) satisfies a geometric drift function. Hence according to (3.11) there exists $R > 0$ and $r > 1$ such that for any starting point $\mathbf{z}_0$ in the set $S_V = \{\mathbf{z} : V(\mathbf{z}) < \infty\}$

$$\sum_t r^t \|P^t(\mathbf{z}_0, .) - \pi\|_V \leq RV(\mathbf{z}_0) \tag{4.24}$$

where $\|\nu\|_V = \sup_{g:|g|\leq V} |\nu(g)|$. Remark that if V satisfies a geometric drift condition, then kV satisfies also a geometric drift condition for any constant $k \geq 1$. Consider the function $g(\mathbf{z}) = \ln \gamma \left(\frac{q+1}{q} E[1_{\{f(\mathbf{z}+\mathcal{N})\leq f(\mathbf{z})\}}] - \frac{1}{q} \right)$. Then $|g(\mathbf{z})| \leq \ln \gamma(q+2)/q$ is bounded and hence $|g(\mathbf{z})| \leq \underbrace{\ln \gamma(q+2)/q}_{k} V(\mathbf{z})$. In addition, $E_{\mathbf{z}_0}[\ln \sigma_{t+1}/\sigma_t] = E_{\mathbf{z}_0}[g(\mathbf{Z}_t)] = P^t(\mathbf{z}_0, .)(g)$ which yields

$$|E_{\mathbf{z}_0}\left[\ln \frac{\sigma_{t+1}}{\sigma_t}\right] - \mathrm{CR}| = |\int g(\mathbf{z}) P^t(\mathbf{z}_0, d\mathbf{z}) - \int \pi(d\mathbf{z}) g(\mathbf{z})| =$$
$$|(P^t(\mathbf{z}_0, .) - \pi)(g)| \leq \|P^t(\mathbf{z}_0, .) - \pi\|_{kV}$$

and thus according to (4.24) there exists $R > 0$ and $r > 1$ such that for any $(\mathbf{x}_0, \sigma_0)$

$$\sum_t r^t |E_{\frac{\mathbf{x}_0}{\sigma_0}}\left[\ln \frac{\sigma_{t+1}}{\sigma_t}\right] - \mathrm{CR}| \leq RV\left(\frac{\mathbf{x}_0}{\sigma_0}\right) \quad . \tag{4.25}$$

This equation implies in particular that, for any initial condition $(\mathbf{x}_0, \sigma_0)$

$$|E_{\frac{\mathbf{x}_0}{\sigma_0}}\left[\ln \frac{\sigma_{t+1}}{\sigma_t}\right] - \mathrm{CR}| r^t \xrightarrow[t\to\infty]{} 0 \tag{4.26}$$

where r is independent of the starting point. □

Remark that we only derived a result for the log-step-size progress in the previous theorem. We believe that a similar result for the log-progress $\ln \|\mathbf{X}_{t+1}\|/\|\mathbf{X}_t\|$ can also be derived. However it appears to be more technical to control the corresponding g-function (see proof) by V.

Our last result derives from the Central Limit Theorem (CLT), which is also a consequence of the the geometric ergodicity. It describes the speed of convergence of the result obtained from applying the LLN. We remind first a CLT result for MC extracted from [12]. Given a function g, $S_t(g) = \sum_{k=1}^t g(\mathbf{Z}_k)$.

THEOREM 4.9 (Theorem 17.0.1, Theorem 16.0.1 in [12]). *Suppose that $\mathbf{Z}$ is a positive Harris chain with invariant probability measure π and is aperiodic. Suppose that $\mathbf{Z}$ satisfies a geometric drift in the sense of (3.10). Let g be a function on $\mathcal{Z}$ that satisfies $g^2 \leq V$ and let $\bar{g}$ denote the centered function $\bar{g} = g - \int g d\pi$. Then the constant*

$$\gamma_g^2 = E_\pi[\bar{g}^2(\mathbf{Z}_0)] + 2\sum_{k=1}^{\infty} E_\pi[\bar{g}(\mathbf{Z}_0)\bar{g}(\mathbf{Z}_k)]$$

is well defined, non-negative and finite, and coincides with the asymptotic variance

$$\lim_{t\to\infty} \frac{1}{t} E_\pi[(S_t(\bar{g}))^2] = \gamma_g^2 \quad .$$

If $\gamma_g^2 > 0$ then the Central Limit Theorem holds for the function g, that is for any initial condition $\mathbf{z}_0$

$$\lim_{t\to\infty} P_{\mathbf{z}_0}\left\{(t\gamma_g^2)^{-1/2} S_t(\bar{g}) \leq x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \ .$$

If $\gamma_g^2 = 0$, then $\lim_{t\to\infty} \frac{1}{\sqrt{t}} S_t(g) = 0$ a.s.

THEOREM 4.10. *Let the (1+1)-ES with generalized one-fifth success rule optimize $h = g \circ f$ where $g \in \mathcal{M}$ and f satisfies Assumptions 2. Assume that $\gamma > 1$ and $\frac{1}{2}(1/\gamma^\alpha + \gamma^{\alpha/q}) < 1$. Then for any initial condition $(\mathbf{x}_0, \sigma_0) \in \mathbb{R}^n_{\neq} \times \mathbb{R}^+_{>}$*

$$\lim_{t\to\infty} P_{\frac{\mathbf{x}_0}{\sigma_0}}\left\{\frac{\sqrt{t}}{\gamma_g}\left(\frac{1}{t}\ln\frac{\sigma_t}{\sigma_0} - \mathrm{CR}\right)\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \exp(-u^2/2) du$$

where $\gamma_g^2 = \frac{1}{t} E_\pi[(\ln\frac{\sigma_t}{\sigma_0} - t\mathrm{CR})^2] > 0$.

Proof. We have seen that $\frac{1}{t}\ln\frac{\sigma_t}{\sigma_0} = \frac{1}{t}\sum_{k=0}^{t-1} \ln\frac{\sigma_{k+1}}{\sigma_k} = \frac{1}{t}\sum_{k=0}^{t-1} g_{4.10}(\mathbf{Z}_k, \mathbf{U}_k^1)$ with

$$g_{4.10}(\mathbf{Z}_k, \mathbf{U}_k^1) = \ln\gamma\left(\left(\frac{1+q}{q}\right) 1_{\{f(\mathbf{Z}_k+\mathbf{U}_k^1)\leq f(\mathbf{Z}_k)\}} - \frac{1}{q}\right) \ .$$

Because the definition domain of $g_{4.10}$ is $\mathbb{R}^n_{\neq} \times \mathbb{R}^n$, we cannot directly compare it to the geometric drift function V and verify whether $g^2 \leq V$. However let us consider not only $\mathbf{Z}_t$ but the couple $(\mathbf{Z}_t, \mathbf{U}_t^1)$ where the $\mathbf{U}_t^i$ are i.i.d. distributed according to $\mathcal{N}(0, I_n)$. Clearly $(\mathbf{Z}_t, \mathbf{U}_t^1)$ is an homogeneous Markov Chain that will inherit the properties of $\mathbf{Z}_t$. Typically the chain is φ-irreducible with respect to the Lebesgue measure on $\mathbb{R}^n_{\neq} \times \mathbb{R}^n$, aperiodic and $D_{[l_1,l_2]} \times \mathbb{R}^n$ are some small sets for the chain. The function $\tilde{V}(\mathbf{z}, \mathbf{u}) = V(\mathbf{z})$ (with V defined in (3.26)) satisfies a geometric drift in the sense of (3.10) as from the small set shape we see that we only need to control the chain outside $D_{[l_1,l_2]}$ while $\mathbf{u} \in \mathbb{R}^n$. The invariant probability distribution of the chain is $\pi \otimes p$.

To be able to apply Theorem 4.9, we need to verify that $g_{4.10}^2 \leq K_{4.10}\tilde{V}$ with $K_{4.10}$ a constant larger 1 (as the same arguments used before holds here as well, namely if $\tilde{V}$ satisfies a geometric drift condition, then every multiple of $\tilde{V}$ also satisfies a geometric drift condition, provided the multiplication constant is larger one to still ensure that the function is larger 1). Let us now remark that $g_{4.10}^2 \leq ((\ln\gamma)(2+q)/q)^2$ and hence $g_{4.10}^2 \leq (((\ln\gamma)(2+q)/q)^2 + 1)\tilde{V}$. Using Theorem 4.9, we know that γ_g^2 is well defined and cannot equal 0 otherwise it would imply that $\frac{1}{\sqrt{t}}\ln\frac{\sigma_t}{\sigma_0} = 0$ that would contradict the fact that $\lim_{t\to\infty}\frac{1}{t}\ln\frac{\sigma_t}{\sigma_0} = -\mathrm{CR} \neq 0$. Hence $\gamma_g^2 > 0$ and we conclude using Theorem 4.9. ∎

5. Discussion. Using the methodology developed in [3], we have proven the global linear convergence of the $(1+1)$-ES with generalized one fifth success rule on continuously differentiable positively homogeneous functions (with degree alpha) composed (to the left) by strictly increasing functions. This class of functions includes non quasi-convex functions and non continuous functions, an untypical setting for proving linear convergence of optimization algorithms in general.

Given a function $h = g \circ f$ where f is positively homogeneous with degree α and continuously differentiable, linear convergence holds under the condition that the step-size increases in case of success, i.e., $\gamma > 1$ and that

$$\frac{1}{2}\left(1/\gamma^\alpha + \gamma^{\alpha/q}\right) < 1 \ . \tag{5.1}$$

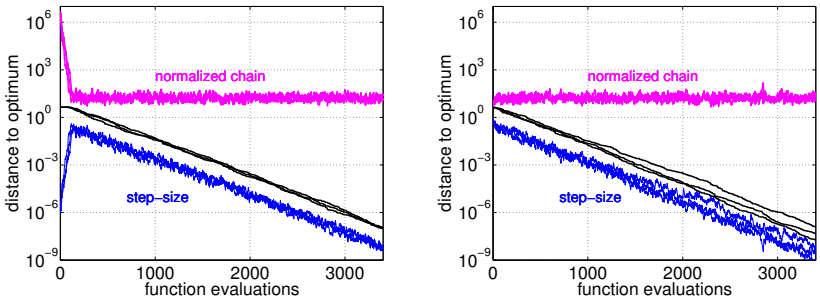


FIG. 5.1. *Convergence simulations of the* $(1+1)$*-ES with generalized one-fifth success rule on spherical functions* $f(\mathbf{x}) = g(\|\mathbf{x}\|)$ *for* $g \in \mathcal{M}$ *in dimension* $n = 20$ *(parameters* $\gamma = \exp(1/3)$ *and* $q = 4$*). Each plot is in log scale and depicts in black the distance to optimum, i.e.,* $\|\mathbf{X}_t\|$*, in blue the respective step-size* σ_t *and in magenta the norm of the normalized chain* $\|\mathbf{Z}_t\|$*. The* x*-axis is the number of function evaluations corresponding to the iteration index* t*. On the left the simulation is voluntarily started with a too small step-size equal to* $\sigma_0 = 10^{-6}$ *to illustrate the adaptivity ability of the algorithm. On the right, the initial step-size equals* 1*.*

Especially, this condition only depends on the function via α and is thus the same for *any* $g \circ f$ with f is positively homogeneous with degree α and continuously differentiable (plus satisfying Assumptions 2).

Because on a linear function the probability of success equals $1/2$, the condition in (5.1) corresponds to the expected inverse of the step-size change to the alpha–on a linear function–being strictly smaller than 1. In other words, the step-size should increase on a linear function. While this latter condition seems a reasonable requirement for an adaptive step-size algorithm, let us point out that some algorithms like the $(1, 2)$-ES with self adaptation fail to satisfy this condition (see [5] for a thorough analysis of this problem). We believe that the fact that linear convergence on the class of functions investigated in the paper is related to increasing the step-size on linear functions illustrates the strong need to study simple models like the linear function when designing CB-SARS algorithms.

Our statements for the linear convergence hold for any initial solution and any initial step-size. This latter property reflects the main advantage of *adaptive* step-size methods: the initial step-size does not need to be too carefully chosen to ensure good convergence properties. Note that methods like simulated annealing or Simultaneous perturbation stochastic approximation (SPSA) [17] do not share this nice property and are very sensitive to the choice of some parameters that unfortunately need to be adjusted by the user.

The adaptation phase, i.e., how long it takes such that the linear convergence is "observed" (see Figure 5.1 left) is related theoretically to the convergence speed of $(\mathbf{Z}_t)_{t\in\mathbb{N}}$ to its stationary distribution. We have proven a geometric drift that ensures that this convergence is geometrically fast and the geometric rate is independent of the initial condition.

Previous attempts to analyze CB-SARS always focused on much smaller classes of functions. The sphere function was analyzed in [1, 10, 9], and a specific class of convex quadratic functions was also analyzed in [7, 8]. Our proof is more general and holds on a wider class of function. Let us however investigate more closely how to apply the analysis on the sphere function or on convex-quadratic functions. Let us note that the sphere function $f(\mathbf{x}) = \|\mathbf{x}\|^2$ can be written for all α as $g_\alpha(\|\mathbf{x}\|^\alpha)$ for $g_\alpha : x \in \mathbb{R}^+ \mapsto x^{2/\alpha}$ ($g_\alpha \in \mathcal{M}$). The function $\mathbf{x} \mapsto \|\mathbf{x}\|^\alpha$ being positively

homogeneous with degree α, for a given dimension n, the condition $\alpha \leq n$ formulated in Assumptions 2 can always be satisfied on the sphere function by choosing the appropriate (α, g_α). The same will be true for any norm function $\mathbf{x} \mapsto N(\mathbf{x})$, in particular for any convex-quadratic function. Assuming that the underlying norm function $N()$ is continuously differentiable, we, in addition, need to have $\alpha > 1$ such that the function $\mathbf{x} \mapsto N(\mathbf{x})^\alpha$ stays continuously differentiable. Hence on the sphere, on a convex-quadratic function or on any norm function (assuming the continuous differentiability of $N()$), linear converge will hold for a given (γ, q) if there exists $1 < \alpha \leq n$ such that $\frac{1}{2}\left(1/\gamma^\alpha + \gamma^{\alpha/q}\right) < 1$.

We have obtained a comprehensive expression for the convergence rate of the algorithm as

$$\mathrm{CR} = -\ln\gamma\left(\frac{q+1}{q}\mathrm{PS} - \frac{1}{q}\right)$$

where PS is the asymptotic probability of success. We have proven that $\mathrm{CR} > 0$ which means that linear convergence indeed occurs. This implies that when convergence occurs the probability of success is strictly smaller than $1/(q+1)$, i.e., strictly smaller than $1/5$ using the traditional $1/5$ as target success probability (corresponding to $q = 4$).

In the case of algorithms that do not guarantee the monotony of $f(\mathbf{X}_t)$, the sign of the "convergence" rate is usually not possible to obtain analytically and one needs to resort to Monte Carlo simulations [1].

Besides the sign of the convergence rate, one would like to extract more properties of CR like the dependence in the dimension, or the dependence in the condition number if one considers convex-quadratic functions. This seems to be hard to achieve with the present approach as CR depends on the stationary distribution of the normalized chain for which little is known besides its existence. However Monte Carlo simulations are natural and always possible to estimate those dependencies. The present paper gives a rigorous framework to perform those Monte Carlo simulations. Note that using more ad-hoc techniques, it is possible to obtain some dependencies in the dimension or condition number [10, 9, 7, 8].

While its convergence proof "resisted" for more than 40 years, the algorithm analyzed is simple and relatively straightforward as witnessed by the fact that it was already proposed very early and by various researchers in parallel. We however want to emphasize that nowadays this algorithm should mainly have an academic purpose. Indeed more robust comparison-based adaptive algorithm exist, namely the CMA-ES algorithm where in addition to the step-size, a full covariance matrix is adapted [6].

Last we want to emphasize two points:

1) A common misconception is that randomized methods are good for global optimization and bad for local optimization. The present paper by proving a global linear convergence for a CB-SARS proves the contrary. In addition, comparisons of the CMA-ES algorithm–the state-of-the-art comparison based adaptive algorithm– with BFGS and NEWUOA show also that CMA-ES is competitive on (unimodal) composite of convex-quadratic functions provided they are significantly non-separable and non-convex [2]. This result does not come as a surprise as CB-SARS and CMA-ES were designed first as robust local search and carefully investigated to optimally solve simple functions like the sphere, the linear function and convex-quadratic functions.

2) The present paper illustrates that the theory of Markov Chains with discrete time and continuous state space is useful and powerful for the analysis of CB-SARS.

We believe that the present analysis can be extended further for the case of stochastic functions or for the case of algorithms where a covariance matrix is adapted in addition to a step-size.

REFERENCES

[1] A. Auger. Convergence results for (1,λ)-SA-ES using the theory of φ-irreducible markov chains. *Theoretical Computer Science*, 334(1-3):35–69, 2005.

[2] A. Auger, N. Hansen, J. M. Perez Zerpa, R. Ros, and M. Schoenauer. Empirical comparisons of several derivative free optimization algorithms. In *Acte du 9ime colloque national en calcul des structures*, volume 1, pages 481–486, 2009.

[3] Anne Auger and Nikolaus Hansen. On Proving Linear Convergence of Comparison-based Step-size Adaptive Randomized Search on Scaling-Invariant Functions via Stability of Markov Chains, 2013. ArXiv eprint.

[4] L. Devroye. The compound random search. In *International Symposium on Systems Engineering and Analysis*, pages 195–110. Purdue University, 1972.

[5] N. Hansen. An analysis of mutative σ-self-adaptation on linear fitness functions. *Evolutionary Computation*, 14(3):255–275, 2006.

[6] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

[7] Jens Jägersküpper. Rigorous runtime analysis of the (1+1)-ES: 1/5-rule and ellipsoidal fitness landscapes. In LNCS, editor, *Foundations of Genetic Algorithms: 8th International Workshop, FoGA 2005*, volume 3469, pages 260–281, 2005.

[8] Jens Jägersküpper. How the (1+1) ES using isotropic mutations minimizes positive definite quadratic forms. *Theoretical Computer Science*, 361(1):38–56, 2006.

[9] Jens Jägersküpper. Probabilistic runtime analysis of ($1\dagger\lambda$) evolution strategies using isotropic mutations. pages 461–468. ACM Press, 2006.

[10] Jens Jägersküpper. Algorithmic analysis of a basic evolutionary algorithm for continuous optimization. *Theoretical Computer Science*, 379(3):329–347, 2007.

[11] J. Matyas. Random optimization. *Automation and Remote control*, 26(2), 1965.

[12] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, New York, 1993.

[13] John Ashworth Nelder and R Mead. A simplex method for function minimization. *The Computer Journal*, pages 308–313, 1965.

[14] Michael J.D. Powell. Developments of newuoa for unconstrained minimization without derivatives. Technical Report DAMTP 2007/NA05, CMS, University of Cambridge, Cambridge CB3 0WA, UK, June 2007.

[15] I. Rechenberg. *Evolutionstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog Verlag, Stuttgart, 1973.

[16] M. Schumer and K. Steiglitz. Adaptive step size random search. *IEEE Transactions on Automatic Control*, 13(3):270–276, 1968.

[17] J.C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *Automatic Control, IEEE Transactions on*, 37(3):332–341, 1992.