



HAL
open science

Scaling-invariant functions versus positively homogeneous functions

Cheikh Touré, Armand Gissler, Anne Auger, Nikolaus Hansen

► **To cite this version:**

Cheikh Touré, Armand Gissler, Anne Auger, Nikolaus Hansen. Scaling-invariant functions versus positively homogeneous functions. *Journal of Optimization Theory and Applications*, 2021. hal-03104436v2

HAL Id: hal-03104436

<https://hal.inria.fr/hal-03104436v2>

Submitted on 7 Sep 2021

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.

Scaling-invariant Functions versus Positively Homogeneous Functions

Cheikh Toure · Armand Gissler · Anne Auger · Nikolaus Hansen

Received: date / Accepted: date

Abstract Scaling-invariant functions preserve the order of points when the points are scaled by the same positive scalar (usually with respect to a unique reference point). Composites of strictly monotonic functions with positively homogeneous functions are scaling-invariant with respect to zero. We prove in this paper that also the reverse is true for large classes of scaling-invariant functions. Specifically, we give necessary and sufficient conditions for scaling-invariant functions to be composites of a strictly monotonic function with a positively homogeneous function. We also study sublevel sets of scaling-invariant functions generalizing well-known properties of positively homogeneous functions.

Communicated by Juan-Enrique Martinez Legaz.

Keywords scaling-invariant function · positively homogeneous function · compact level set.

Mathematics Subject Classification (2000) 49J52 · 54C35

1 Introduction

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is scaling-invariant (SI) with respect to a reference point $x^* \in \mathbb{R}^n$ if for all $x, y \in \mathbb{R}^n$ and $\rho > 0$:

$$f(x^* + x) \leq f(x^* + y) \iff f(x^* + \rho x) \leq f(x^* + \rho y) , \quad (1)$$

that is, the f -order of any two points is invariant under a multiplicative change of their distance to the reference point—the order only depends on their direction and their *relative* distance to the reference. Scaling-invariant functions

Inria and CMAP, Ecole Polytechnique, IP Paris, France
firstname.lastname@inria.fr
cheikh.toure@polytechnique.edu

appear naturally when studying the convergence of comparison-based optimization algorithms where the update of the state of the algorithm is using f only through comparisons of candidate solutions [3, 8]. A famous example of a comparison-based optimization algorithm is the Nelder-Mead method [16].

A function $p : \mathbb{R}^n \rightarrow \mathbb{R}$ is positively homogeneous (PH) with degree $\alpha > 0$ (PH_α) if for all $x \in \mathbb{R}^n$ and $\rho > 0$:

$$p(\rho x) = \rho^\alpha p(x) . \quad (2)$$

Positively homogeneous functions are scaling-invariant with respect to $x^* = 0$. We also consider that $x \mapsto p(x - x^*)$ is positively homogeneous w.r.t. x^* when p is positively homogeneous. Linear functions, norms, and convex quadratic functions are positively homogeneous. We can define PH functions piecewise on cones or half-lines, because a function is PH_α if and only if (2) is satisfied within each cone or half-line (which is not the case with SI functions where x and y in (1) can belong to different cones). For example, the function $p : \mathbb{R}^n \rightarrow \mathbb{R}$ defined as $p(x) = x_1$ if $x_1 x_2 > 0$ and $p(x) = 0$ otherwise, is PH_1 . Positively homogeneous functions and in particular increasing positively homogeneous functions are well-studied in the context of Monotonic Analysis [7, 17, 18] or nonsmooth analysis and nonsmooth optimization [9]. Specifically, non-linear programming problems where the objective function and constraints are positively homogeneous are analyzed in [13] whereas saddle representations of continuous positively homogeneous functions by linear functions are established in [10]. The (left) composition of a PH function with a strictly monotonic function is SI while this composite function is in general not PH. One of the questions we investigate in this paper is to which extent SI functions and composites of PH functions with strictly monotonic functions are the same. We prove that a *continuous* SI function is always the composite of a strictly monotonic function with a PH function. We give necessary and sufficient conditions for an SI function to be the composite of a strictly monotonic function with a PH function in the general case.

Only level sets or sublevel sets matter to determine the difficulty of an SI problem optimized with a comparison-based algorithm. We investigate different properties of level sets thereby generalizing properties that are known for PH functions, including a formulation of the Euler homogenous function theorem that holds for PH functions.

Notation

We denote \mathbb{R}_+ the interval $[0, +\infty)$, $\mathbb{R}_- = (-\infty, 0]$, \mathbb{Z} the set of all integers, \mathbb{Z}_+ the set of all non-negative integers and \mathbb{Q} the set of rational numbers. The Euclidean norm is denoted by $\|\cdot\|$. For $x \in \mathbb{R}^n$ and $\rho > 0$, we denote by $\mathcal{B}(x, \rho) = \{y \in \mathbb{R}^n; \|x - y\| < \rho\}$ the open ball centered at x and of radius ρ , $\overline{\mathcal{B}}(x, \rho)$ its closure and $\mathcal{S}(x, \rho)$ its boundary. When they are centered at 0, we denote $\mathcal{B}_\rho = \mathcal{B}(0, \rho)$, $\overline{\mathcal{B}}_\rho = \overline{\mathcal{B}}(0, \rho)$ and $\mathcal{S}_\rho = \mathcal{S}(0, \rho)$. We refer to a proper interval containing more than a single element as *nontrivial* interval. For a

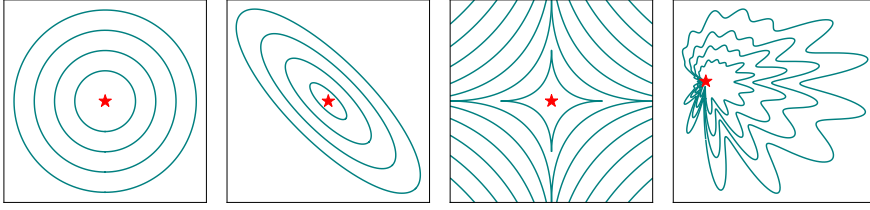


Fig. 1: Level sets of SI functions with respect to the red star x^* . The four functions are strictly increasing transformations of $x \mapsto p(x - x^*)$ where p is a PH function. From left to right: $p(x) = \|x\|$; $p(x) = x^\top A x$ for A symmetric positive and definite; $p(x) = \left(\sum_i \sqrt{|x_i|}\right)^2$ the $\frac{1}{2}$ -norm; a randomly generated SI function from a “smoothly” randomly perturbed sphere function. The two first functions from the left have convex sublevel sets, contrary to the last two.

nontrivial interval $I \subset \mathbb{R}$ and a function $\varphi : I \rightarrow \mathbb{R}$, we use the terminology of strictly increasing (respectively strictly decreasing) if for all $a, b \in I$ with $a < b$, $\varphi(a) < \varphi(b)$ (respectively $\varphi(a) > \varphi(b)$). For a real number ρ and a subset $A \subset \mathbb{R}^n$, we define $\rho A = \{\rho x; x \in A\}$. For a function f , we denote by $\text{Im}(f)$ the image of f .

2 Preliminaries

Given a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $x \in \mathbb{R}^n$, we denote the level set going through x as $\mathcal{L}_{f,x} = \{y \in \mathbb{R}^n, f(y) = f(x)\}$ and the sublevel set as $\mathcal{L}_{f,x}^{\leq} = \{y \in \mathbb{R}^n, f(y) \leq f(x)\}$.

If f is SI with respect to x^* , then the function $x \mapsto f(x + x^*) - f(x^*)$ is scaling invariant with respect to 0. Hence, if a function f is SI, we assume in the following that f is SI with respect to the reference point 0 and that $f(0) = 0$, without loss of generality.

We can immediately imply from (1) that if x and y belong to the same level set, then ρx and ρy belong to the same level set. Hence the level set of x and ρx are scaled from one another, i.e., $\mathcal{L}_{f,\rho x} = \rho \mathcal{L}_{f,x}$.

Similarly, since for any $x, y \in \mathbb{R}^n$ and $\rho > 0$, $f(\frac{y}{\rho}) \leq f(x)$ if and only if $f(y) \leq f(\rho x)$,

$$\mathcal{L}_{f,\rho x}^{\leq} = \rho \mathcal{L}_{f,x}^{\leq} \text{ and } \mathcal{L}_{f,\rho x} = \rho \mathcal{L}_{f,x} . \quad (3)$$

These properties are visualized in Figure 1.

Given an SI function f , we define surjective restrictions of f to half-lines along a vector $x \in \mathbb{R}^n$ as

$$f_x : t \in [0, \infty) \mapsto f(tx) . \quad (4)$$

It is immediate to see that the f_x are also SI¹. However, f may not be SI even when all f_x are².

Scaling invariant functions have at most one isolated local optimum [3] where an isolated local optimum, say, an isolated argmin, x , for a function $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined in that there exists $\epsilon > 0$ such that for all $y \in \mathcal{B}(x, \epsilon) \setminus \{x\}$, $g(y) > g(x)$. This result is reminded in the following proposition.

Proposition 2.1 (see [3, Proposition 3.2]) *Let f be an SI function. Then f can admit an isolated local optimum only in $f(0) = 0$ and this local optimum is also the global optimum. In addition, the functions f_x cannot admit a local plateau, i.e., a ball where the function is locally constant, unless the function is equal to 0 everywhere.*

We characterize in the following the functions f_x of an SI function f under different conditions.

Proposition 2.2 *If f is a continuous SI function on \mathbb{R}^n , then for all $x \in \mathbb{R}^n$, f_x is either constant equal to 0 or strictly monotonic.*

More specifically, if $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$ is a 1-dimensional continuous SI function, then φ is either constant equal to 0 or strictly monotonic.

Proof Assume that φ is not strictly monotonic on $[0, \infty)$. Then φ is not strictly monotonic on $(0, \infty)$. Since continuous injective functions are strictly monotonic, φ is not injective on $(0, \infty)$. Therefore there exists $0 < s < t$ such that $\varphi(s) = \varphi(t)$. By scaling-invariance, it follows that $\varphi(\frac{s}{t}) = \varphi(1)$. It follows iteratively that for all integer $k > 0$, $\varphi((\frac{s}{t})^k) = \varphi(1)$. Taking the limit for $k \rightarrow \infty$, we obtain that $\varphi(0) = \varphi(1)$. Thereby by scaling-invariance again, it follows for all $\rho > 0$ that $\varphi(0) = \varphi(\rho)$. Hence we have shown that if φ is not strictly monotonic, it is a constant function.

Now if f is a continuous SI function on \mathbb{R}^n and $x \in \mathbb{R}^n$, then f_x is also scaling invariant and continuous. Then it follows that f_x is either constant or strictly monotonic. \square

We deduce from Proposition 2.2 the next corollary.

Corollary 2.1 *Let f be a continuous SI function. If f has a local optimum at x , then for all $t \geq 0$, $f(tx) = f(0)$. In particular, if f has a global argmin (resp. argmax), then 0 is a global argmin (resp. argmax).*

Proof Assume that there exists a local optimum at x . Then f_x has a local optimum at 1. Therefore f_x is not strictly monotonic, and thanks to Proposition 2.2, f_x is necessarily a constant function. In other words, $f(tx) = f(0)$ for all $t \geq 0$. \square

¹ This directly follows because for $s, t \in \mathbb{R}_+$ and $\rho > 0$, $f_x(t) \leq f_x(s) \iff f(tx) \leq f(sx) \iff f(\rho tx) \leq f(\rho sx) \iff f_x(\rho t) \leq f_x(\rho s)$.

² For example, define $f : \mathbb{R} \rightarrow \mathbb{R}$ as $t \mapsto t$ on \mathbb{R}_+ and $t \mapsto t^2$ on \mathbb{R}_- . Then $f_1(t) = t$ and $f_{-1}(t) = t^2$, for $t \in \mathbb{R}_+$, are both SI and even PH with degree 1 and 2, respectively. But f is not SI, and hence also not PH, because $f(\frac{1}{2}) = \frac{1}{2} > \frac{1}{4} = f(-\frac{1}{2})$ but $f(4 \times \frac{1}{2}) = 2 < 4 = f(4 \times (-\frac{1}{2}))$.

We derive another proposition with the same conclusions as Proposition 2.2 but under a different assumption. We start by showing the following lemma.

Lemma 2.1 *Let $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$ be an SI function continuous at 0 and strictly monotonic on a nontrivial interval $I \subset \mathbb{R}_+$, then φ is strictly monotonic.*

Proof Assume without loss of generality that φ is strictly increasing on I and that $I = (a, b)$ with $0 < a < b$, up to replacing I with a subset of I . Denote $\rho = \frac{b}{a}$. Then $\{\rho^k, \rho^{k+1}\}_{k \in \mathbb{Z}}$ covers $(0, \infty)$. To prove that φ is strictly increasing on $(0, \infty)$, it is enough to prove that φ is strictly increasing on $[\rho^k, \rho^{k+1}]$ for all integer k .

Let k be an integer and (x, y) two real numbers such that $\rho^k \leq x < y \leq \rho^{k+1}$. Then $a \leq \frac{ax}{\rho^k} < \frac{ay}{\rho^k} \leq a\rho = b$. Therefore $\varphi(\frac{ax}{\rho^k}) < \varphi(\frac{ay}{\rho^k})$. And by scaling-invariance, $\varphi(x) < \varphi(y)$.

With the continuity at 0, it follows that φ is strictly increasing on \mathbb{R}_+ . \square

We derive from Lemma 2.1 the following proposition.

Proposition 2.3 *Let f be an SI function continuous at 0. Assume that each f_x is on some nontrivial interval either strictly monotonic or constant. Then for all $x \in \mathbb{R}^n$, f_x is either constant equal to 0 or strictly monotonic.*

Note that the continuity of the function f_x alone does not suffice to conclude that f_x is either constant or strictly monotonic on some nontrivial interval. Indeed there exist 1-D continuous functions (even differentiable functions) that are not monotonic on any nontrivial interval [4, 6, 11].

For the sake of completeness, we construct SI functions in \mathbb{R}_+ that are not monotonic on any nontrivial interval. The construction of such functions is based on the nonlinear solutions of the Cauchy functional equation: for all $x, y \in \mathbb{R}$, $g(x+y) = g(x) + g(y)$, called Hamel functions [12]. A Hamel function f also satisfies $f(q_1x + q_2y) = q_1f(x) + q_2f(y)$ for all real numbers x, y and rational numbers q_1, q_2 [1, Chapter 2]. Since g is nonlinear, there exist real numbers x and y such that the vectors $\{(x, g(x)), (y, g(y))\}$ form a basis of \mathbb{R}^2 over the field \mathbb{R} . Then the graph of g , which is a vector subspace of \mathbb{R}^2 over the field \mathbb{Q} , contains $\{q_1 \cdot (x, g(x)) + q_2 \cdot (y, g(y)); (q_1, q_2) \in \mathbb{Q}^2\}$ which is dense in \mathbb{R}^2 . Therefore a 1-D Hamel function is highly pathological, since its graph is dense in \mathbb{R}^2 .

Lemma 2.2 *There exist SI functions on \mathbb{R}_+ that are neither monotonic nor continuous on any nontrivial interval.*

Proof We start by choosing a nonlinear solution of the Cauchy's functional equation denoted by $g : \mathbb{R} \rightarrow \mathbb{R}$, knowing that there are uncountably many ways to pick such a g [12]. Then for all real numbers a and b , $g(a+b) = g(a) + g(b)$. And since g is not linear, we also know that g is neither continuous nor monotonic on any nontrivial interval [12]. Let us define $f = \exp \circ g \circ \log$ on $(0, \infty)$ and $f(0) = 0$. Then $f(x) > 0$ for all $x > 0$ and f is still not monotonic on any nontrivial interval. We also have for all $\rho > 0$ and $x > 0$, $f(\rho x) = \exp(g(\log(x) + \log(\rho))) = \exp(g(\log(x))) \exp(g(\log(\rho))) = f(x)f(\rho)$. This last result gives the scaling-invariance property. \square

Based on Lemma 2.2, we derive the next proposition.

Proposition 2.4 *There exist SI functions f on \mathbb{R}^n such that for all non-zero x , f_x is neither monotonic nor continuous on any nontrivial interval.*

Proof Based on Lemma 2.2, there exists $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$ SI on \mathbb{R}_+ which is neither monotonic nor continuous on any nontrivial interval. We construct f as follows. For all $x \in \mathbb{R}^n$, $f(x) = \varphi(\|x\|)$. Then f is SI because for $x, y \in \mathbb{R}^n$ and for $\rho > 0$, $f(x) \leq f(y) \iff \varphi(\|x\|) \leq \varphi(\|y\|) \iff \varphi(\rho\|x\|) \leq \varphi(\rho\|y\|) \iff f(\rho x) \leq f(\rho y)$. In addition for a non-zero x and $t \geq 0$, $f_x(t) = f(tx) = \varphi(t\|x\|)$ and then f_x is neither monotonic nor continuous on any nontrivial interval. \square

Now assume that f is a continuous scaling invariant function and we can write $f = \varphi \circ g$ where φ is a continuous bijection and g is a positively homogeneous function. As a direct consequence of the bijection theorem given in Appendix A, φ^{-1} is continuous if φ is a continuous bijection defined on an interval. Therefore $g = \varphi^{-1} \circ f$ is also continuous. This result is stated in the following corollary.

Corollary 2.2 *Let f be a continuous SI function, φ a continuous bijection defined on an interval in \mathbb{R} and p a positively homogeneous function such that $f = \varphi \circ p$. Then p is also continuous.*

3 Scaling-invariant Functions as Composite of Strictly Monotonic Functions with Positively Homogeneous Functions

As underlined in the introduction, compositions of strictly monotonic functions with positively homogeneous functions are scaling-invariant (SI) functions. We investigate in this section under which conditions the converse is true, that is, when SI functions are compositions of strictly monotonic functions with PH functions. Section 3.1 shows that continuity is a sufficient condition, whereas Section 3.2 gives some necessary *and* sufficient condition on f to be decomposable in this way.

3.1 Continuous SI Functions

We prove in this section a main result of the paper: any continuous SI function f can be written as $f = \varphi \circ p$ where p is PH_1 and φ is a homeomorphism (and in particular strictly monotonically increasing and continuous). The proof relies on the following proposition where we do not assume yet that f is continuous but only the restrictions of f to the half-lines originating in 0, the f_x functions.

Proposition 3.1 *Let f be an SI function such that for any $x \in \mathbb{R}^n$, f_x as defined in (4) is continuous and strictly monotonic or constant. Then for all $\alpha > 0$, there exist a PH_α function p and a strictly increasing, continuous bijection (thus a homeomorphism) φ such that $f = \varphi \circ p$. For a non-zero f and $\alpha > 0$, the choice of (φ, p) is unique up to a left composition of p with a piece-wise linear function.*

- (i) In addition, if all non-constant f_x have the same monotonicity for all $x \in \mathbb{R}^n$, then for any $x_0 \in \mathbb{R}^n$ such that $f(x_0) \neq 0$, the homeomorphism φ corresponding to a PH_1 function can be chosen as f_{x_0} and is at least as smooth as f .
- (ii) Otherwise, there exist $x_1, x_{-1} \in \mathbb{R}^n$ such that f_{x_1} is strictly increasing and $f_{x_{-1}}$ is strictly decreasing. And for any such x_1 and x_{-1} we can choose as homeomorphism φ corresponding to a PH_1 function the function f_{x_1} on \mathbb{R}_+ and $t \mapsto f_{x_{-1}}(-t)$ on \mathbb{R}_- .

Proof Let f be an SI function such that for any $x \in \mathbb{R}^n$, f_x is either a constant or a strictly monotonic continuous function.

In the case where all the f_x are constant for all $x \in \mathbb{R}^n$, then $f = 0$ and therefore we can take $p_\alpha = 0$ as a candidate for a continuous PH_α and $\varphi_\alpha : t \mapsto t$ as the candidate for the corresponding homeomorphism.

From now on, at least one of the $\{f_x\}_{x \in \mathbb{R}^n}$ is non-constant. We now split the proof in two parts, the case where all the non-constant f_x have the same monotonicity and the case where there exist $x_1, x_{-1} \in \mathbb{R}^n$ such that f_{x_1} is strictly increasing and $f_{x_{-1}}$ is strictly decreasing.

Part 1. Assume here that all the non-constant f_x have the same monotonicity for all $x \in \mathbb{R}^n$. And up to a transformation $x \mapsto -f(x)$, we can assume without loss of generality that they are increasing. Therefore 0 is a global argmin and since we have assumed $f(0) = 0 : f(x) \geq 0$ for all $x \in \mathbb{R}^n$. Then there exists $x_0 \in \mathbb{R}^n$ such that $f(x_0) > 0$.

For any $x \in \mathcal{L}_{f,x_0} = \{y \in \mathbb{R}^n, f(y) = f(x_0)\}$, and any $\lambda > 0$ different than 1, $\lambda x \notin \mathcal{L}_{f,x_0}$. Indeed, as $x \in \mathcal{L}_{f,x_0}$, we know from Proposition 2.2 that f_x is strictly increasing on \mathbb{R}_+ , since f_x cannot be constant equal to 0.

Moreover, for all $x \in \mathbb{R}^n$ such that $f(x) \neq 0$, there exists $\lambda > 0$ such that $\lambda x \in \mathcal{L}_{f,x_0}$. Indeed, if $f(x) < f(x_0)$, the intermediate value theorem applied to the continuous function f_{x_0} shows that there exists $0 < t < 1$ such that $f(tx_0) = f_{x_0}(t) = f(x)$, and then $f(\frac{1}{t}x) = f(x_0)$. And by interchanging x and x_0 , the same argument holds if $f(x) > f(x_0)$.

The two last paragraphs ensure that for all x such that $f(x) \neq 0$, there exists a unique positive number λ_x such that $\lambda_x x \in \mathcal{L}_{f,x_0}$. Let us define the function p for all $x \in \mathbb{R}^n$ as follows: if $f(x) \neq 0$ then $p(x) = \frac{1}{\lambda_x}$, otherwise $p(x) = 0$. We prove in the following that p is PH_1 .

Let $x \in \mathbb{R}^n$ and $\rho > 0$. If $f(x) = 0$ (hence $f(\rho x) = 0$), then $p(\rho x) = 0 = \rho p(x)$. Otherwise $f(x) > 0$ (hence $f(\rho x) > 0$), and $p(\rho x) = \frac{\rho}{\lambda_x}$ since $\frac{\lambda_x}{\rho}$ is the (unique) positive number such that $\frac{\lambda_x}{\rho} \rho x = \lambda_x x \in \mathcal{L}_{f,x_0}$. And thereby $p(\rho x) = \rho p(x)$.

We prove that $f = f_{x_0} \circ p$, where f_{x_0} is a continuous strictly increasing function and p is PH_1 . Let $x \in \mathbb{R}^n$. If $f(x) = 0$, then $p(x) = 0$, and then $f(x) = 0 = f(0) = f_{x_0}(0) = (f_{x_0} \circ p)(x)$. Otherwise, we have by construction that $\frac{x}{p(x)} \in \mathcal{L}_{f,x_0}$. Therefore $f(\frac{x}{p(x)}) = f(x_0)$ and then $f(x) = f(p(x)x_0) = f_{x_0}(p(x))$. By Theorem A.1, $\varphi = f_{x_0}$ is a homeomorphism. Let $\alpha > 0$, define $\tilde{\varphi} = t \mapsto \varphi(t^{1/\alpha})$ and $\tilde{p} = p^\alpha$. Then \tilde{p} is PH_α , $\tilde{\varphi}$ is a homeomorphism and $f = \tilde{\varphi} \circ \tilde{p}$.

Assume that we have two couples of solutions (φ, p) and $(\bar{\varphi}, \bar{p})$ such that $f = \varphi \circ p = \bar{\varphi} \circ \bar{p}$ where $\varphi, \bar{\varphi}$ are homeomorphisms and p, \bar{p} are PH_α . For all $t > 0$ and $x \in \mathbb{R}^n$, we have for instance $p(tx) = t^\alpha p(x)$. Therefore $\text{Im}(p) = \mathbb{R}_+$. Denote $\psi = \bar{\varphi}^{-1} \circ \varphi$. For all $\lambda > 0$ and $x \in \mathbb{R}^n$, $\psi(\lambda^\alpha p(x)) = \psi(p(\lambda x)) = \bar{p}(\lambda x) = \lambda^\alpha \psi(p(x))$. Hence ψ is PH_1 on \mathbb{R}_+ . For all $t > 0$, $\psi(t) = t\psi(1)$. Therefore ψ is linear.

Part 2. Assume now that there exist $x_1, x_{-1} \in \mathbb{R}^n$ such that f_{x_1} is strictly increasing and $f_{x_{-1}}$ is strictly decreasing. Then $f(x_1) > 0$ and $f(x_{-1}) < 0$. Then thanks to the intermediate value theorem, if $f(x) > 0$, there exists a unique positive number λ_x such that $\lambda_x x \in \mathcal{L}_{f, x_1}$, and if $f(x) < 0$, there exists a unique positive number λ_x such that $\lambda_x x \in \mathcal{L}_{f, x_{-1}}$. We define now p for all $x \in \mathbb{R}^n$ as follows: if $f(x) = 0$ then $p(x) = 0$, if $f(x) > 0$ then $p(x) = \frac{1}{\lambda_x}$, and finally if $f(x) < 0$ then $p(x) = -\frac{1}{\lambda_x}$. Let us show that p is PH_1 . Indeed for any $\rho > 0$ and $x \in \mathbb{R}^n$, if $f(x) = 0$ (hence $f(\rho x) = 0$), then $p(\rho x) = 0 = \rho p(x)$. If $f(x) > 0$ (hence $f(\rho x) > 0$), then $p(\rho x) = \frac{\rho}{\lambda_x} = \rho p(x)$ since $\frac{\lambda_x}{\rho}$ is the (unique) positive number such that $\frac{\lambda_x}{\rho} \rho x = \lambda_x x \in \mathcal{L}_{f, x_1}$. And finally if $f(x) < 0$ (hence $f(\rho x) < 0$), then $p(\rho x) = -\frac{\rho}{\lambda_x} = \rho p(x)$ since $\frac{\lambda_x}{\rho}$ is the (unique) positive number such that $\frac{\lambda_x}{\rho} \rho x = \lambda_x x \in \mathcal{L}_{f, x_{-1}}$. Hence p is PH_1 .

We define now the function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ such that if $t \geq 0$, $\varphi(t) = f_{x_1}(t)$ and if $t \leq 0$, $\varphi(t) = f_{x_{-1}}(-t)$. Then, φ is well defined ($f_{x_1}(0) = 0 = f_{x_{-1}}(0)$), continuous and strictly increasing.

Let $x \in \mathbb{R}^n$. If $f(x) = 0$, then $p(x) = 0$, and then $f(x) = 0 = (\varphi \circ p)(x)$. If $f(x) > 0$, $\varphi(p(x)) = f_{x_1}(p(x)) = f(p(x)x_1) = f(x)$ since $\frac{x}{p(x)} \in \mathcal{L}_{f, x_1}$. And finally if $f(x) < 0$, $\varphi(p(x)) = f_{x_{-1}}(-p(x)) = f(-p(x)x_{-1}) = f(x)$ since $-\frac{x}{p(x)} = \lambda_x x \in \mathcal{L}_{f, x_{-1}}$. Thereby, $f = \varphi \circ p$. Theorem A.1 ensures that φ is a homeomorphism. By defining for all $\alpha > 0$, $\tilde{\varphi}(t) = \varphi(t^{1/\alpha})$ if $t \geq 0$, $\tilde{\varphi}(t) = \varphi(-(-t)^{1/\alpha})$ if $t < 0$, $\tilde{p}(x) = p(x)^\alpha$ if $p(x) \geq 0$ and $\tilde{p}(x) = -(-p(x))^\alpha$ if $p(x) < 0$, it follows that $f = \tilde{\varphi} \circ \tilde{p}$.

Assume here again that we have two couples of solutions (φ, p) and $(\bar{\varphi}, \bar{p})$ such that $f = \varphi \circ p = \bar{\varphi} \circ \bar{p}$ where $\varphi, \bar{\varphi}$ are homeomorphisms and p, \bar{p} are PH_α . For all $t > 0$ and $x \in \mathbb{R}^n$, we have $p(tx) = t^\alpha p(x)$. Therefore $\text{Im}(p) = \mathbb{R}$ since $p(x_1)$ and $p(x_2)$ have opposite signs. Denote $\psi = \bar{\varphi}^{-1} \circ \varphi$. For all $\lambda > 0$ and $x \in \mathbb{R}^n$, $\psi(\lambda^\alpha p(x)) = \psi(p(\lambda x)) = \bar{p}(\lambda x) = \lambda^\alpha \psi(p(x))$. Hence ψ is PH_1 on \mathbb{R} . For all $t > 0$, $\psi(t) = t\psi(1)$ and $\psi(-t) = t\psi(-1)$. Therefore depending on the values of $\psi(1)$ and $\psi(-1)$, ψ is either linear or piece-wise linear. \square

We now use the previous proposition to prove that a continuous SI function is a homeomorphic transformation of a continuous PH_1 function. The proof relies on the result that for a continuous SI function, the f_x are either constant or strictly monotonic and continuous (see Proposition 2.2). We distinguish the case where f has a global optimum as in Proposition 3.1 (i) and the case where f does not have a global optimum as in Proposition 3.1 (ii). Overall the following result holds.

Theorem 3.1 *Let f be a continuous SI function. Then for all $\alpha > 0$, there exists a continuous PH_α function p and a strictly increasing and continuous bijection (thus a homeomorphism) φ such that $f = \varphi \circ p$.*

For a non-zero f and $\alpha > 0$, the choice of (φ, p) is unique up to a left composition of p with a piece-wise linear function. If f admits a global optimum, then 0 is also a global optimum. For any $x_0 \in \mathbb{R}^n$ such that $f(x_0) \neq 0$, in the case where p is a PH_1 function, the homeomorphism φ can be chosen as f_{x_0} and is at least as smooth as f .

If f does not admit a global optimum, then there exist $x_1, x_{-1} \in \mathbb{R}^n$ such that $f(x_1) > 0$ and $f(x_{-1}) < 0$. For any such x_1 and x_{-1} , in the case where p is a PH_1 function, the homeomorphism φ can be chosen as the function equal to f_{x_1} on \mathbb{R}_+ and equal to $t \mapsto f_{x_{-1}}(-t)$ on \mathbb{R}_- .

Proof Let f be a continuous SI function. Thanks to Proposition 2.2, for all $x \in \mathbb{R}^n$, f_x is either constant equal to 0 or strictly monotonic.

Part 1. Assume that f has a global optimum. Corollary 2.1 shows that 0 is also a global optimum. Then we can apply Proposition 3.1 in the case where the non-constant f_x have the same monotonicity. Let $x_0 \in \mathbb{R}^n$ such that $f(x_0) \neq 0$ and define $\varphi = f_{x_0}$. Then $f = \varphi \circ p$ and φ is a homeomorphism. That settles the continuity of the PH_1 function $\varphi^{-1} \circ f$ thanks to Corollary 2.2.

Part 2. Assume in this part that f has no global optimum. Since 0 is not a global optimum, we can find x_1 and x_{-1} such that $f(x_1) > 0$ and $f(x_{-1}) < 0$. Therefore f_{x_1} is strictly increasing and $f_{x_{-1}}$ is strictly decreasing. We apply Proposition 3.1 in the case where the non-constant f_x do not have the same monotonicity. If φ is the function equal to f_{x_1} on \mathbb{R}_+ and to $t \mapsto f_{x_{-1}}(-t)$ on \mathbb{R}_- , then $f = \varphi \circ p$ where φ is a homeomorphism. That settles the continuity of the PH_1 function $\varphi^{-1} \circ f$ thanks to Corollary 2.2. For all $\alpha > 0$, the unique construction of (φ, p) up to a piece-wise linear function in both parts is a consequence of Proposition 3.1. \square

3.2 Sufficient and Necessary Condition to be the Composite of a PH Function

We have seen in the previous section that a *continuous* SI function can be written as $\varphi \circ p$ with φ strictly monotonic and p PH. Relaxing continuity, we prove in the next theorem some necessary *and* sufficient condition under which an SI function is the composite of a PH function with a strictly monotonic function.

Theorem 3.2 *Let f be an SI function. There exist a PH_1 function p and a strictly increasing function φ such that $f = \varphi \circ p$ if and only if for all $x \in \mathbb{R}^n$, the function f_x is either constant or strictly monotonic and the strictly increasing f_x share the same image (i.e., if $\lambda \in \mathbb{R}$ is reached for one of these functions, then it is reached for all of them) and the strictly decreasing ones too.*

For a non-zero f , up to a left composition of p with a piece-wise linear function, the choice of (φ, p) is unique.

Proof We prove first the forward implication. Suppose there is a PH_1 function p and a strictly monotonic function φ such that $f = \varphi \circ p$. Consider $x \in \mathbb{R}^n$. Either $p(x) = 0$ and then for any $t \geq 0$ we have that $p(tx) = 0$, so that $f_x(t) = f(tx) = \varphi(p(tx)) = \varphi(0)$ and f_x is constant on \mathbb{R}_+ , or $p(x) \neq 0$, and then $t \in \mathbb{R}_+ \mapsto p(tx) = tp(x)$ is strictly monotonic, and $f_x(t) = \varphi(p(tx))$ is strictly monotonic too on \mathbb{R}_+ . Moreover, consider $x_1 \neq x_2$ such that f_{x_1} and f_{x_2} are increasing. Then $p(x_1)$ and $p(x_2)$ are of the same sign, so there is some $t_* > 0$ such that $p(x_1) = t_*p(x_2) = p(t_*x_2)$, so the functions $t \mapsto f(tx_1)$ and $t \mapsto f(tt_*x_2)$ are equal, so the functions f_{x_1} and f_{x_2} take the same values. The same applies on the strictly decreasing functions.

We now prove the backward implication. Suppose that the functions f_x are either constant or strictly monotonic and the increasing ones share the same values and the decreasing ones too.

If all the f_x are constant, then for all $x \in \mathbb{R}^n$, $f(x) = f(0) = 0$ and it is enough to write $f = \varphi \circ p$ with $p = t \mapsto t$ on \mathbb{R}_+ and $p = 0$. We assume from now on that at least one f_x is not constant.

Consider that all the non-constant f_x have the same monotonicity. Let us choose x_0 such that $f(x_0) \neq 0$. Then for all $x \neq 0$, f_x and f_{x_0} have the same monotonicity. Since they have the same image and are injective, there exists a unique $\lambda_x > 0$ such that $\lambda_x x \in \mathcal{L}_{f, x_0}$. We then define p and φ as in the Part 1 of Theorem 3.1 to ensure that $f = \varphi \circ p$ where p is PH_1 and φ is strictly monotonic.

Consider finally that all the non-constant f_x do not have the same monotonicity. Then there exist x_1 and x_{-1} such that $f(x_1) > 0$ and $f(x_{-1}) < 0$. Then, thanks to the assumption that all increasing f_x share the same values and the strictly decreasing f_x too, if $f(x) > 0$, then there exists a unique positive number λ_x such that $\lambda_x x \in \mathcal{L}_{f, x_1} = \{y \in \mathbb{R}^n, f(y) = f(x_1)\}$, and if $f(x) < 0$, then there exists a unique (thanks to the assumption of strict monotonicity for the non-constant f_x) positive number λ_x such that $\lambda_x x \in \mathcal{L}_{f, x_{-1}}$. Therefore, we can define p as in Theorem 3.1. As before, p is PH_1 . Define also the function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ as in Theorem 3.1. It is still increasing, but not necessarily continuous. Then, as in Theorem 3.1, $f = \varphi \circ p$.

The proof of the unicity of (φ, p) up to a piece-wise real linear function is similar to the proof in Proposition 3.1. \square

Complementing Theorem 3.2, we construct an example of an SI function that can not be decomposed as $f = \varphi \circ p$, because the strictly increasing f_x do not share the same image. Define f such that for all $x \in \mathbb{R}^n$, $f(x) = \tanh(x_1)$ if the first coordinate $x_1 \geq 0$ and $f(x) = 1 + \exp(-x_1)$ otherwise. Then f is SI and if $x_1 \neq 0$, f_x is strictly increasing. However for all x such that $x_1 > 0$ then $\text{Im}(f_x) = [0, 1)$ and otherwise for x such that $x_1 \leq 0$, $\text{Im}(f_x) = \{0\} \cup (2, \infty)$.

4 Level Sets of SI Functions

Scaling-invariant functions appear naturally when studying the convergence of comparison-based optimization algorithms [3]. In this specific context, the

difficulty of a problem is entirely determined by its level sets whose properties are studied in this section.

4.1 Identical Sublevel Sets

Level sets and sublevel sets of a function f remain unchanged if we compose the function with a strictly increasing function φ since

$$f(x) \leq f(y) \iff \varphi(f(x)) \leq \varphi(f(y)) . \quad (5)$$

We prove in the next theorem that two arbitrary functions f and p have the same level sets if and only if $f = \varphi \circ p$ where φ is strictly increasing.

Theorem 4.1 *Two functions f and p have the same sublevel sets if and only if there exists a strictly increasing function φ such that $f = \varphi \circ p$.*

Proof If $f = \varphi \circ p$ with φ strictly increasing, since sublevel sets are invariant by φ , f and p have the same sublevel sets. Now assume that f and p have the same sublevel sets. Then for all $x \in \mathbb{R}^n$, there exists $T(x) \in \mathbb{R}^n$ such that $\mathcal{L}_{f,x}^{\leq} = \mathcal{L}_{p,T(x)}^{\leq}$. In other words, for all $y \in \mathbb{R}^n$, $f(y) \leq f(x) \iff p(y) \leq p(T(x))$. We define the function

$$\phi : \begin{cases} \text{Im}(f) \longrightarrow \text{Im}(p) \\ f(x) \mapsto p(T(x)) \end{cases} .$$

The function ϕ is well-defined because for $x, y \in \mathbb{R}^n$ such that $f(x) = f(y)$, $\mathcal{L}_{f,x}^{\leq} = \mathcal{L}_{f,y}^{\leq}$. And since $\mathcal{L}_{f,x}^{\leq} = \mathcal{L}_{p,T(x)}^{\leq}$ and $\mathcal{L}_{f,y}^{\leq} = \mathcal{L}_{p,T(y)}^{\leq}$, then $\mathcal{L}_{p,T(x)}^{\leq} = \mathcal{L}_{p,T(y)}^{\leq}$, and then $p(T(x)) = p(T(y))$. Therefore $\phi(f(x)) = \phi(f(y))$. By construction we have that $\phi \circ f = p \circ T$.

Let us show that $p \circ T = p$. We have $\mathcal{L}_{p \circ T, x}^{\leq} = \mathcal{L}_{f, x}^{\leq} = \mathcal{L}_{p, T(x)}^{\leq}$. Then $x \in \mathcal{L}_{p, T(x)}^{\leq}$ and then $p(x) \leq p(T(x))$. Therefore $p \leq p \circ T$. In addition for all $y \in \mathbb{R}^n$, there exists x such that $\mathcal{L}_{p, y}^{\leq} = \mathcal{L}_{f, x}^{\leq} = \mathcal{L}_{p, T(x)}^{\leq} = \mathcal{L}_{p \circ T, x}^{\leq}$. Then $y \in \mathcal{L}_{p \circ T, x}^{\leq}$, which induces that $p(T(y)) \leq p(T(x))$. Plus, $\mathcal{L}_{p, y}^{\leq} = \mathcal{L}_{p, T(x)}^{\leq}$, therefore $p(T(x)) = p(y)$. Thereby $p(T(y)) \leq p(y)$, and then $p \circ T \leq p$. Finally $p \circ T = p$. Hence $\phi \circ f = p$.

Let us prove now that ϕ is strictly increasing. Consider $x, y \in \mathbb{R}^n$ such that $f(x) < f(y)$. Then $\mathcal{L}_{f, x}^{\leq} \subset \mathcal{L}_{f, y}^{\leq}$ with a strict inclusion, which means that $\mathcal{L}_{p, T(x)}^{\leq} \subset \mathcal{L}_{p, T(y)}^{\leq}$ with a strict inclusion. Thereby $p(T(x)) < p(T(y))$, i.e., $\phi(f(x)) < \phi(f(y))$. Hence ϕ is strictly increasing. And up to restricting ϕ to its image, we can assume without loss of generality that ϕ is a strictly increasing bijection. We finally denote $\varphi = \phi^{-1}$ and it follows that $f = \varphi \circ p$. \square

Theorem 4.1 and Theorem 3.2 give both equivalence conditions for an SI function f to be equal to $\varphi \circ p$ where φ is strictly increasing and p is positively homogeneous³. One condition is that there exists a PH function with the same sublevel sets as f , while the other condition is that the f_x are either constant or strictly monotonic, and the strictly increasing and decreasing ones have the same image, respectively.

4.2 Compactness of the Sublevel Sets

Compactness of sublevel sets is relevant for analyzing step-size adaptive randomized search algorithms [2, 14]. We investigate here how compactness properties shown for positively homogeneous functions extend to scaling-invariant functions. For an SI function f , we have $\mathcal{L}_{f,tx}^{\leq} = t\mathcal{L}_{f,x}^{\leq}$. When $\psi : y \mapsto ty$ is a homeomorphism, we have that $\psi(\mathcal{L}_{f,x}^{\leq})$ equals $t\mathcal{L}_{f,x}^{\leq}$ and is compact if and only if $\mathcal{L}_{f,x}^{\leq}$ is compact. Therefore, for all $t > 0$:

$$\mathcal{L}_{f,tx}^{\leq} \text{ is compact if and only if } \mathcal{L}_{f,x}^{\leq} \text{ is compact.} \quad (6)$$

Furthermore, if p is a lower semi-continuous positively homogeneous function such that $p(x) > 0$ for all nonzero x then the sublevel sets of p are compact [2, Lemma 2.7]. We recall it with all the details in the following proposition:

Proposition 4.1 (Lemma 2.7 in [2]) *Let p be a positively homogeneous function with degree $\alpha > 0$ and $p(x) > 0$ for all $x \neq 0$ (or equivalently, 0 is the unique global argmin of p) and $p(x)$ finite for every $x \in \mathbb{R}^n$. Then for every $x \in \mathbb{R}^n$, the following holds:*

- (i) $\lim_{t \rightarrow 0} p(tx) = 0$ and for all $x \neq 0$ the function $p_x : t \in [0, \infty) \mapsto p(tx) \in \mathbb{R}_+$ is continuous, strictly increasing and converges to ∞ when t goes to ∞ .
- (ii) If p is lower semi-continuous, the sublevel set $\mathcal{L}_{p,x}^{\leq}$ is compact.

We prove a similar theorem for lower semi-continuous SI functions f with continuous f_x functions, showing in particular that the unicity of the global argmin is equivalent to items similar to the above. Note that we need to assume the continuity of the functions f_x , while this property is unconditionally satisfied for positively homogeneous functions where $p_x(t) = t^\alpha p_x(1)$ for all $x \in \mathbb{R}^n$ and for all $t > 0$.

Theorem 4.2 *Let f be SI. Then the conditions*

- $f(x) > 0$ for all $x \neq 0$ and
- 0 is the unique global argmin

³ Note that in Theorem 3.2, we can assume without loss of generality that φ is always strictly increasing by replacing if needed φ by $t \rightarrow \varphi(-t)$ and p by $-p$.

are equivalent. Let f be additionally lower semi-continuous and for all $x \in \mathbb{R}^n$, f_x is continuous on a neighborhood of 0. Then the following are equivalent:

- (i) 0 is the unique global argmin.
- (ii) For any $x \in \mathbb{R}^n \setminus \{0\}$, the function f_x is strictly increasing.
- (iii) The sublevel sets $\mathcal{L}_{\bar{f},x}^{\leq}$ for all x are compact.

Proof Since, w.l.o.g., f is given such that $f(0) = 0$, its unique global argmin is 0 if and only if $f(x) > 0$ for all $x \neq 0$. Now we prove first that (i) \Rightarrow (ii): Let $x \in \mathbb{R}^n \setminus \{0\}$. Assume (by contraposition) that there exists $0 < t_1 < t_2$ such that $f_x(t_1) = f_x(t_2)$. Then by scaling-invariance, $f_x(1) = f_x\left(\frac{t_1}{t_2}\right)$. It follows by multiplying iteratively by $\frac{t_1}{t_2}$ that for all $k \in \mathbb{Z}_+$, $f_x(1) = f_x\left(\left(\frac{t_1}{t_2}\right)^k\right)$. Therefore if we take the limit when $k \rightarrow \infty$, it follows thanks to the continuity of f_x at 0 that $f(x) = f_x(1) = f(0)$, which contradicts the assumption (i). Hence, f_x is an injective function. Plus, there exists $\epsilon > 0$ such that f_x is continuous on $[0, \epsilon]$. Therefore f_x is an injective continuous function on $[0, \epsilon]$, which implies that f_x is a strictly monotonic function on $[0, \epsilon]$. Lemma 2.1 implies therefore that f_x is strictly monotonic. And since 0 is an argmin of f_x , then f_x is strictly increasing.

(ii) \Rightarrow (iii): f is lower semi-continuous on the compact \mathcal{S}_1 , then it reaches its minimum on that compact: there exists $s \in \mathcal{S}_1$ such that $f(s) = \min_{z \in \mathcal{S}_1} f(z)$.

Also, since sublevel sets of lower semi-continuous functions are closed, then $\mathcal{L}_{\bar{f},s}^{\leq}$ is closed. Now let us show that it is also bounded.

If $y \in \mathcal{L}_{\bar{f},s}^{\leq} \setminus \{0\}$, then $f(y) \leq f(s) \leq f\left(\frac{y}{\|y\|}\right)$. And since f_y is strictly increasing, we obtain that $1 \leq \frac{1}{\|y\|}$, thereby $\|y\| \leq 1$. We have shown that $\mathcal{L}_{\bar{f},s}^{\leq} \subset \bar{\mathcal{B}}_1$.

Then $\mathcal{L}_{\bar{f},s}^{\leq}$ is a compact set, as it is a closed and bounded subset of \mathbb{R}^n . By (6), it follows that $\mathcal{L}_{\bar{f},ts}^{\leq}$ is compact for all $t > 0$.

For all $x \in \mathbb{R}^n \setminus \{0\}$, $f(x) > f(0)$ thanks to (ii). Then $\mathcal{L}_{\bar{f},0}^{\leq} = \{0\}$ and is compact.

Let $x \in \mathbb{R}^n \setminus \{0\}$. Then there exists $\epsilon > 0$ such that $f_{\frac{x}{\|x\|}}$ is continuous on $[0, \epsilon]$. We have that $f_{\frac{x}{\|x\|}}(0) = 0 < f(s) \leq f_{\frac{x}{\|x\|}}(1)$, then by scaling-invariance, $f_{\frac{x}{\|x\|}}(0) < f(\epsilon s) \leq f_{\frac{x}{\|x\|}}(\epsilon)$. Therefore by the intermediate value theorem applied to $f_{\frac{x}{\|x\|}}$ continuous on $[0, \epsilon]$, there exists $t \in (0, \epsilon]$ such that $f_{\frac{x}{\|x\|}}(t) = f(\epsilon s)$. Then $\mathcal{L}_{\bar{f},t\frac{x}{\|x\|}}^{\leq} = \mathcal{L}_{\bar{f},\epsilon s}^{\leq}$ and is compact. We apply again (6) to observe that $\mathcal{L}_{\bar{f},x}^{\leq}$ is compact.

(iii) \Rightarrow (i): Let $x \in \mathcal{L}_{\bar{f},0}^{\leq}$, then $tx \in \mathcal{L}_{\bar{f},0}^{\leq}$ for all $t \geq 0$, and then $\{tx\}_{t \in \mathbb{R}_+} \subset \mathcal{L}_{\bar{f},0}^{\leq}$ which is a compact set. This is only possible if $x = 0$, otherwise the set $\{tx\}_{t \in \mathbb{R}_+}$ would not be bounded. Hence $\mathcal{L}_{\bar{f},0}^{\leq} = \{0\}$ which implies that 0 is the unique global argmin. \square

We derive from Theorem 4.2 the next corollary, stating for a lower semi-continuous SI function with a unique global argmin, when the intersection of any half-line of origin 0 and a level set is a singleton.

Corollary 4.1 *Let f be a lower semi-continuous SI function with 0 as unique global argmin. Assume that for all $x \in \mathbb{R}^n$, f_x is continuous on a neighborhood of 0, and the f_x share the same image. Then for all $x \in \mathbb{R}^n$, any half-line of origin 0 intersects $\mathcal{L}_{f,x}$ at a unique point.*

Proof For all non-zero x , Theorem 4.2 ensures that f_x is strictly increasing. Therefore for a non-zero x , f_x is injective. And then the intersection of a level set and a half-line of origin 0 contains at most one point. In addition the f_x share the same image for all non-zero x . Then for two non-zero vectors x, y , there exists $t \geq 0$ such that $f_y(t) = f_x(1)$. In other words, there exists $t \geq 0$ such that $ty \in \mathcal{L}_{f,x}$. We end this proof by noticing that $\mathcal{L}_{f,0} = \{0\}$ and then intersects any half-line of origin 0 only at 0. \square

4.3 Sufficient Condition for Lebesgue Negligible Level Sets

We assume that f is lower semi-continuous SI admitting a unique global argmin and all f_x are continuous and prove that f has Lebesgue negligible level sets.

Proposition 4.2 *Let f be an SI function with 0 as unique global argmin. Assume that f is lower semi-continuous and for all $x \in \mathbb{R}^n$, f_x is continuous. Then the level sets of f are Lebesgue negligible.*

Proof Let $x \in \mathbb{R}^n$. Let us denote by μ the Lebesgue measure. For all $t > 0$, $\mu(\mathcal{L}_{f,tx}) = \mu(t\mathcal{L}_{f,x}) = t^n \mu(\mathcal{L}_{f,x})$, thanks to (3). Therefore, if $t \geq 1$, $\mu(\mathcal{L}_{f,tx}) \geq \mu(\mathcal{L}_{f,x})$. In addition, for all $k \geq 1$, $\mathcal{L}_{f,(1+\frac{1}{k})x} \subset \{y \in \mathbb{R}^n, f(x) \leq f(y) \leq f(2x)\} \subset \mathcal{L}_{f,2x}^{\leq}$ because if $x \neq 0$, f_x is strictly increasing thanks to Theorem 4.2. And the same theorem induces that $\mathcal{L}_{f,x}^{\leq}$ is compact and hence $\mu(\mathcal{L}_{f,x}^{\leq}) < \infty$.

It follows that $\sum_{k=1}^{\infty} \mu(\mathcal{L}_{f,x}) \leq \sum_{k=1}^{\infty} \mu(\mathcal{L}_{f,(1+\frac{1}{k})x}) \leq \mu(\mathcal{L}_{f,2x}^{\leq}) < \infty$. Hence, $\mu(\mathcal{L}_{f,x}) = 0$. \square

4.4 Balls Containing and Balls Contained in Sublevel Sets

The sublevel sets of continuous PH functions include and are embedded in balls whose construction is scaling-invariant. Given that continuous SI functions are monotonic transformation of PH functions, those properties are naturally transferred to SI functions. This is what we formalize in this section.

From the definition of a PH function with degree α , for all $x \neq 0$ we have $p(x) = \|x\|^\alpha p(x/\|x\|)$ for all $x \neq 0$. Therefore, p is continuous on $\mathbb{R}^n \setminus \{0\}$ if and only if p is continuous on \mathcal{S}_1 . For such p , we denote $m_p = \min_{x \in \mathcal{S}_1} p(x)$, and

$M_p = \max_{x \in \mathcal{S}_1} p(x)$. We have the following propositions:

Proposition 4.3 (Lemma 2.8 in [2]) *Let p be a PH function with degree α such that $p(x) > 0$ for all $x \neq 0$. Assume that p is continuous on \mathcal{S}_1 , then for all $x \neq 0$, the following holds*

$$\|x\|m_p^{1/\alpha} \leq p(x)^{1/\alpha} \leq \|x\|M_p^{1/\alpha} . \quad (7)$$

Proposition 4.4 (Lemma 2.9 in [2]) *Let p be a PH function with degree α such that $g(x) > 0$ for all $x \neq 0$. Assume that p is continuous on \mathcal{S}_1 . Then for all $\rho > 0$, the ball centered in 0 and of radius ρ is included in the sublevel set of degree $\rho^\alpha M_p$, i.e., $\mathcal{B}(0, \rho) \subset \mathcal{L}_{p, \rho x_{M_p}}^\leq$, with $p(x_{M_p}) = M_p$. For all $x \neq 0$, the sublevel set of degree $p(x)$ is included into the ball centered in 0 and of radius $(p(x)/m_p)^\alpha$, i.e.,*

$$\mathcal{L}_{p,x}^\leq \subset \mathcal{B}\left(0, \left(\frac{p(x)}{m_p}\right)^\alpha\right).$$

We can generalize both propositions to continuous scaling-invariant functions using Theorem 3.1.

Proposition 4.5 *Let f be a continuous SI function such that $f(x) > 0$ for all $x \neq 0$. Then there exist an increasing homeomorphism φ on \mathbb{R}_+ and two positive numbers $0 < m \leq M$ such that*

- (i) *for all $x \neq 0$, $\varphi(m\|x\|) \leq f(x) \leq \varphi(M\|x\|)$;*
- (ii) *for all $\rho > 0$, the ball centered in 0 and of radius ρ is included in the sublevel set of degree $\varphi(\rho\varphi^{-1}(M))$, i.e., $\mathcal{B}(0, \rho) \subset \mathcal{L}_{f, \rho x_M}^\leq$ with $f(\rho x_M) = \varphi(\rho\varphi^{-1}(M))$;*
- (iii) *for all $x \neq 0$, the sublevel set of degree $f(x)$ is included into the ball centered in 0 and of radius $\frac{\varphi^{-1}(f(x))}{\varphi^{-1}(m)}$, i.e.,*

$$\mathcal{L}_{f,x}^\leq \subset \mathcal{B}\left(0, \frac{\varphi^{-1}(f(x))}{\varphi^{-1}(m)}\right) . \quad (8)$$

Proof Thanks to Theorem 3.1, we can write $f = f_{x_0} \circ p$ where $x_0 \neq 0$, p is PH₁ and φ defined as f_{x_0} is an increasing homeomorphism. Then $p = \varphi^{-1} \circ f$ and hence verifies: for all $x \neq 0$, $p(x) > 0$. Define $m = \varphi(m_p)$ and $M = \varphi(M_p)$, where $m_p = \min_{x \in \mathcal{S}_1} p(x)$ and $M_p = \max_{x \in \mathcal{S}_1} p(x)$.

For $x \neq 0$, Proposition 4.3 ensures that $m_p\|x\| \leq p(x) \leq M_p\|x\|$. Taking the image of this equation with respect to φ proves (i).

For all $\rho > 0$, $\mathcal{B}(0, \rho) \subset \mathcal{L}_{p, \rho x_{M_p}}^\leq$, with $p(x_{M_p}) = M_p$. Since sublevel sets are invariant with respect to an increasing bijection, it follows that $\mathcal{L}_{p, \rho x_{M_p}}^\leq = \mathcal{L}_{f, \rho x_{M_p}}^\leq$. In addition, $f(\rho x_{M_p}) = \varphi(p(\rho x_{M_p})) = \varphi(\rho p(x_{M_p})) = \varphi(\rho\varphi^{-1}(M))$ such that we have proven (ii).

Let $x \neq 0$. Again by invariance of the sublevel set, $\mathcal{L}_{p,x}^{\leq} = \mathcal{L}_{f,x}^{\leq}$. And Proposition 4.4 says that $\mathcal{L}_{p,x}^{\leq} \subset \mathcal{B}\left(0, \frac{p(x)}{m_p}\right)$. We obtain the results with the facts that $p = \varphi^{-1} \circ f$ and $m_p = \varphi^{-1}(m)$. \square

4.5 A Generalization of a Weak Formulation of Euler's Homogeneous Function Theorem

For a function $p : \mathbb{R}^n \rightarrow \mathbb{R}$ continuously differentiable on $\mathbb{R}^n \setminus \{0\}$, Euler's homogeneous function theorem states that there is equivalence between p is PH with degree α and for all $x \neq 0$

$$\alpha p(x) = \nabla p(x) \cdot x . \quad (9)$$

If in addition p is continuously differentiable in zero, then $\alpha p(0) = 0 = \nabla p(0) \cdot 0$. Along with (9), this latter equation implies that at each point y of a level set $\mathcal{L}_{p,x}$, the scalar product between $\nabla p(y)$ and y is constant equal to $\nabla p(x) \cdot x$ or that the level sets of p and of the function $x \mapsto \nabla p(x) \cdot x$ are the same, that is, the level sets of a continuously differentiable PH function satisfy

$$\mathcal{L}_{p,x} = \mathcal{L}_{z \mapsto \nabla p(z) \cdot z, x} = \{y \in \mathbb{R}^n, \nabla p(y) \cdot y = \nabla p(x) \cdot x\} . \quad (10)$$

We call this a weak formulation of Euler's homogeneous function theorem.

If f is a continuous SI function, we can write f as $\varphi \circ p$ where p is PH and φ is a homeomorphism, according to Theorem 3.1. We have the following proposition in the case where φ and p are also continuously differentiable.

Proposition 4.6 *Let f be a continuously differentiable SI function that can be written as $\varphi \circ p$ where p is PH_α , φ is a homeomorphism, and φ and φ^{-1} are continuously differentiable (and thus p is continuously differentiable). Then for all $x \in \mathbb{R}^n$,*

$$\nabla f(x) \cdot x = \alpha \varphi'(p(x)) p(x) . \quad (11)$$

Proof Since $p = \varphi^{-1} \circ f$, it is continuously differentiable. From the chain rule, for all $x \in \mathbb{R}^n : \nabla f(x) \cdot x = \varphi'(p(x)) \nabla p(x) \cdot x = \alpha \varphi'(p(x)) p(x)$. The last equality results from the Euler's homogeneous theorem applied to p . \square

Yet, the assumptions of the previous proposition are not necessarily satisfied when f is a continuously differentiable SI function. Indeed, we exhibit in the next proposition an example of an SI and continuously differentiable function f such that $f = \varphi \circ p$ but either p or φ is non-differentiable.

Proposition 4.7 *Define $\varphi : t \mapsto \int_0^t \frac{1}{1 + \log^2(u)} du$ on \mathbb{R}_+ and $p : x \mapsto |x_1|$. Then $f = \varphi \circ p$ is continuously differentiable and SI. Yet, for any $\tilde{\varphi}$ strictly increasing and \tilde{p} PH such that $f = \tilde{\varphi} \circ \tilde{p}$ (including φ and p above), either \tilde{p} is not differentiable on any point of the set $\{x; x_1 = 0\}$ or $\tilde{\varphi}$ is not differentiable at 0.*

Proof Let us prove that f is continuously differentiable. For $x \neq 0$, $\nabla f(x) = \frac{1}{1+\log^2(|x_1|)} \frac{x_1}{|x_1|} e_1$ where e_1 is the unit vector $(1, 0, \dots, 0)$. Then $\lim_{x \rightarrow 0} \nabla f(x)$ exists and is equal to 0, hence f is continuously differentiable.

Assume that $(\tilde{\varphi}, \tilde{p})$ is such that $\varphi \circ p = \tilde{\varphi} \circ \tilde{p}$, with $\tilde{\varphi}$ strictly increasing and \tilde{p} PH_α . Denote $\psi = \tilde{\varphi}^{-1} \circ \varphi$. For all $\lambda > 0$ and $x \in \mathbb{R}^n$, $\psi(\lambda p(x)) = \psi(p(\lambda x)) = \tilde{p}(\lambda x) = \lambda^\alpha \psi(p(x))$. Therefore ψ is PH_α on $\text{Im}(p) = \mathbb{R}_+$, hence for all $t > 0$, $\psi(t) = t^\alpha \psi(1)$. Then up to a positive constant multiplicative factor, $\tilde{p}(x) = |x_1|^\alpha$ and $\tilde{\varphi}(t) = \varphi(t^{1/\alpha})$. And then if \tilde{p} is differentiable, we necessarily have that $\alpha > 1$.

In the case where $\alpha > 1$, for all $t > 0$, $\tilde{\varphi}'(t) = \frac{1}{\alpha} \frac{t^{\frac{1}{\alpha}-1}}{1+\log^2(t^{1/\alpha})}$ and then $\tilde{\varphi}$ is not differentiable at 0. \square

Yet we can prove for all continuously differentiable SI functions, f , that the level set of f going through x , i.e., $\mathcal{L}_{f,x}$, is included in the level set of $z \mapsto \nabla f(z) \cdot z$ going through x .

Lemma 4.1 *For a continuously differentiable SI function f and for $x \in \mathbb{R}^n$,*

$$\mathcal{L}_{f,x} \subset \mathcal{L}_{z \mapsto \nabla f(z) \cdot z, x} = \{y \in \mathbb{R}^n, \nabla f(y) \cdot y = \nabla f(x) \cdot x\}. \quad (12)$$

That is, each level set of f has a single value of $\nabla f(x) \cdot x$ while also different level sets of f can have the same value of $\nabla f(x) \cdot x$.

Proof Let $y \in \mathcal{L}_{f,x}$. Since $f(y) = f(x)$, then for all $t \geq 0$, $f(ty) = f(tx)$. We define the function h on \mathbb{R}_+ such that for all $t \geq 0$, $h(t) = f(tx) - f(ty)$. Then h is the zero function, so is its derivative: $h'(t) = \nabla f(tx) \cdot x - \nabla f(ty) \cdot y = 0$ for all $t \geq 0$. In particular we have the result for $t = 1$. \square

We exhibit in the next proposition a continuously differentiable SI function where the inclusion in the above lemma is strict (another example is Lemma 4.2).

Proposition 4.8 *Let p be the PH_2 function $x \in \mathbb{R}^n \mapsto \|x\|^2$ and φ the strictly monotonic function $\varphi(t) = \exp(-t)$ for all $t \geq 0$. Then $f : x \mapsto \varphi(p(x)) = \exp(-\|x\|^2)$ is continuously differentiable. For any $0 < r < 1$, there is a unique $s > 1$ such that for any $x \in \mathcal{S}_r$:*

$$\mathcal{L}_{z \mapsto \nabla f(z) \cdot z, x} = \mathcal{L}_{f,x} \cup \mathcal{L}_{f, \frac{s}{r}x}$$

where $\mathcal{L}_{f,x}$ and $\mathcal{L}_{f, \frac{s}{r}x}$ are disjoint.

Proof Remark that the transformation could be chosen to obtain a degree equal to 1 as in Theorem 3.1, but the differentiability of p would not be guaranteed.

We notice that $t \rightarrow t\varphi'(t)$ is not injective on \mathbb{R}_+ . It is injective on $[0, 1)$ and on $[1, \infty)$ and for any $0 < r < 1$ there is a unique $s > 1$ such that

$$r^2 \varphi'(r^2) = s^2 \varphi'(s^2). \quad (13)$$

We will prove that for any such r, s and any $x \in \mathcal{S}_r$,

$$\{y \in \mathbb{R}^n, \nabla f(y) \cdot y = \nabla f(x) \cdot x\} = \mathcal{L}_{f,x} \cup \mathcal{L}_{f,\frac{s}{r}x} \quad (14)$$

Let $x \in \mathbb{R}^n$ such that $\|x\| = r$. By the chain rule, for all $y \in \mathbb{R}^n$ we have

$$\nabla f(y) \cdot y = \varphi'(p(y))\nabla p(y) \cdot y = 2\varphi'(p(y))p(y) \ .$$

Therefore $y \in \{y \in \mathbb{R}^n, \nabla f(y) \cdot y = \nabla f(x) \cdot x\}$ if and only if $\|y\|^2\varphi'(\|y\|^2) = r^2\varphi'(r^2)$. From (13), we know that this is possible only if $\|y\| = r$ or $\|y\| = s$, i.e., only if $f(y) = f(x)$ or $f(y) = f(\frac{s}{r}x)$. Hence the equality in (14).

We remark that $\mathcal{L}_{f,x}$ and $\mathcal{L}_{f,\frac{s}{r}x}$ are disjoint whenever $f(x) \neq f(\frac{s}{r}x)$. If $x \in \mathcal{S}_r$, $f(x) = e^{-r^2} \neq e^{-s^2} = f(\frac{s}{r}x)$ which implies that $\mathcal{L}_{f,x}$ and $\mathcal{L}_{f,\frac{s}{r}x}$ are disjoint. \square

The non-injectivity of $t \rightarrow t\varphi'(t)$ is essential in the above example to obtain a non-strict inclusion in (12) for some SI functions. We obtain a weak formulation of Euler's homogeneous function theorem for some SI functions in the following proposition.

Proposition 4.9 *Let f be a continuously differentiable SI function that can be written as $\varphi \circ p$ where φ is a homeomorphism, p is PH_1 and φ and φ^{-1} are continuously differentiable. Assume that the function $t \in \mathbb{R}_+ \mapsto t\varphi'(t) \in \mathbb{R}$ is injective. Then for $x \in \mathbb{R}^n$,*

$$\mathcal{L}_{f,x} = \mathcal{L}_{z \mapsto \nabla f(z) \cdot z, x} \ . \quad (15)$$

Proof It follows from Proposition 4.6 that for all $x \in \mathbb{R}^n$, $\nabla f(x) \cdot x = \varphi'(p(x))p(x)$. Thanks to the bijectivity of φ along with the injectivity of $t \rightarrow t\varphi'(t)$, we have:

$$\varphi'(p(y))p(y) = \varphi'(p(x))p(x) \iff p(x) = p(y) \iff f(x) = f(y).$$

In other words, $\mathcal{L}_{f,x} = \{y \in \mathbb{R}^n, \nabla f(y) \cdot y = \nabla f(x) \cdot x\}$. \square

4.6 Compact Neighborhoods of Level Sets with Non-Vanishing Gradient

We prove in this section that any continuously differentiable SI function f with a unique global argmin has level sets, for example \mathcal{L}_{f,z_0} , such that for some compact neighborhood of the level set, $\mathcal{N} \supset \mathcal{L}_{f,z_0}$, the gradient does not vanish and $\nabla f(z) \cdot z > 0$ for all $z \in \mathcal{N}$.

For a continuously differentiable PH function p such that $p(x) > 0$ for all $x \neq 0$, i.e., such that 0 is the unique global argmin of p , this result is a consequence of Euler's homogeneous function theorem which implies that

$$\nabla p(x) \cdot x > 0 \text{ for all } x \neq 0 \ . \quad (16)$$

In particular, (16) is true on any compact neighborhood of any level set of p , if that compact does not contain 0.

We now remark that the property that $\nabla f \neq 0$ for all $x \neq 0$ is not necessarily true if f is a continuously differentiable SI function with a unique global argmin. Namely, f can have level sets that contain only saddle points.

Lemma 4.2 *Let $p(z) = \|z\|^2$ and $\varphi(t) = \int_0^t \sin^2(u)du$ for $t \geq 0$. Then $f = \varphi \circ p$ is a continuously differentiable SI function with a unique global argmin and an infinite number of z belonging to different level sets of f , such that $\nabla f(z) = 0$.*

Proof The function φ is strictly increasing since \sin^2 is non-negative and has zeros on isolated points. Also, for all $t \geq 0$, $\varphi(t) = \frac{t}{2} - \frac{\sin(2t)}{4}$, where we use that $\cos(2t) = 1 - 2\sin^2(t)$.

For any natural integer n , $n\pi$ is a stationary point of inflection of φ : $\varphi'(n\pi) = 0$ and $\varphi''(t) = \sin(2t)$ has opposite signs in the neighborhood of $n\pi$. For all z with $\|z\|^2 \in \pi\mathbb{Z}_+$, $\nabla f(z) = \varphi'(g(z))\nabla g(z) = 2\varphi'(\|z\|^2)z = 0$.

Hence there exists an infinite number of level sets $\mathcal{L}_{f,z}$ for which $\nabla f(z) = 0$. \square

Yet, a consequence of Theorem 4.2 and Lemma 4.1 is the existence of a level set of f such that $\nabla f(z) \cdot z > 0$ for all z in that level set as shown in the next proposition.

Proposition 4.10 *Let f be a continuously differentiable SI function with 0 as unique global argmin. There exists $z_0 \in \overline{\mathcal{B}}_1$ with $\mathcal{L}_{f,z_0} \subset \overline{\mathcal{B}}_1$, such that for all $z \in \mathcal{L}_{f,z_0}$, $\nabla f(z) \cdot z > 0$.*

Proof Since f is a continuous SI function, we have all the equivalences in Theorem 4.2.

Inside the proof of Theorem 4.2, we have shown that there exists $s \in \mathcal{S}_1$ such that $\mathcal{L}_{f,s}^{\leq} \subset \overline{\mathcal{B}}_1$, with $f(s) = \min_{z \in \mathcal{S}_1} f(z)$. Since f_s is strictly increasing and differentiable, there exists $t \in (0, 1]$ such that $f'_s(t) > 0$. Let us denote $z_0 = ts$. We have that $\mathcal{L}_{f,z_0} \subset \mathcal{L}_{f,s}^{\leq} \subset \overline{\mathcal{B}}_1$. And with the chain rule, $0 < f'_s(t) = \nabla f(z_0) \cdot \frac{z_0}{t}$. Therefore along with Lemma 4.1, it follows that for all $z \in \mathcal{L}_{f,z_0}$, $\nabla f(z) \cdot z = \nabla f(z_0) \cdot z_0 > 0$. \square

From the uniform continuity of $z \mapsto \nabla f(z) \cdot z$ on a compact we deduce the announced result.

Proposition 4.11 *Let f be a continuously differentiable SI function with 0 as unique global argmin. There exists $\delta > 0$, $z_0 \in \overline{\mathcal{B}}_1$ with $\mathcal{L}_{f,z_0} \subset \overline{\mathcal{B}}_1$ such that for all $z \in \mathcal{L}_{f,z_0} + \overline{\mathcal{B}}(0, \delta)$, $\nabla f(z) \cdot z > 0$.*

Proof Since $\nabla f(z) \cdot z > 0$ for all z in the compact \mathcal{L}_{f,z_0} , then $z \mapsto \nabla f(z) \cdot z$ has a positive minimum (that is reached) denoted by $\epsilon = \min_{z \in \mathcal{L}_{f,z_0}} \nabla f(z) \cdot z > 0$. The continuous function $z \mapsto \nabla f(z) \cdot z$ is uniformly continuous on the compact $\mathcal{L}_{f,z_0} + \overline{\mathcal{B}}(0, 1)$, therefore there exists a positive number $\delta < 1$ such that if $y, z \in \mathcal{L}_{f,z_0} + \overline{\mathcal{B}}(0, 1)$ with $\|y - z\| \leq \delta$ then $|\nabla f(z) \cdot z - \nabla f(y) \cdot y| < \frac{\epsilon}{2}$. Then for all $z \in \mathcal{L}_{f,z_0} + \overline{\mathcal{B}}(0, \delta)$, there exists $y \in \mathcal{L}_{f,z_0}$ such that $|\nabla f(z) \cdot z - \nabla f(y) \cdot y| < \frac{\epsilon}{2}$. Then $\nabla f(z) \cdot z > \nabla f(y) \cdot y - \frac{\epsilon}{2} \geq \frac{\epsilon}{2} > 0$. Hence $z \mapsto \nabla f(z) \cdot z$ is positive on the compact set $\mathcal{L}_{f,z_0} + \overline{\mathcal{B}}(0, \delta)$. \square

5 Summary and Conclusion

This paper reveals that *continuous* scaling-invariant functions are strictly monotonic transformations of *continuous* positively homogeneous functions. Moreover, we present necessary and sufficient conditions for any scaling-invariant function to be a strictly monotonic transformation of a positively homogeneous function. The derivation is solely based on analyzing restrictions to the half-lines starting from zero that need to be strictly monotonic on a nontrivial interval (or entirely flat). We also highlight counter-intuitive examples of scaling-invariant functions that are not monotonic on any nontrivial interval.

We then present different properties of the level sets of a scaling-invariant function. In particular, Proposition 4.11 shows that continuously differentiable scaling-invariant functions with a unique argmin have a compact level set in a compact neighborhood with non-vanishing gradient. The level set intersects any half-line with origin zero at a single point—forming a “star-shaped” manifold.

Scaling-invariant functions play a central role in the analysis of the convergence of some comparison-based stochastic optimization algorithms [3]. On this function class, for some translation and scale invariant comparison-based algorithms, linear convergence can be deduced when analyzing the stability of a normalized process⁴. When linear convergence occurs, the step-size and the distance of the current solution to the optimum decrease geometrically fast to zero at the same (linear) rate.

A stability analysis leading to linear convergence can be carried out for composites of strictly increasing functions with continuously differentiable scaling-invariant functions. To obtain basic stability properties deduced from a connection to a deterministic control model [5], one can exploit that these functions have Lebesgue negligible level sets as a consequence of Proposition 4.2. In addition, the stability study relies on proving that when the normalized process diverges, the step-size multiplicative factor converges in distribution to the factor on nontrivial linear functions. The proof exploits level set properties shown in Proposition 4.11 and Corollary 4.1.

Acknowledgements

Part of this research has been conducted in the context of a research collaboration between Storengy and Inria. We particularly thank F. Huguet and A. Lange from Storengy for their strong support. We would like to thank the anonymous referee for their very careful reviewing of the paper and their suggestions that lead to an improved version.

⁴ In the case of step-size adaptive algorithms where the state of the algorithm equals a current solution and a step-size, the normalized process equals to the solution minus the reference point x^* of the scaling-invariant function, divided by the step-size. Stability of the normalized process is key to imply linear convergence of the adaptive algorithm.

References

1. Aczél, J.: Lectures on Functional Equations and their Applications. Academic Press, New York (1966)
2. Auger, A., Hansen, N.: 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. arXiv e-prints 1310.8397 [cs.NA] (2013)
3. Auger, A., Hansen, N.: Linear convergence of comparison-based step-size adaptive randomized search via stability of Markov chains. *SIAM Journal on Optimization* **26**(3), 1589–1624 (2016)
4. Buskes, G., van Rooij, A.: Topological Spaces. Springer-Verlag New York (1997)
5. Chotard, A., Auger, A.: Verifiable conditions for the irreducibility and aperiodicity of Markov chains by analyzing underlying deterministic models. *Bernoulli* **25**(1), 112–147 (2019)
6. Denjoy, A.: Sur les fonctions dérivées sommables. *Bulletin de la Société Mathématique de France* **43**, 161–248 (1915)
7. Dutta, J., Martinez-Legaz, J., Rubinov, A.: Monotonic analysis over cones: I. *Optimization* **53**(2), 129–146 (2004)
8. Fournier, H., Teytaud, O.: Lower bounds for comparison based evolution strategies using VC-dimension and sign patterns. *Algorithmica* **59**(3), 387–408 (2011)
9. Gorokhovich, V.V., Trafimovich, M.: Positively homogeneous functions revisited. *Journal of Optimization Theory and Applications* **171**(2), 481–503 (2016)
10. Gorokhovich, V.V., Trafimovich, M.: Saddle representations of positively homogeneous functions by linear functions. *Optimization Letters* **12**(8), 1971–1980 (2018)
11. Hardy, G.H.: Weierstrass’s non-differentiable function. *Trans. Amer. Math. Soc.* **17**(3), 301–325 (1916)
12. Kuczma, M.: An Introduction to the Theory of Functional Equations and Inequalities: Cauchy’s Equation and Jensen’s Inequality. Birkhäuser (2009)
13. Lasserre, J.B., Hiriart-Urruty, J.B.: Mathematical properties of optimization problems defined by positively homogeneous functions. *Journal of optimization theory and applications* **112**(1), 31–52 (2002)
14. Morinaga, D., Akimoto, Y.: Generalized drift analysis in continuous domain: linear convergence of (1+1)-ES on strongly convex functions with Lipschitz continuous gradients. In: Proceedings of the 15th ACM/SIGEVO Conference on Foundations of Genetic Algorithms, pp. 13–24. Association for Computing Machinery (2019)
15. Muresan, M.: A Concrete Approach to Classical Analysis. CMS Books in Mathematics. Springer-Verlag New York (2009)
16. Nelder, J.A., Mead, R.: A simplex method for function minimization. *The Computer Journal* **7**(4), 308–313 (1965)
17. Rubinov, A., Gasimov, R.: Strictly increasing positively homogeneous functions with application to exact penalization. *Optimization* **52**(1), 1–28 (2003)
18. Rubinov, A., Glover, B.: Duality for increasing positively homogeneous functions and normal sets. *RAIRO - Operations Research* **32**(2), 105–123 (1998)

A Bijection Theorem

This standard theorem is reminded for the sake of completeness.

Theorem A.1 (Bijection theorem, [15, Theorem 2.20]) *Let $I \subset \mathbb{R}$ be a nontrivial interval, $J \subset \mathbb{R}$ and $\varphi : I \rightarrow J$ be a continuous bijection (and therefore strictly monotonic). Then J is an interval and φ is a homeomorphism, i.e., $\varphi^{-1} : J \rightarrow I$ is also a continuous bijection, and if φ is strictly increasing (respectively strictly decreasing), then φ^{-1} is strictly increasing (respectively strictly decreasing).*