Analysis of Evolution Strategies with the Optimal Weighted Recombination

Chun-kit Au
Toronto, Ontario, Canada
vincentau@alumni.cuhk.net

Ho-fung Leung
The Chinese University of Hong Kong
Shatin, Hong Kong
lhf@cuhk.edu.hk

ABSTRACT

This paper studies the performance for evolution strategies with the optimal weighed recombination on spherical problems in finite dimensions. We first discuss the different forms of functions that are used to derive the optimal recombination weights and step size, and then derive an inequality that establishes the relationship between these functions. We prove that using the expectation of random variables to derive the optimal recombination weights and step size can be disappointing in terms of the expected performance of evolution strategies. We show that using the realizations of random variables is a better choice. We generalize the results to any convex functions and establish an inequality for the normalized quality gain. We prove that the normalized quality gain of the evolution strategies have a better and robust performance when they use the optimal recombination weights and the optimal step size that are derived from the realizations of random variables rather than using the expectations of random variables.

CCS CONCEPTS

• **Mathematics of computing** → **Evolutionary algorithms**; • **Theory of computation** → **Bio-inspired optimization**;

KEYWORDS

Evolution Strategies, Recombination Weights

ACM Reference Format:

Chun-kit Au and Ho-fung Leung. 2018. Analysis of Evolution Strategies with the Optimal Weighted Recombination. In *GECCO '18: Genetic and Evolutionary Computation Conference, July 15–19, 2018, Kyoto, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3205455.3205632

1 INTRODUCTION

Evolution strategies (ESs) are random search heuristics that are designed to optimize objective functions that map a n-dimensional space $\mathbb{R}^n$ to $\mathbb{R}$. In the evolution strategies with weighted recombination, the parent $\mathbf{x}_k \in \mathbb{R}^n$ at iteration k generates λ number of new candidate solutions $\left(\mathbf{y}_k^i\right)_{1\leq i\leq\lambda} \in \mathbb{R}^n$. These new candidate solutions, which are called *offspring*, are generated by adding the parent the independent random vectors $\left(\mathbf{z}_k^i\right)_{1\leq i\leq\lambda} \in \mathbb{R}^n$ that are drawn from the n-dimensional standard normal distribution $\mathcal{N}(\mathbf{0},\mathbf{I})$, where $\mathbf{0} \in \mathbb{R}^n$ is a n-dimensional zero vector and $\mathbf{I}$ is a $n\times n$ identity matrix. These offspring are computed as $\mathbf{y}_k^i = \mathbf{x}_k + \sigma_k \cdot \mathbf{z}_k^i$ for $1 \leq i \leq \lambda$ where $\sigma_k \in \mathbb{R}_{>0}$ is called the step size or mutation strength. Every offspring is evaluated on a given objective optimization function $f : \mathbb{R}^n \to \mathbb{R}$ and then ordered so it gives $f(\mathbf{y}_k^{1:\lambda}) \leq \cdots \leq f(\mathbf{y}_k^{\lambda:\lambda})$ where $i : \lambda$ denotes the index of the i-th best offspring. The new parent $\mathbf{x}_{k+1}$ at the next iteration $k+1$ is then formed by computing $\mathbf{x}_{k+1} = \mathbf{x}_k + \sigma_k \mathcal{Z}\mathbf{w}$ where $\mathcal{Z} \in \mathbb{R}^{n\times\lambda}$ is a $n\times\lambda$ matrix $\mathcal{Z} := \left[\mathbf{z}_k^{1:\lambda} \cdots \mathbf{z}_k^{\lambda:\lambda}\right]$ and the column vector $\mathbf{z}_k^{i:\lambda}$ denotes the random vector used by the i-th best offspring. The λ-th dimensional vector $\mathbf{w} \in \mathbb{R}^\lambda$ is the recombination weights of the evolution strategies such that the conditions $[\mathbf{w}]_1 \geq \cdots \geq [\mathbf{w}]_\lambda$ and $\sum_{i=1}^{\lambda}[\mathbf{w}]_i = 1$ hold.

In the literature, there are many works that analyze the performance of evolution strategies empirically and theoretically. Empirical methods always provide us the easiest way to understand how evolution strategies perform. On the other hand, theoretical studies help us to better understand evolution strategies, give us an idea of the 'optimal' scenario in which evolution strategies perform the best and these 'optimal' scenarios cannot be empirically found. For instance, the study [6] analyzed the optimal values for the parent size μ in the $(\mu/\mu_w, \lambda)$ evolution strategies on spherical functions. Another theoretical study [5] proved that the perpendicularity condition does not guarantee the optimal performance on spherical functions when evolution strategies use the cumulative step size adaptation mechanism.

In this paper, we are interested in studying the recombination weights that give the optimal expected performance of evolution strategies on spherical functions. The first analysis of the evolution strategies with the optimal weighted recombination was done in [2, 3]. It derives the optimal recombination weights on the spherical function in infinite dimensions, and the optimal performance of the evolution stratgies is attained when the recombination weights are equal to the expectations of order statistics of a standard normal distribution. Another paper [1] investigated the optimal recombination weights on general convex quadratic functions and found that the optimal recombination weights are independent of the Hessian of the objective function and the optimal step size depends heavily on the Hessian of the objective function.

While all these theoretical studies have rigorously derived the optimal recombination weights where the objective is to maximize the so-called *quality gain* over one iteration, the function for the optimal recombination weights and step size involves the expectations of random variables used in the evolution strategies. Precisely, the optimal recombination weights $\mathbf{w}^* \in \mathbb{R}^\lambda$ and the optimal step size $\sigma^* \in \mathbb{R}_{>0}$ are derived by solving the following problem in the

GECCO '18, July 15–19, 2018, Kyoto, Japan

ACM ISBN 978-1-4503-5618-3/18/07…$15.00
https://doi.org/10.1145/3205455.3205632

form of:

$$\max_{\sigma \in \mathbb{R}_{>0}, \mathbf{w} \in \mathbb{R}^\lambda} g\Big(\mathbf{x}, \sigma, \mathbf{w}, \mathbb{E}[\xi(\omega)]\Big)$$

where g is a function mapping defined as $g : \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^\lambda \times \mathbb{R}^{n\times\lambda} \to \mathbb{R}$ and is the function for deriving the optimal recombination weights and step size. The notation $\xi : \Omega \to \mathbb{R}^{n\times\lambda}$ is a $n \times \lambda$ random matrix used in the evolution strategies and is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The notation $\mathbb{E}$ denotes the expectation on the random matrix ξ. We call all these problems of finding the optimal recombination weights and step size in [1–3] the *Expected Value (EV) Problems*. Using the expectations of random variables in the function g probably provides an excellent starting point to find out any improvement techniques which might be useful for evolution strategies. It is therefore interested to investigate whether the optimal recombination weights and step size are limited to the scenario when the expectations of random variables are realized only. In addition, the expectation of the function g might turn out to be different from the value of the function using the optimal $\mathbf{w}^*$ and σ^*. It is therefore of interest to examine under what conditions the following equality will hold:

$$\mathbb{E}\left[g\Big(\mathbf{x}, \sigma^*, \mathbf{w}^*, \xi(\omega)\Big)\right] = g\Big(\mathbf{x}, \sigma^*, \mathbf{w}^*, \mathbb{E}[\xi(\omega)]\Big)$$

In this paper, we call the value on the left hand side of the previous equation the *Expectation of using EV solution (EEV)*.

Besides studying the EV problem for the optimal recombination weights and step size, we also analyze the performance of the evolution strategies when the optimal recombination weights and step size are derived from the realizations of random variables. That is, we want to understand the expected performance of evolution strategies in the form of

$$\mathbb{E}\left[\max_{\sigma \in \mathbb{R}_{>0}, \mathbf{w} \in \mathbb{R}^\lambda} g\Big(\mathbf{x}, \sigma, \mathbf{w}, \xi(\omega)\Big)\right]$$

We call the above expectation the *Expected Perfect Information Value (EPIV)*. The EPIV calculates the expected performance when the information for the realizations of random variables are *perfectly* known. This is a rather ideal scenario because in practice we would have to decide the value of step size before the random variables are realized. We can "wait and see" until the random variables are realized and then derive the optimal recombination weights for evolution strategies. Therefore, it is interesting to investigate the optimal recombination weights and step size where the problem is in the form of

$$\max_{\sigma \in \mathbb{R}_{>0}} \mathbb{E}\left[\max_{\mathbf{w} \in \mathbb{R}^\lambda} g\Big(\mathbf{x}, \sigma, \mathbf{w}, \xi(\omega)\Big)\right]$$

We call this the *Expected Imperfect Information Value (EIIV)*.

This paper is organized as follows. Section (2) discusses the inequality for different forms of functions for deriving the optimal recombination weights and step size. Section (3) reviews different forms of the normalized quality gain for evolution strategies and derives the optimal recombination weights which are based on the realizations of random variables. Numerical simulations are presented in Section (4). We conclude this paper in Section (5).

1.1 Notations

We use the following mathematical notations throughout the paper. The inverse of a square matrix $\mathbf{A}$ is denoted as $\mathbf{A}^{-1}$ such that the equation $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ holds, where $\mathbf{I}$ is an identity matrix. The pseudo inverse (or Moore-Penrose inverse) of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}^+$ that fulfils these four conditions: (I) $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$, (II) $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$, (III) $(\mathbf{A}\mathbf{A}^+)^T = \mathbf{A}\mathbf{A}^+$, and (IV) $(\mathbf{A}^+\mathbf{A})^T = \mathbf{A}^+\mathbf{A}$. The pseudo inverse $\mathbf{A}^+$ exists for any matrix $\mathbf{A}$. Let $\xi : \Omega \to \mathbb{R}^{n\times\lambda}$ be a $n \times \lambda$ random matrix such that it is a measurable mapping from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathbb{R}^{n\times\lambda}$, where Ω denotes the set of possible outcomes, $\mathcal{F}$ is a sigma algebra of the set of events, and $\mathbb{P}$ denotes the probability measure. Sometimes we also let $\xi : \Omega \to \mathbb{R}^{n\times\lambda}$ be a $n \times \lambda$ random matrix such that it is a measurable mapping from the probability space $(\Xi, \mathcal{B}, \mathbb{P})$ to $\mathbb{R}^{n\times\lambda}$, where $\Xi \subset \mathbb{R}^{n\times\lambda}$ denotes the support of measure $\mathbb{P}$ and $\mathcal{B}$ is the sigma algebra of all Borel subsets of Ξ[1].

2 BOUNDS ON GENERAL FUNCTION

2.1 Preliminaries

In this section, we prove the inequality for the expected performance of evolution strategies when different forms of the functions for deriving the optimal recombination weights and step size are used. We start by formally defining the EV problem.

Definition 2.1 (Expected Value (EV) Problem for Recombination Weights and Step size). Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the random matrix $\xi : \Omega \to \mathbb{R}^{n\times\lambda}$ and the function mapping $g : \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^\lambda \times \mathbb{R}^{n\times\lambda} \to \mathbb{R}$ as the function for deriving the optimal recombination weights and step size. Assume that the expectation of random matrix $\mathbb{E}[\xi(\omega)] \in \mathbb{R}^{n\times\lambda}$ is well defined and finite. Then for a fixed $\mathbf{x} \in \mathbb{R}^n$, the EV Problem for the optimal recombination weights and step size is in the form of

$$\max_{\sigma \in \mathbb{R}_{>0}, \mathbf{w} \in \mathbb{R}^\lambda} g\Big(\mathbf{x}, \sigma, \mathbf{w}, \mathbb{E}[\xi(\omega)]\Big) \tag{1}$$

For a fixed $\mathbf{x} \in \mathbb{R}^n$, we call its optimal value the *Expected Value (EV)* of the function g and it is denoted by $\mathrm{EV}_{\mathbf{x}}$. We also call its optimal recombination weights and step size the *EV solution* and they are denoted by $\mathbf{w}^*$ and σ^* respectively. We can interchange the order of maximization of the EV problem to get the same optimal recombination weights and step size.[2]

$$\mathrm{EV}_{\mathbf{x}} := \max_{\sigma \in \mathbb{R}_{>0}, \mathbf{w} \in \mathbb{R}^\lambda} g\Big(\mathbf{x}, \sigma, \mathbf{w}, \mathbb{E}[\xi(\omega)]\Big) \tag{2}$$

$$\{\mathbf{w}^*_{\mathbf{x}}, \sigma^*_{\mathbf{x}}\} := \arg\max_{\sigma \in \mathbb{R}_{>0}, \mathbf{w} \in \mathbb{R}^\lambda} g\Big(\mathbf{x}, \sigma, \mathbf{w}, \mathbb{E}[\xi(\omega)]\Big) \tag{3}$$

We now define the *Expectation of Using EV solution* (EEV) that will be useful when we evaluate the quality of EV solutions.

[1] We use the same notation ξ to denote a random matrix and its particular realization. The meaning of such notation will usually be clear from the context. If in doubt, then we write $\xi = \xi(\omega)$ to emphasize that ξ is a random matrix defined on a corresponding probability space.

[2] Suppose g is continuously differentiable with respect to $\mathbf{w}$ and σ. The optimal recombination weights $\mathbf{w}^*$ and step size σ^* satisfy the first order necessary optimality conditions if and only if the equations $\mathbf{0} \in \partial_{\mathbf{w}} g\Big(\mathbf{x}, \sigma, \mathbf{w}^*, \mathbf{Z}\Big)$ and $0 \in \partial_\sigma g\Big(\mathbf{x}, \sigma^*, \mathbf{w}, \mathbf{Z}\Big)$ hold. If the function $g\Big(\mathbf{x}, \sigma, \mathbf{w}, \mathbb{E}[\xi(\omega)]\Big)$ is convex, both equations represent the sufficient conditions for maximizing the equation (1) over $\mathbf{w} \in \mathbb{R}^\lambda$ or $\sigma \in \mathbb{R}_{>0}$

Definition 2.2 (Expectation of using EV solution (EEV) for Recombination Weights and Step size). Consider the probability space $(\Xi, \mathcal{B}, \mathbb{P})$, the random matrix $\xi : \Xi \to \mathbb{R}^{n\times\lambda}$, the function mapping $g : \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^\lambda \times \mathbb{R}^{n\times\lambda} \to \mathbb{R}$ and the optimal recombination weights $\mathbf{w}^*$ and step size σ^* for EV problem in equation (3). Assume that for every fixed $\mathbf{x} \in \mathbb{R}^n, \sigma \in \mathbb{R}_{>0}$ and $\mathbf{w} \in \mathbb{R}^\lambda$, the function $g(\mathbf{x}, \sigma, \mathbf{w}, \cdot)$ is $\mathcal{B}$-measurable. Then for a fixed $\mathbf{x} \in \mathbb{R}^n$, we define the Expectation of Using EV solution (EEV) as

$$\mathrm{EEV}_{\mathbf{x}} := \mathbb{E}\left[g\Big(\mathbf{x}, \sigma^*, \mathbf{w}^*, \xi\Big)\right] \tag{4}$$

The following defines the Expected Perfect Information Value (EPIV) and the Expected Imperfect Information Value (EIIV) .

Definition 2.3 (Expected Perfect Information Value (EPIV) for Recombination Weights and Step size). Consider the probability space $(\Xi, \mathcal{B}, \mathbb{P})$, the random matrix $\xi : \Xi \to \mathbb{R}^{n\times\lambda}$ and the function mapping $g : \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^\lambda \times \mathbb{R}^{n\times\lambda} \to \mathbb{R}$. Assume that for every fixed $\mathbf{x} \in \mathbb{R}^n, \sigma \in \mathbb{R}_{>0}$ and $\mathbf{w} \in \mathbb{R}^\lambda$, the function $g(\mathbf{x}, \sigma, \mathbf{w}, \cdot)$ is $\mathcal{B}$-measurable. Then for a fixed $\mathbf{x} \in \mathbb{R}^n$ and each $\xi \in \Xi$, the $\mathcal{B}$-measurable functions $v_{\mathbf{x}} : \Xi \mapsto \mathbb{R}$, $\mathbf{w}^*_{\mathbf{x}} : \Xi \mapsto \mathbb{R}^\lambda$ and $\sigma^*_{\mathbf{x}} : \Xi \mapsto \mathbb{R}_{>0}$ are defined as

$$v_{\mathbf{x}}(\xi) := \max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \xi) \tag{5}$$

$$\mathbf{w}^*_{\mathbf{x}}(\xi) := \arg\max_{\mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \xi) \tag{6}$$

$$\sigma^*_{\mathbf{x}}(\xi) := \arg\max_{\sigma\in\mathbb{R}_{>0}} g(\mathbf{x}, \sigma, \mathbf{w}, \xi) \tag{7}$$

We call the expectation of v where the problem in the form of equation (5) is solved for each $\xi \in \Xi$ the *Expected Perfect Information Value (EPIV)*

$$\mathrm{EPIV}_{\mathbf{x}} := \mathbb{E}\left[v_{\mathbf{x}}(\xi)\right] = \mathbb{E}\left[\max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g\Big(\mathbf{x}, \sigma, \mathbf{w}, \xi\Big)\right] \tag{8}$$

We call the set of the optimal recombination weights $\mathbf{w}^*_{\mathbf{x}}(\xi)$ and the set of the optimal step size $\sigma^*_{\mathbf{x}}(\xi)$ for all $\xi \in \Xi$ as the *EPIV solutions.*

Definition 2.4 (Expected Imperfect Information Value (EIIV) for Recombination Weights and Step size). Following Definition (2.3), for a fixed $\mathbf{x} \in \mathbb{R}^n$ and each $\xi \in \Xi$, the $\mathcal{B}$-measurable function $\alpha_{\mathbf{x}} : \Xi \mapsto \mathbb{R}$ is defined as

$$\alpha_{\mathbf{x}}(\xi) := \max_{\mathbf{w}\in\mathbb{R}^\lambda} g\Big(\mathbf{x}, \sigma, \mathbf{w}, \xi\Big) \tag{9}$$

We call the problem of maximizing the expectation of α with respect to step size σ the *Expected Imperfect Information Value (EIIV).*

$$\mathrm{EIIV}_{\mathbf{x}} := \max_{\sigma\in\mathbb{R}_{>0}} \mathbb{E}\left[\alpha_{\mathbf{x}}(\xi)\right] = \max_{\sigma\in\mathbb{R}_{>0}} \mathbb{E}\left[\max_{\mathbf{w}\in\mathbb{R}^\lambda} g\Big(\mathbf{x}, \sigma, \mathbf{w}, \xi\Big)\right] \tag{10}$$

$$\bar{\sigma}^*_{\mathbf{x}} := \arg\max_{\sigma\in\mathbb{R}_{>0}} \mathbb{E}\left[\alpha_{\mathbf{x}}(\xi)\right] \tag{11}$$

We call the set of the optimal recombination weights $\mathbf{w}^*_{\mathbf{x}}(\xi)$ in equation (6) for all $\xi \in \Xi$ and the optimal step size $\bar{\sigma}^*_{\mathbf{x}}$ in equation (11) as *EIIV solutions.*

The difference between EV and EPIV is that in EV problem the objective is to maximize the function g with respect to the expectation of random matrix $\xi(\omega)$. In EPIV, the objective is to maximize the function g with respect to the realizations of the random matrix $\xi \in \Xi$ and then take the expectation of the resulting maximum of all realizations. We will show in the next section how we can relate these two measures by assuming the convexity and the continuity of the function g with respect to the random variable ξ. The difference between EPIV and EIIV is that in EIIV, the objective is to maximize the function g for the realizations with respect to the recombination weights and then take the expectation of the resulting maximum and then maximize it with respect to the step size. We will show in the next section that EPIV is a rather ideal scenario and EIIV is more practical when we want to maximize the expected performance of evolution strategies when the random variables are realized.

2.2 The Inequality EPIV ≤ EV

In this section, we derive an inequality that establishes the relationship between EV, EPIV, EIIV and EEV. Throughout this section, we assume that the maxima of function g with respect to recombination weights and step size is a convex function of ξ. That is, we assume that for a fixed $\mathbf{x} \in \mathbb{R}^n$, $\max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \xi)$ is convex over the set $\xi \in \mathbb{R}^{n\times\lambda}$. We establish our first lemma that proves the EV is always larger than or equal to EPIV.

Lemma 2.5 (EPIV ≤ EV). *Consider the EV in equation* (2)*, the EPIV in equation* (8) *and the function* $g : \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^\lambda \times \mathbb{R}^{n\times\lambda} \to \mathbb{R}$ *for the optimal recombination weights and step size. Assume that for a fixed* $\mathbf{x} \in \mathbb{R}^n$*, the maxima of function* g *with respect to recombination weights and step size* $\max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \xi)$ *is convex over the set* $\xi \in \mathbb{R}^{n\times\lambda}$*. Then for a fixed* $\mathbf{x} \in \mathbb{R}^n$*, the EV is always greater than or equal to the EPIV. That is, the following inequality holds:*

$$\mathrm{EPIV}_{\mathbf{x}} \leq \mathrm{EV}_{\mathbf{x}} \tag{12}$$

Proof. To prove the inequality, we first prove the convexity and the continuity of function g with respect to the set $\xi \in \mathbb{R}^{n\times\lambda}$. We know by assumption that the maxima of function g with respect to recombination weights and step size $\max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \xi)$ is convex over the set $\xi \in \mathbb{R}^{n\times\lambda}$. To prove the continuity, we let ξ be the boundary of the convex set. Consider the sequence $\{\xi_i\}$ such that $\lim_{i\to\infty} \xi_i = \xi$. Now we can write

$$\lim_{i\to\infty} \max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \xi_i)$$
$$\leq \max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \lim_{i\to\infty} \xi_i) = \max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \xi)$$

Also by convexity of $\max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \xi)$,

$$\max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \xi) \leq \lim_{i\to\infty} \max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \xi_i)$$

Therefore we can write

$$\lim_{i\to\infty} \max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \xi_i) = \max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \lim_{i\to\infty} \xi_i)$$

and hence the continuity of g over the convex set of $\xi \in \mathbb{R}^{n\times\lambda}$. Finally, we establish that the function g is a continuous convex function over the convex set of ξ. When ξ is a random matrix, we see by Jensen's inequality that the inequality (12) holds. □

We can immediately establish a corollary about the equality of the EV and the EPIV.

Corollary 2.6 (EPIV = EV). *Following Lemma* (2.5)*, equality holds for equation* (12) *if and only if* $\max_{\sigma\in\mathbb{R}_{>0}, \mathbf{w}\in\mathbb{R}^\lambda} g(\mathbf{x}, \sigma, \mathbf{w}, \xi)$ *is* *linear over the set $\xi \in \mathbb{R}^{n\times\lambda}$ with respect to recombination weights and step size.*

Proof. Consider Jensen's inequality that the equality holds if the function is a linear function on the random variables ξ hence the results. □

2.3 The Inequality EIIV ≤ EPIV

In this section, we prove an inequality that the EPIV is always larger than or equal to the EIIV.

Lemma 2.7 (EIIV ≤ EPIV). *Consider the EPIV in equation* (8)*, the EIIV in equation* (10) *and the function* $g : \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^\lambda \times \mathbb{R}^{n\times\lambda} \to \mathbb{R}$ *for the optimal recombination weights and step size. Then for a fixed* $\mathbf{x} \in \mathbb{R}^n$*, the EPIV is always greater than or equal to the EIIV. That is, the following inequality holds:*

$$\mathrm{EIIV}_{\mathbf{x}} \leq \mathrm{EPIV}_{\mathbf{x}} \tag{13}$$

Proof. First, we know the fact that for any $\mathbf{w} \in \mathbb{R}^\lambda$ and $\sigma \in \mathbb{R}_{>0}$, the inequality

$$\max_{\mathbf{w}\in\mathbb{R}^\lambda} g\Big(\mathbf{x},\sigma,\mathbf{w},\xi\Big) \leq \max_{\sigma\in\mathbb{R}_{>0},\mathbf{w}\in\mathbb{R}^\lambda} g\Big(\mathbf{x},\sigma,\mathbf{w},\xi\Big)$$

clearly holds. Take the expectation on both side of the inequality so we have

$$\mathbb{E}\left[\max_{\mathbf{w}\in\mathbb{R}^\lambda} g\Big(\mathbf{x},\sigma,\mathbf{w},\xi\Big)\right] \leq \mathbb{E}\left[\max_{\sigma\in\mathbb{R}_{>0},\mathbf{w}\in\mathbb{R}^\lambda} g\Big(\mathbf{x},\sigma,\mathbf{w},\xi\Big)\right]$$

It then follows that

$$\max_{\sigma\in\mathbb{R}_{>0}} \mathbb{E}\left[\max_{\mathbf{w}\in\mathbb{R}^\lambda} g\Big(\mathbf{x},\sigma,\mathbf{w},\xi\Big)\right] \leq \mathbb{E}\left[\max_{\sigma\in\mathbb{R}_{>0},\mathbf{w}\in\mathbb{R}^\lambda} g\Big(\mathbf{x},\sigma,\mathbf{w},\xi\Big)\right]$$

hence it proves the results. □

The following lemma establishes the equality of the EPIV and the EIIV.

Lemma 2.8 (EIIV=EPIV). *Following Lemma* (2.7)*, equality holds for equation* (13) *if and only if the function* $g(\mathbf{x},\sigma,\mathbf{w},\xi)$ *has a maximum point* $\bar{\sigma}^*$ *that is independent of* $\xi \in \Xi$*.*

Proof. Assume that for all $\xi \in \Xi$ there exists the optimal recombination weights $\mathbf{w}^*_{\mathbf{x}}(\xi)$ defined in equation (6), and that there exists the optimal step size $\bar{\sigma}^*$ defined in equation (11). Then for fixed $\mathbf{x} \in \mathbb{R}^n$, the following inequality holds

$$\max_{\sigma\in\mathbb{R}_{>0}} g\Big(\mathbf{x},\sigma,\mathbf{w}^*_{\mathbf{x}}(\xi),\xi\Big) - g\Big(\mathbf{x},\bar{\sigma}^*_{\mathbf{x}},\mathbf{w}^*_{\mathbf{x}}(\xi),\xi\Big) \geq 0$$

for all $\xi \in \Xi$. Then we can write

$$\mathbb{E}\left[g\Big(\mathbf{x},\bar{\sigma}^*_{\mathbf{x}},\mathbf{w}^*_{\mathbf{x}}(\xi),\xi\Big)\right] = \mathbb{E}\left[\max_{\sigma\in\mathbb{R}_{>0}} g\Big(\mathbf{x},\sigma,\mathbf{w}^*_{\mathbf{x}}(\xi),\xi\Big)\right] \tag{14}$$

if and only if $g\Big(\mathbf{x},\bar{\sigma}^*_{\mathbf{x}},\mathbf{w}^*_{\mathbf{x}}(\xi),\xi\Big) = \max_{\sigma\in\mathbb{R}_{>0}} g\Big(\mathbf{x},\sigma,\mathbf{w}^*_{\mathbf{x}}(\xi),\xi\Big)$ with probability 1. That means that the equality in equation (14) holds if and only if

$$g\Big(\mathbf{x},\bar{\sigma}^*_{\mathbf{x}},\mathbf{w}^*_{\mathbf{x}}(\xi),\xi\Big) = \max_{\sigma\in\mathbb{R}_{>0}} g\Big(\mathbf{x},\sigma,\mathbf{w}^*_{\mathbf{x}}(\xi),\xi\Big)$$

for a.e. $\xi \in \Xi$. This happens if $g\Big(\mathbf{x},\sigma,\mathbf{w}^*_{\mathbf{x}}(\xi),\xi\Big)$ has a maximum point $\bar{\sigma}^*_{\mathbf{x}}$ independent of $\xi \in \Xi$. □

2.4 The Inequality EEV ≤ EIIV ≤ EPIV ≤ EV

We now establish the following theorem that proves the inequality for the expected performance of evolution strategies when different forms of the function g for deriving the optimal recombination weights and step size are used.

Theorem 2.9 (EEV ≤ EIIV ≤ EPIV ≤ EV). *Consider the EV in equation* (2)*, the EEV in equation* (4)*, the EPIV in equation* (8)*, the EIIV in equation* (10) *and the function* $g : \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^\lambda \times \mathbb{R}^{n\times\lambda} \to \mathbb{R}$ *for the optimal recombination weights and step size. Then for fixed* $\mathbf{x} \in \mathbb{R}^n$*, the following inequality holds:*

$$\mathrm{EEV}_{\mathbf{x}} \leq \mathrm{EIIV}_{\mathbf{x}} \leq \mathrm{EPIV}_{\mathbf{x}} \leq \mathrm{EV}_{\mathbf{x}} \tag{15}$$

Proof. We first prove the inequality $\mathrm{EEV}_{\mathbf{x}} \leq \mathrm{EIIV}_{\mathbf{x}}$. By definition (2.2) and equation (4), we know that EEV is always less than or equal to the maximum of the expectation of the function g, i.e. the following inequality always holds:

$$\mathbb{E}\left[g\Big(\mathbf{x},\sigma^*_{\mathbf{x}},\mathbf{w}^*_{\mathbf{x}},\xi\Big)\right] \leq \max_{\sigma\in\mathbb{R}_{>0},\mathbf{w}\in\mathbb{R}^\lambda} \mathbb{E}\left[g\Big(\mathbf{x},\sigma,\mathbf{w},\xi\Big)\right] \tag{16}$$

On the other hand, we know that for any $\mathbf{w} \in \mathbb{R}^\lambda$ and $\sigma \in \mathbb{R}_{>0}$, the inequality $g\Big(\mathbf{x},\sigma,\mathbf{w},\xi\Big) \leq \max_{\mathbf{w}\in\mathbb{R}^\lambda} g\Big(\mathbf{x},\sigma,\mathbf{w},\xi\Big)$ always holds. Take the expectation on both side of the inequality and then get the maximum value respect to the step size so we can write

$$\max_{\sigma\in\mathbb{R}_{>0}} \mathbb{E}\left[g\Big(\mathbf{x},\sigma,\mathbf{w},\xi\Big)\right] \leq \max_{\sigma\in\mathbb{R}_{>0}} \mathbb{E}\left[\max_{\mathbf{w}\in\mathbb{R}^\lambda} g\Big(\mathbf{x},\sigma,\mathbf{w},\xi\Big)\right]$$

It then follows that

$$\max_{\sigma\in\mathbb{R}_{>0},\mathbf{w}\in\mathbb{R}^\lambda} \mathbb{E}\left[g\Big(\mathbf{x},\sigma,\mathbf{w},\xi\Big)\right] \leq \max_{\sigma\in\mathbb{R}_{>0}} \mathbb{E}\left[\max_{\mathbf{w}\in\mathbb{R}^\lambda} g\Big(\mathbf{x},\sigma,\mathbf{w},\xi\Big)\right] \tag{17}$$

The right hand side of the previous inequality is $\mathrm{EIIV}_{\mathbf{x}}$. Combining the inequality in equation (16) and (17). Hence the inequality $\mathrm{EEV}_{\mathbf{x}} \leq \mathrm{EIIV}_{\mathbf{x}}$ holds. Lastly by Lemma (2.5) and (2.7), we know that the inequalities $\mathrm{EPIV}_{\mathbf{x}} \leq \mathrm{EV}_{\mathbf{x}}$ and $\mathrm{EIIV}_{\mathbf{x}} \leq \mathrm{EPIV}_{\mathbf{x}}$ hold. Combining all together, the inequality in equation (15) holds. □

3 BOUNDS ON NORMALIZED QUALITY GAIN

3.1 Preliminaries

This section derives the inequality for the expected performance of evolution strategies when different forms of the normalized quality gain are used for the optimal recombination weights and step size. Quality gain [4, 7] is the expected relative improvement on the function value in an iteration of the evolution strategies and is commonly used for deriving the optimal values of the algorithms' parameters. We start by defining the normalized quality gain of evolution strategies which is taken from [4].

Definition 3.1 (Normalized Quality Gain on Spherical Function). The quality gain on spherical function for an evolution strategy with weighted recombination is a function mapping $\varphi : \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^\lambda \times \mathbb{R}^{n\times\lambda} \mapsto \mathbb{R}$ and is defined as

$$\varphi(\mathbf{x},\sigma,\mathbf{w},\mathcal{Z}) := \frac{f(\mathbf{x}) - f(\mathbf{x}+\sigma\mathcal{Z}\mathbf{w})}{f(\mathbf{x})} \tag{18}$$

where $f(\mathbf{x}) := ||\mathbf{x}-\mathbf{x}^*||^2$ is the objective function for the spherical function and $\mathbf{x}^* \in \mathbb{R}^n$ is the optimum of the objective function. The matrix $\mathcal{Z}$ is a $n\times\lambda$ random matrix in the form of $\mathcal{Z} := [\mathbf{z}^{1:\lambda} \cdots \mathbf{z}^{\lambda:\lambda}]$

defined on the probability space $(\Xi, \mathcal{B}, \mathbb{P})$ where the column vector $\mathbf{z}^{i:\lambda}$ denotes the n-dimensional random vector that is drawn from the n-dimensional standard normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$ and is used by the i-th best offspring. The *normalized quality gain* is a function mapping $\tilde{\varphi}: \mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^\lambda \times \mathbb{R}^{n\times\lambda} \mapsto \mathbb{R}$ and is defined as

$$\begin{aligned}\tilde{\varphi}(\mathbf{x}, \tilde{\sigma}, \mathbf{w}, \mathcal{Z}) &:= \frac{n}{2}\varphi\left(\mathbf{x}, \sigma = \frac{\tilde{\sigma}||\mathbf{x}-\mathbf{x}^*||}{n}, \mathbf{w}, \mathcal{Z}\right) \\ &= -\tilde{\sigma}\frac{(\mathbf{x}-\mathbf{x}^*)^T}{||\mathbf{x}-\mathbf{x}^*||}\mathcal{Z}\mathbf{w} - \frac{\tilde{\sigma}^2}{2n}\mathbf{w}^T\mathcal{Z}^T\mathcal{Z}\mathbf{w} \qquad (19)\end{aligned}$$

where $\tilde{\sigma} = \frac{\sigma \cdot n}{||\mathbf{x}-\mathbf{x}^*||}$ is the *normalized step size*.

The following proposition provides an important property of the function $\tilde{\varphi}$ that is useful when we derive different forms of the expected performance of the normalized quality gain.

Proposition 3.2. *Let $\mathbf{O} \in \mathbb{R}^{n\times n}$ be an orthogonal matrix, i.e. $\mathbf{O}^\mathrm{T}\mathbf{O} = \mathbf{O}\mathbf{O}^\mathrm{T} = \mathbf{I}$ where $\mathbf{I}$ is an identity matrix, such that for vectors $\mathbf{x}, \mathbf{x}^* \in \mathbb{R}^n$, we have $\bar{\mathbf{x}} = \mathbf{O}(\mathbf{x}-\mathbf{x}^*)$ and the vector $\bar{\mathbf{x}} \in \mathbb{R}^n$ is in the form of $(||\mathbf{x}-\mathbf{x}^*||, 0, \ldots, 0)$. Assume that for every fixed $\mathbf{x} \in \mathbb{R}^n$, $\tilde{\sigma} \in \mathbb{R}_{>0}$ and $\mathbf{w} \in \mathbb{R}^\lambda$, the function $\tilde{\varphi}(\mathbf{x}, \tilde{\sigma}, \mathbf{w}, \cdot)$ is $\mathcal{B}$-measurable, then the following equation holds almost surely:*

$$\tilde{\varphi}(\mathbf{x}, \tilde{\sigma}, \mathbf{w}, \mathcal{Z}) = -\tilde{\sigma}\mathbf{e}_1^T\mathbf{O}\mathcal{Z}\mathbf{w} - \frac{\tilde{\sigma}^2}{2n}\mathbf{w}^T\mathcal{Z}^T\mathcal{Z}\mathbf{w} \ \textit{a.s.} \qquad (20)$$

where $\mathbf{e}_1$ is a column vector $[1, 0, \ldots, 0]^T \in \mathbb{R}^n$.

Proof. We know that $\mathbf{O}$ is an orthogonal matrix so we have $||\mathbf{O}\mathbf{x}||^2 = ||\mathbf{x}||^2$ for all $\mathbf{x} \in \mathbb{R}^n$. Then the equation $||\mathbf{x}-\mathbf{x}^* + \sigma\mathcal{Z}\mathbf{w}||^2 = ||\mathbf{O}\left(\mathbf{x}-\mathbf{x}^* + \sigma\mathcal{Z}\mathbf{w}\right)||^2$ holds almost surely. Since $||\bar{\mathbf{x}}||^2$ equals $||\mathbf{x}-\mathbf{x}^*||^2$, the equation $||\mathbf{O}\bar{\mathbf{x}}||^2 = ||\mathbf{O}(\mathbf{x}-\mathbf{x}^*)||^2$ holds. Then $||\mathbf{x}-\mathbf{x}^* + \sigma\mathcal{Z}\mathbf{w}||^2 = ||\bar{\mathbf{x}} + \sigma\mathbf{O}\mathcal{Z}\mathbf{w}||^2$ holds almost surely. Now consider the quality gain in equation (18), the following equation holds almost surely

$$\varphi(\mathbf{x}, \sigma, \mathbf{w}, \mathcal{Z}) = 1 - \frac{||\mathbf{x}-\mathbf{x}^* + \sigma\mathcal{Z}\mathbf{w}||^2}{||\mathbf{x}-\mathbf{x}^*||^2} = 1 - \frac{||\bar{\mathbf{x}} + \sigma\mathbf{O}\mathcal{Z}\mathbf{w}||^2}{||\bar{\mathbf{x}}||^2}$$

The RHS of the previous equation equals to $\varphi(\bar{\mathbf{x}} + \mathbf{x}^*, \sigma, \mathbf{w}, \mathbf{O}\mathcal{Z})$. Therefore we can further derive that the equation

$$\tilde{\varphi}(\mathbf{x}, \tilde{\sigma}, \mathbf{w}, \mathcal{Z}) = \tilde{\varphi}(\bar{\mathbf{x}} + \mathbf{x}^*, \tilde{\sigma}, \mathbf{w}, \mathbf{O}\mathcal{Z})$$

holds almost surely. Replace the RHS of the previous equation with equation (19) and then substitute $\frac{\bar{\mathbf{x}}^T}{||\bar{\mathbf{x}}||} = \mathbf{e}_1$ into it. Hence the results in equation (20) holds. □

The following proposition derives the optimal value of the normalized quality gain given a fixed $\mathbf{x} \in \mathbb{R}^n$ and $\mathcal{Z} \in \mathbb{R}^{n\times\lambda}$.

Proposition 3.3. *Consider the normalized quality gain $\tilde{\varphi}$ in equation (19). Given a matrix $\mathcal{Z} \in \mathbb{R}^{n\times\lambda}$ such that the determinant of $\mathcal{Z}^T\mathcal{Z}$ is non-zero, i.e. $det(\mathcal{Z}^T\mathcal{Z}) \neq 0$. For $\mathbf{x} \in \mathbb{R}^n$, let $\upsilon_\mathbf{x}: \mathcal{Z} \mapsto \mathbb{R}$ be the function mapping for the optimal value for the normalized quality gain $\tilde{\varphi}$ as*

$$\upsilon_\mathbf{x}(\mathcal{Z}) := \tilde{\varphi}\left(\mathbf{x}, \tilde{\sigma}_\mathbf{x}^*(\mathcal{Z}), \mathbf{w}_\mathbf{x}^*(\mathcal{Z}), \mathcal{Z}\right) \qquad (21)$$

where $\mathbf{w}_\mathbf{x}^*: \mathbb{R}^{n\times\lambda} \mapsto \mathbb{R}^\lambda$ and $\tilde{\sigma}_\mathbf{x}^*: \mathbb{R}^{n\times\lambda} \mapsto \mathbb{R}_{>0}$ are the function mappings for the optimal recombination weights and the optimal step size respectively. Then for fixed $\mathbf{x} \in \mathbb{R}^n$ and $\mathcal{Z} \in \mathbb{R}^{n\times\lambda}$, the normalized quality gain $\mathbf{x}$ attains its maximum $\upsilon_\mathbf{x}(\mathcal{Z})$ and the following equations hold:

$$\upsilon_\mathbf{x}(\mathcal{Z}) = -\frac{n}{2}\mathbf{e}_1^T\mathbf{O}\mathcal{Z}(\mathcal{Z}^T\mathcal{Z})^+\mathcal{Z}^T\mathbf{O}^T\mathbf{e}_1 \qquad (22)$$

$$\mathbf{w}_\mathbf{x}^*(\mathcal{Z}) = -(\mathcal{Z}^T\mathcal{Z})^+\mathcal{Z}^T\mathbf{O}^T\mathbf{e}_1 \qquad (23)$$

$$\tilde{\sigma}_\mathbf{x}^*(\mathcal{Z}) = n \qquad (24)$$

where $(\mathcal{Z}^T\mathcal{Z})^+$ is the pseudo inverse of the $\lambda\times\lambda$ matrix $\mathcal{Z}^T\mathcal{Z}$.

Proof. We first know that for a fixed $\mathbf{x} \in \mathbb{R}^n$ and $\mathcal{Z} \in \mathbb{R}^{n\times\lambda}$ the normalized quality gain $\tilde{\varphi}$ is continuously differentiable with respect to $\mathbf{w}$ and $\tilde{\sigma}$. It is also convex over the set $\mathbf{w} \in \mathbb{R}^\lambda$ or the set $\tilde{\sigma} \in \mathbb{R}_{>0}$. Then for a fixed $\mathbf{x} \in \mathbb{R}^n$ and $\mathcal{Z} \in \mathbb{R}^{n\times\lambda}$, the optimal recombination weights $\mathbf{w}_\mathbf{x}^*(\mathcal{Z})$ and step size $\tilde{\sigma}_\mathbf{x}^*(\mathcal{Z})$ satisfy the first order necessary optimality conditions. By proposition (3.2), we know that the equation $\tilde{\varphi}(\mathbf{x}, \tilde{\sigma}, \mathbf{w}, \mathcal{Z}) = \tilde{\varphi}(\mathbf{e}_1, \tilde{\sigma}, \mathbf{w}, \mathbf{O}\mathcal{Z})$ holds almost surely. Therefore we take the partial derivatives of the normalized quality gain with respect to $\mathbf{w}$ and $\tilde{\sigma}$ in equation (20) so we can write

$$\partial_\mathbf{w}\tilde{\varphi} = -\tilde{\sigma}\mathcal{Z}^T\mathbf{O}^T\mathbf{e}_1 - \frac{\tilde{\sigma}^2}{n}\mathcal{Z}^T\mathcal{Z}\mathbf{w} \qquad (25)$$

$$\partial_{\tilde{\sigma}}\tilde{\varphi} = -\mathbf{e}_1^T\mathbf{O}\mathcal{Z}\mathbf{w} - \frac{\tilde{\sigma}}{n}\mathbf{w}^T\mathcal{Z}^T\mathcal{Z}\mathbf{w} \qquad (26)$$

Setting equation (25) to λ-dimensional zero vector $\mathbf{0}$ and equation (26) to 0, rearranging the terms and it follows that we get $\mathbf{w}_\mathbf{x}^* = -(\mathcal{Z}^T\mathcal{Z})^+\mathcal{Z}^T\mathbf{O}^T\mathbf{e}_1$ and $\tilde{\sigma}_\mathbf{x}^* = n$. Hence the equations (23) and (24) hold. Lastly, substitute these optimal recombination weights and step size to the RHS of equation (20) and calculate the optimal value. This proves equation (22). □

3.2 Normalized Quality Gain in Different Forms

This section derives different forms of the normalized quality gain. We start by deriving the EV of the normalized quality gain.

Lemma 3.4. *Consider the EV problem in Definition (2.1) and the normalized quality gain $\tilde{\varphi}$ in Definition (3.1). Then for any vector $\mathbf{x} \in \mathbb{R}^n$, the EV for the normalized quality gain $\mathrm{EV}_\mathbf{x}$ and the EV solutions for the optimal recombination weights $\mathbf{w}_\mathbf{x}^*$ and step size $\sigma_\mathbf{x}^*$ satisfy*

$$\mathrm{EV}_\mathbf{x} = -\frac{n}{2}\mathbf{p}^T\left(\mathbb{E}[\mathcal{Z}]^T\mathbb{E}[\mathcal{Z}]\right)^+\mathbf{p} \qquad (27)$$

$$\mathbf{w}_\mathbf{x}^* = -\left(\mathbb{E}[\mathcal{Z}]^T\mathbb{E}[\mathcal{Z}]\right)^+\mathbf{p} \qquad (28)$$

$$\tilde{\sigma}_\mathbf{x}^* = n \qquad (29)$$

where $\mathbb{E}[\mathcal{Z}] = \left[\mathbb{E}[\mathbf{z}^{1:\lambda}]\cdots\mathbb{E}[\mathbf{z}^{\lambda:\lambda}]\right]$ is a $n\times\lambda$ matrix and $\mathbf{p} := \left(\mathbb{E}[z_1^{1:\lambda}], \ldots, \mathbb{E}[z_1^{\lambda:\lambda}]\right)^T$ is a λ-th dimensional column vector. The notation $\mathbf{z}^{i:\lambda}$ is a n-dimensional column random vector used by the i-th best offspring and $z_1^{1:\lambda}$ is the first component of the random vector $\mathbf{z}^{1:\lambda}$. The $\lambda\times\lambda$ matrix $\left(\mathbb{E}[\mathcal{Z}]^T\mathbb{E}[\mathcal{Z}]\right)^+$ is the pseudo inverse of the $\lambda\times\lambda$ matrix $\mathbb{E}[\mathcal{Z}]^T\mathbb{E}[\mathcal{Z}]$.

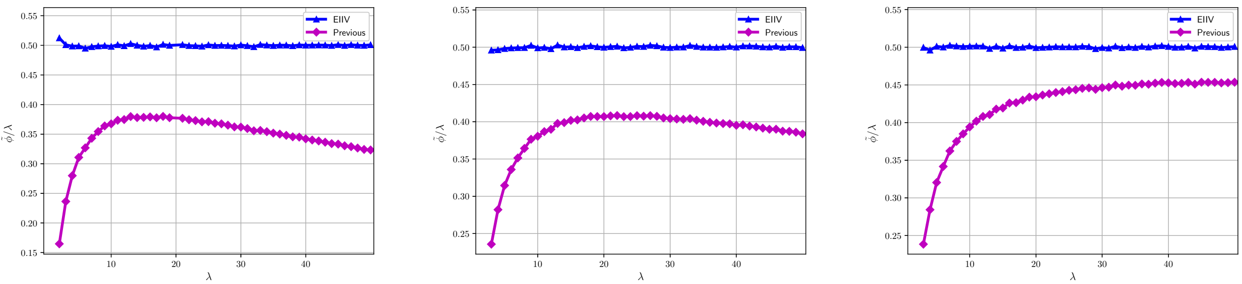

Figure 1: The serial efficiency (normalized quality gain $\tilde{\varphi}$ divided by λ) against the population size (λ) on spherical function of dimension $n = 100, 200, 1000$ (from left to right). The EIIV shows the optimal value obtained by equation (40). The bottom line in each graph uses the recombination weights in the work [2, 3].

Proof. Consider the definition of $\mathcal{Z}$ and the orthogonal matrix $\mathbf{O}$, we know that the equation $\mathbf{O}\mathbb{E}[\mathcal{Z}] = \left[\mathbf{O}\mathbb{E}[\mathbf{z}^{1:\lambda}] \cdots \mathbf{O}\mathbb{E}[\mathbf{z}^{\lambda:\lambda}]\right]$ holds. Since $\mathbf{O}$ is an orthogonal matrix so we have $\mathbf{O}\mathbb{E}[\mathbf{z}] = \mathbb{E}[\mathbf{O}\mathbf{z}]$ for any n-dimensional random vector $\mathbf{z} \in \mathbb{R}^n$ drawn from a n-dimensional standard normal distribution. Subtitute it to the previous equation so we have $\mathbf{O}\mathbb{E}[\mathcal{Z}] = \left[\mathbb{E}[\mathbf{O}\mathbf{z}^{1:\lambda}] \cdots \mathbb{E}[\mathbf{O}\mathbf{z}^{\lambda:\lambda}]\right]$. Now for any random vector $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^N$ drawn from a n-dimensional standard normal distribution, if $||\mathbf{x} - \mathbf{x}^* + \sigma\mathbf{z}_1||^2 \leq ||\mathbf{x} - \mathbf{x}^* + \sigma\mathbf{z}_2||^2$ holds, the inequality $||\bar{\mathbf{x}} + \sigma\mathbf{O}\mathbf{z}_1||^2 \leq ||\bar{\mathbf{x}} + \sigma\mathbf{O}\mathbf{z}_2||^2$ holds. This implies that for the expectation of the random vector used by the i-th best offspring, the equality $\mathbb{E}[\mathbf{O}\mathbf{z}^{i:\lambda}] = \mathbb{E}[\mathbf{z}^{i:\lambda}]$ holds for $1 \leq i \leq \lambda$. Therefore the equation $\mathbf{O}\mathbb{E}[\mathcal{Z}] = \mathbb{E}[\mathcal{Z}]$ hold.

Consider the EV in equation (2) for the normalized quality gain in equation (19). By proposition (3.2), the equation $\tilde{\varphi}(\mathbf{x}, \tilde{\sigma}, \mathbf{w}, \mathcal{Z}) = \tilde{\varphi}(\mathbf{e}_1, \tilde{\sigma}, \mathbf{w}, \mathbf{O}\mathcal{Z})$ holds almost surely. Therefore we can write

$$\max_{\sigma \in \mathbb{R}_{>0}, \mathbf{w} \in \mathbb{R}^\lambda} \tilde{\varphi}\left(\mathbf{x}, \tilde{\sigma}, \mathbf{w}, \mathbb{E}[\mathcal{Z}]\right) = \max_{\sigma \in \mathbb{R}_{>0}, \mathbf{w} \in \mathbb{R}^\lambda} \tilde{\varphi}\left(\mathbf{e}_1, \tilde{\sigma}, \mathbf{w}, \mathbf{O}\mathbb{E}[\mathcal{Z}]\right) \tag{30}$$

By Proposition (3.3), we can take differentiation to the RHS of equation (30) with respect to the recombination weights and step size so it follows from equations (23) and (24), we have

$$\mathbf{w}_{\mathbf{x}}^* = -\left(\mathbb{E}[\mathcal{Z}]^T \mathbb{E}[\mathcal{Z}]\right)^+ \mathbb{E}[\mathcal{Z}]^T \mathbf{O}^T \mathbf{e}_1 \tag{31}$$

$$\tilde{\sigma}_{\mathbf{x}}^* = n \tag{32}$$

Consider that the equations $\mathbf{O}\mathbb{E}[\mathcal{Z}] = \mathbb{E}[\mathcal{Z}]$ and $\mathbb{E}[\mathcal{Z}]^T \mathbf{O}^T \mathbf{e}_1 = \left(\mathbb{E}[z_1^{1:\lambda}], \ldots, \mathbb{E}[z_1^{\lambda:\lambda}]\right)^T$ hold. Substitute them into the previous equations and hence equations (28) and (29) hold. To prove equation (27), substitute these optimal recombination weights and step size into the RHS of equation (20) and hence the results in equation (27) hold. □

The following two lemmas derive the EPIV and the EIIV of the normalized quality gain.

Lemma 3.5. *Consider the EPIV in Definition* (2.3) *and the normalized quality gain $\tilde{\varphi}$ in Definition* (3.1). *Then for any vector $\mathbf{x} \in \mathbb{R}^n$, the EPIV for the normalized quality gain* $\text{EPIV}_{\mathbf{x}}$, *and the EPIV solutions for the optimal recombination weights $\mathbf{w}_{\mathbf{x}}^*(\mathcal{Z})$ and step size $\sigma_{\mathbf{x}}^*(\mathcal{Z})$ for all $\mathcal{Z} \in \Xi$ satisfy*

$$\text{EPIV}_{\mathbf{x}} = -\frac{n}{2}\mathbb{E}\left[\mathbf{q}^T \left(\mathcal{Z}^T \mathcal{Z}\right)^+ \mathbf{q}\right] \tag{33}$$

$$\mathbf{w}_{\mathbf{x}}^*(\mathcal{Z}) = -\left(\mathcal{Z}^T \mathcal{Z}\right)^+ \mathbf{q} \tag{34}$$

$$\tilde{\sigma}_{\mathbf{x}}^*(\mathcal{Z}) = n \tag{35}$$

where $\mathcal{Z} = \left[\mathbf{z}^{1:\lambda} \cdots \mathbf{z}^{\lambda:\lambda}\right]$ is a $n \times \lambda$ matrix and $\mathbf{q} := \left(z_1^{1:\lambda}, \ldots, z_1^{\lambda:\lambda}\right)^T$ is a λ-dimensional column vector. The notation $\mathbf{z}^{i:\lambda}$ is a n-dimensional column random vector used by the i-th best offspring and $z_1^{1:\lambda}$ is the first component of the random vector $\mathbf{z}^{1:\lambda}$. The $\lambda \times \lambda$ matrix $(\mathcal{Z}^T \mathcal{Z})^+$ is the pseudo inverse of the $\lambda \times \lambda$ matrix $\mathcal{Z}^T \mathcal{Z}$.

Proof. We first derive the EPIV solution. Consider the function $v_{\mathbf{x}}(\xi)$ for the normalized quality gain in the equation (5) in Definition (2.3). For a fixed $\mathbf{x} \in \mathbf{R}$ and each $\mathcal{Z} \in \Xi$, we have

$$v_{\mathbf{x}}(\mathcal{Z}) = \max_{\sigma \in \mathbb{R}_{>0}, \mathbf{w} \in \mathbb{R}^\lambda} \tilde{\varphi}(\mathbf{x}, \tilde{\sigma}, \mathbf{w}, \mathcal{Z}) \tag{36}$$

By proposition (3.2), we know that the equation $\tilde{\varphi}(\mathbf{x}, \tilde{\sigma}, \mathbf{w}, \mathcal{Z}) = \tilde{\varphi}(\mathbf{e}_1, \tilde{\sigma}, \mathbf{w}, \mathbf{O}\mathcal{Z})$ holds almost surely. We can rewrite equation (36) into

$$v_{\mathbf{x}}(\mathcal{Z}) = \max_{\sigma \in \mathbb{R}_{>0}, \mathbf{w} \in \mathbb{R}^\lambda} \tilde{\varphi}(\mathbf{e}_1, \tilde{\sigma}, \mathbf{w}, \mathbf{O}\mathcal{Z}) \tag{37}$$

By Proposition (3.3), we can take differentiation to the previous equation so it follows from equations (23) and (24) that for every $\mathcal{Z} \in \Xi$, the following equations hold

$$\mathbf{w}_{\mathbf{x}}^*(\mathcal{Z}) = -\left(\mathcal{Z}^T \mathcal{Z}\right)^+ \mathcal{Z}^T \mathbf{O}^T \mathbf{e}_1 \tag{38}$$

$$\tilde{\sigma}_{\mathbf{x}}^*(\mathcal{Z}) = n \tag{39}$$

Consider that the equation $\mathcal{Z}^T \mathbf{O}^T \mathbf{e}_1 = \left(z_1^{1:\lambda}, \ldots, z_1^{\lambda:\lambda}\right)$ holds. Substitute into equations (38) and (39) hence we prove the equations (34) and (35). To prove the $\text{EPIV}_{\mathbf{x}}$ in equation (33), we subtitute the optimal recombination weights and step size in equation (38) and (39) for every $\mathcal{Z} \in \Xi$ and take the expectation, then we get the $\text{EPIV}_{\mathbf{x}}$ in equation (33). □

Lemma 3.6. *Consider the EIIV in Definition* (2.4) *and the normalized quality gain $\tilde{\varphi}$ in Definition* (3.1). *Then for any vector $\mathbf{x} \in \mathbb{R}^n$, the EIIV for the normalized quality gain $EIIV_{\mathbf{x}}$, the EIIV solutions* *for the optimal recombination weights* $\mathbf{w}^*_{\mathbf{x}}(\mathcal{Z})$ *for all* $\mathcal{Z} \in \Xi$*, and the optimal step size* $\sigma^*_{\mathbf{x}}$ *satisfy*

$$\mathrm{EIIV}_{\mathbf{x}} = -\frac{n}{2}\mathbb{E}\left[\mathbf{q}^T\left(\mathcal{Z}^T\mathcal{Z}\right)^+\mathbf{q}\right] \tag{40}$$

$$\mathbf{w}^*_{\mathbf{x}}(\mathcal{Z}) = -\left(\mathcal{Z}^T\mathcal{Z}\right)^+\mathbf{q} \tag{41}$$

$$\tilde{\sigma}^*_{\mathbf{x}} = n \tag{42}$$

where $\mathcal{Z} = \left[\mathbf{z}^{1:\lambda}\cdots\mathbf{z}^{\lambda:\lambda}\right]$ *is a* $n\times\lambda$ *matrix and* $\mathbf{q} := \left(z_1^{1:\lambda},\ldots,z_1^{\lambda:\lambda}\right)^T$ *is a* λ*-dimensional column vector. The notation* $\mathbf{z}^{i:\lambda}$ *is a* n*-dimensional column random vector used by the* i*-th best offspring and* $z_1^{1:\lambda}$ *is the first component of the random vector* $\mathbf{z}^{1:\lambda}$*. The* $\lambda\times\lambda$ *matrix* $(\mathcal{Z}^T\mathcal{Z})^+$ *is the pseudo inverse of the* $\lambda\times\lambda$ *matrix* $\mathcal{Z}^T\mathcal{Z}$*.*

Proof. Consider the equation (9) in Definition (2.4). By proposition (3.2), the equation $\tilde{\varphi}(\mathbf{x},\tilde{\sigma},\mathbf{w},\mathcal{Z}) = \tilde{\varphi}(\mathbf{e}_1,\tilde{\sigma},\mathbf{w},\mathbf{O}\mathcal{Z})$ holds almost surely. We can then write

$$\alpha_{\mathbf{x}}(\mathcal{Z}) = \max_{\mathbf{w}\in\mathbb{R}^\lambda}\tilde{\varphi}\left(\mathbf{e}_1,\tilde{\sigma},\mathbf{w},\mathbf{O}\mathcal{Z}\right) \tag{43}$$

By Proposition (3.3), we take differentiation to the previous equation with respect to the recombination weights $\mathbf{w}$ so we get

$$\mathbf{w}^*_{\mathbf{x}}(\mathcal{Z}) = -\left(\mathcal{Z}^T\mathcal{Z}\right)^+\mathcal{Z}^T\mathbf{O}^T\mathbf{e}_1$$

Substitute the equation $\mathcal{Z}^T\mathbf{O}^T\mathbf{e}_1 = \left(z_1^{1:\lambda},\ldots,z_1^{\lambda:\lambda}\right)$ into the previous equation hence we obtain the result in equation (41). To obtain $\tilde{\sigma}^*_{\mathbf{x}}$ in equation (42), consider the equation (43). The RHS of the equation equals $\tilde{\varphi}\left(\mathbf{e}_1,\tilde{\sigma},\mathbf{w}^*_{\mathbf{x}}(\mathcal{Z}),\mathbf{O}\mathcal{Z}\right)$. By equation (20), for every $\mathcal{Z}\in\Xi$ we can write $\alpha_{\mathbf{x}}(\mathcal{Z})$

$$\alpha_{\mathbf{x}}(\mathcal{Z}) = -\tilde{\sigma}\mathbf{e}_1^T\mathbf{O}\mathcal{Z}\mathbf{w}^*_{\mathbf{x}}(\mathcal{Z}) - \frac{\tilde{\sigma}^2}{2n}\mathbf{w}^*_{\mathbf{x}}(\mathcal{Z})^T\mathcal{Z}^T\mathcal{Z}\mathbf{w}^*_{\mathbf{x}}(\mathcal{Z})$$

Taking expectation on both side of the previous equation yields

$$\mathbb{E}\left[\alpha_{\mathbf{x}}(\mathcal{Z})\right] = -\tilde{\sigma}\mathbb{E}\left[\mathbf{q}^T\mathbf{w}^*_{\mathbf{x}}(\mathcal{Z})\right] - \frac{\tilde{\sigma}^2}{2n}\mathbb{E}\left[\mathbf{w}^*_{\mathbf{x}}(\mathcal{Z})^T\mathcal{Z}^T\mathcal{Z}\mathbf{w}^*_{\mathbf{x}}(\mathcal{Z})\right] \tag{44}$$

Differentiate the previous equation with respect to $\tilde{\sigma}$, set the resulting equation to 0 and rearrange the terms, we get the result in equation (42). To prove the $\mathrm{EIIV}_{\mathbf{x}}$ in equation (40), we subtitute the optimal recombination weights and step size into equation (44) and hence the result in equation (40) hold. □

Lastly we derive the EEV of the normalized quality gain.

Lemma 3.7. *Consider the EEV in Definition* (2.2) *and the normalized quality gain* $\tilde{\varphi}$ *in Definition* (3.1)*. Then for any vector* $\mathbf{x}\in\mathbb{R}^n$*, the EEV for the normalized quality gain* $\mathrm{EEV}_{\mathbf{x}}$ *satisfies*

$$\mathrm{EEV}_{\mathbf{x}} = -n\mathbb{E}\left[\mathbf{q}^T\mathbf{w}^*_{\mathbf{x}}\right] - \frac{n}{2}\mathbb{E}\left[\mathbf{w}^{*T}_{\mathbf{x}}\mathcal{Z}^T\mathcal{Z}\mathbf{w}^*_{\mathbf{x}}\right] \tag{45}$$

where $\mathcal{Z} = \left[\mathbf{z}^{1:\lambda}\cdots\mathbf{z}^{\lambda:\lambda}\right]$ *is a* $n\times\lambda$ *matrix and* $\mathbf{q} := \left(z_1^{1:\lambda},\ldots,z_1^{\lambda:\lambda}\right)^T$ *is a* λ*-dimensional column vector. The notation* $\mathbf{z}^{i:\lambda}$ *is a* n*-dimensional column random vector used by the* i*-th best offspring and* $z_1^{1:\lambda}$ *is the first component of the random vector* $\mathbf{z}^{1:\lambda}$*. The optimal recombination weights* $\mathbf{w}^*_{\mathbf{x}}$ *is defined in equation* (28)*.*

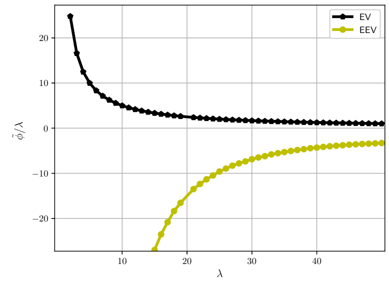

Figure 2: The comparison of the serial efficiencies for EV and EEV against the population size (λ) on spherical function of dimension $n = 100$. The EV and EEV show the optimal values obtained by equation (27) and (45) respectively.

Proof. Consider $\mathrm{EEV}_{\mathbf{x}}$ in equation (4). By proposition (3.2), the equation $\tilde{\varphi}(\mathbf{x},\tilde{\sigma},\mathbf{w},\mathcal{Z}) = \tilde{\varphi}(\mathbf{e}_1,\tilde{\sigma},\mathbf{w},\mathbf{O}\mathcal{Z})$ holds almost surely. We can then write

$$\mathrm{EEV}_{\mathbf{x}} = \mathbb{E}\left[\tilde{\varphi}(\mathbf{e}_1,\tilde{\sigma}^*_{\mathbf{x}},\mathbf{w}^*_{\mathbf{x}},\mathbf{O}\mathcal{Z})\right]$$

By equation (20), we can write the RHS of the previous equation as

$$\begin{aligned}&\mathbb{E}\left[\tilde{\varphi}(\mathbf{e}_1,\tilde{\sigma}^*_{\mathbf{x}},\mathbf{w}^*_{\mathbf{x}},\mathbf{O}\mathcal{Z})\right]\\&= -n\mathbb{E}\left[\mathbf{e}_1^T\mathbf{O}\mathcal{Z}\mathbf{w}^*_{\mathbf{x}}\right] - \frac{n}{2}\mathbb{E}\left[\mathbf{w}^{*T}_{\mathbf{x}}\mathcal{Z}^T\mathcal{Z}\mathbf{w}^*_{\mathbf{x}}\right]\end{aligned} \tag{46}$$

Substitute the equation $\mathcal{Z}^T\mathbf{O}^T\mathbf{e}_1 = \left(z_1^{1:\lambda},\ldots,z_1^{\lambda:\lambda}\right)$ into the previous equation hence we obtain the result in equation (45). □

The following theorem establishes the inequality for evolution strategies when different forms of the normalized quality gain are used.

Theorem 3.8. *Consider the EV problem in Definition* (2.1)*, the EEV in Definition* (2.2)*, the EPIV in Definition* (2.3) *and the EIIV in Definition* (2.4)*. Then the following inequality holds for the normalized quality gain* $\tilde{\varphi}$ *in Definition* (3.1)

$$\mathrm{EEV}_{\mathbf{x}} \le \mathrm{EIIV}_{\mathbf{x}} \le \mathrm{EPIX}_{\mathbf{x}} \le \mathrm{EV}_{\mathbf{x}} \tag{47}$$

Proof. It follows easily from Theorem (2.9) and hence the inequality in equation (47) holds. □

4 SIMULATION

We now compare the expected performance of the normalized quality gain in different forms. As the normalized quality gain is expressed in terms of the expectations of random variables, we simulate the normalized quality gain in finite dimensions by means of Monte-Carlo method. For every normalized quality gain expression, we simulate 10^6 times for an expectation of a random variable in order to estimate the normalized quality gain for different normalized step size $\tilde{\sigma}$ and different population sizes.

Figure (1) shows the serial efficiency (the normalized quality gain divided by the population size λ) of the EIIV for the normalized quality gain. The optimal values for EIIV are taken from the equation (40). We also simulate the normalized quality gain in the

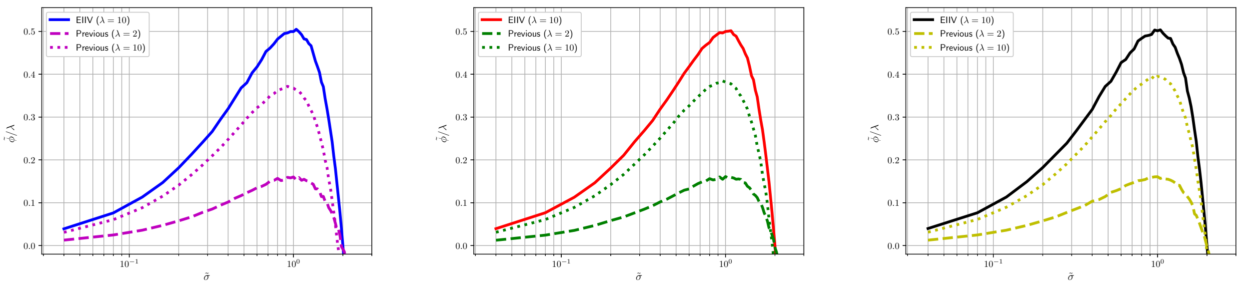

Figure 3: The serial efficiency against normalized step size on spherical function of dimension $n = 10, 100, 1000$ (from left to right). The population sizes λ are 2 and 10. The optimal recombination weights in equation (41) are used for EIIV. The bottom two lines in each graph use the recombination weights in the work [2, 3].

previous work in [2, 3]. It is seen that the EIIV has the better serial efficiency than that in the previous works. The EIIV always achieves its best serial efficiency for all population sizes . Better serial efficiencies can be observed only on large population sizes in the previous work.

Figure (2) illustrates the serial efficiency of the EV and the EEV for the normalized quality gain. The optimal values for the EV and EEV are taken from the equations (27) and (45) respectively. We can see that the EV has the best serial efficiency. However, when its optimal recombination weights and step size are used, the serial efficiency degrades in the EEV and in all cases we observe a negative serial efficiency.

The graphs in Figure (3) show the serial efficiency against the normalized step size over a range of 0.01 to 3. All graphs show that in EIIV, using the realization of random variables to compute the recombination weights has a better serial efficiency than that of using the expectations of random variables in the previous work.

5 CONCLUSION

This paper performs an analysis on the evolution strategies with the optimal recombination weights and step size on spherical functions in finite dimensions. We first discuss the different forms of function for deriving the optimal recombination weights and step size, and then we derive an inequality for general functions as well as the normalized quality gain. The results provide us a useful insight into the algorithmic behavior. Firstly, we prove that in the EV problem, using the expectations of random variables to derive the optimal recombination weights and step size provides us the best expected performance of evolution strategies. However, this only happens when certain conditions are met, and the expected performance of using the solutions obtained from the EV problem can be disappointing. Secondly, we prove that in the EPIV and EIIV problems, using the realizations of random variables in the evolution strategies to derive the optimal recombination weights is better than using expectations of random variables. This suggests that a fitness-based weight schema is better than a rank-based weight schema. Thirdly, the inequality suggests us that when we derive the optimal parameters in the evolution strategies, the choice of the function for deriving the optimal parameters plays a significant part. Different forms of functions for optimal parameters can yield a different result.

Although the theorems theoretically prove that the evolution strategy achieves a better performance when it uses realizations of random variable to derive its recombination weights, this has become a question how this can be applied in practice. Additionally, extending this work to investigate the spherical functions in infinite dimensions or other quadratic functions is of our interest.

REFERENCES

[1] Youhei Akimoto, Anne Auger, and Nikolaus Hansen. 2017. Quality gain analysis of the weighted recombination evolution strategy on general convex quadratic functions. In *Proceedings of the 14th ACM/SIGEVO Conference on Foundations of Genetic Algorithms*. ACM, 111–126.
[2] Dirk V Arnold. 2005. Optimal weighted recombination. In *FOGA*. Springer, 215–237.
[3] Dirk V Arnold. 2006. Weighted multirecombination evolution strategies. *Theoretical computer science* 361, 1 (2006), 18–37.
[4] Hans-Georg Beyer. 2013. *The theory of evolution strategies*. Springer Science & Business Media.
[5] Hans-Georg Beyer and Dirk V Arnold. 2003. Qualms regarding the optimality of cumulative path length control in CSA/CMA-evolution strategies. *Evolutionary Computation* 11, 1 (2003), 19–28.
[6] Mohamed Jebalia and Anne Auger. 2010. Log-linear Convergence of the Scale-invariant (μ/μ w, λ)-ES and Optimal μ for Intermediate Recombination for Large Population Sizes. *Parallel Problem Solving from Nature, PPSN XI* (2010), 52–62.
[7] I. Rechenberg. 1973. *Evolutionsstrategie : Optimierung technischer Systeme nach Prinzipien der biologischen Evolution.* Number 15 in Problemata. Frommann-Holzboog, Stuttgart-Bad Cannstatt.