Discussiones Mathematicae
Differential Inclusions, Control and Optimization 20 (2000) 257–278

A PRIMAL-DUAL INTEGRAL METHOD IN GLOBAL OPTIMIZATION

Jens Hichert

Geomagic GmbH
Harbergstr. 73, D-99099 Erfurt, Germany

Armin Hoffmann

Institute of Mathematics
Technical University Ilmenau
PF 100565, D-98684 Ilmenau, Germany

e-mail: armin.hoffmann@mathematik.tu-ilmenau.de

Huan Xoang Phú

Institute of Mathematics Hanoi
P.O. Box 631 Bo Ho, 10 000 Hanoi, Vietnam

and

Rüdiger Reinhardt

Institute of Mathematics
Technical University Ilmenau
PF 100565, D-98684 Ilmenau, Germany

Abstract

Using the Fenchel conjugate F^c of Phú's Volume function F of a given essentially bounded measurable function f defined on the bounded box $D \subset \mathbb{R}^n$, the integral method of Chew and Zheng for global optimization is modified to a superlinearly convergent method with respect to the level sequence. Numerical results are given for low dimensional functions with a strict global essential supremum.

Keywords: global optimization, integral method, Monte Carlo method, primal dual algorithm, level set method.

1991 Mathematics Subject Classification: 90C30, 65C05, 65K05, 49M29.

1 Introduction and historical remarks

Let $D = [a, b] = \prod_{j=1}^{n} [a^j, b^j] \subset \mathbb{R}^n$ be a bounded box and let $f : D \to \mathbb{R}$ be a Lebesgue measurable function. We denote with $\mu(A)$ the n-dimensional Lebesgue measure of a measurable set A. Throughout this paper we use the following abbreviations

$$\alpha^* = \operatorname{ess\,sup} f := \inf_{\mu(B)=0} \sup_{x \in D \setminus B} f(x), \qquad \max f := \max_{x \in D} f(x),$$
$$\alpha_* = \operatorname{ess\,inf} f := -\operatorname{ess\,sup}(-f), \qquad \min f := -\max(-f),$$
$$\arg\max f := \{x \in D \mid f(x) = \max f\}, \quad [f \geq \alpha] := \{x \in D \mid f(x) \geq \alpha\}.$$

Analogously to the upper level set $[f \geq \alpha]$ we define $[f \leq \alpha]$, $[f = \alpha]$, $[f > \alpha]$ and $[\alpha < f < ß]$. We denote with $\int_{[f \geq \alpha]} f(x)\, d\mu(x)$ the n-dimensional Lebesgue integral of the function f over the upper level set $[f \geq \alpha]$ and assume that ess sup $|f| < \infty$.

We want to study the determination of the ess sup$f$ from the theoretical and numerical point of view. The classical integral global optimization method of Chew and Zheng [1] and its modifications developed by Phú and Hoffmann [10], Hichert, Hoffmann, Phú [5] and Hichert [3, 4] play a central role. We propose a modification which generates a superlinearly convergent level sequence. Using a Monte-Carlo implementation we compare our modification with the original **Zheng-method** (further cited as **ZM**) [1] for test examples with strict global maximizer. Straightforward simple programming under MATLAB already shows that improvements of **ZM** are obtained by using the above mentioned acceleration.

In our paper branching strategies (see e.g. [3, 4]) are not considered to avoid mixing of influences of several strategies. Branching strategies are necessary in the cases when more than one global maximizer exists, when there are some far away local maximizers which have a function value close to the essential supremum or when the dimension of the problem is not low.

In Section 2 we give a short review of the main ideas of **ZM** and its implementation. Referring essential results of [10] Section 3 is devoted to another description of **ZM** as a Newton-method replacing the mean value of Zheng by a Newton-step for a convex volume function F. In Section 4 we consider the Fenchel conjugate F^c of the volume function F. We have found that the smallest zero of F^c again characterizes the essential supremum of f. Using the Newton algorithm for finding this first zero and reformulating the algorithm with respect to the function F we obtain in Section 5 a

primal-dual-method (further cited as **PDM**) for the determination of the essential supremum of an almost all bounded measurable function f. Under mild conditions, the associated level sequence is superlinearly convergent to $\operatorname{ess\,sup} f$. Straightforward implementations and tests in Section 6 show that **ZM** and **PDM** run very stable and find the global maximum with similar probability and accuracy. However, **PDM** needs in the average much less function evaluations.

2 Integral global optimization (ZM)

2.1 Theoretical results

In [1] Chew and Zheng proposed the following theoretical method for the determination of the global maximum of a continuous function $f : D \to \mathbb{R}$.

Algorithm 1. (**ZM**)

Initialization: Choose $\alpha_0 < \alpha^*$

Iteration:

$$\alpha_{k+1} = \frac{\int_{[f \geq \alpha_k]} f(x)\, d\mu(x)}{\mu[f \geq \alpha_k]} =: M_f(\alpha_k) \quad k = 0, 1, 2, ... \tag{1}$$

Formula (1) can be interpreted that α_{k+1} is the mean value $M_f(\alpha_k)$ of f over the upper level set $[f \geq \alpha_k]$. The monotonicity, upper estimation and continuity of $\alpha \to M_f(\alpha)$ for all α with $\mu[f \geq \alpha] > 0$ justify the definition of $M_f(\bar{\alpha}) := \lim_{\alpha \uparrow \bar{\alpha}} M_f(\alpha)$ in the case $\mu[f \geq \alpha] > 0$ for all $\alpha < \bar{\alpha}$ and $\mu[f \geq \bar{\alpha}] = 0$. It is easy to verify that $M_f(\alpha) \leq \alpha^*$ for all levels $\alpha \leq \alpha^*$ and that $M_f(\alpha) = \alpha$ implies $\alpha = \alpha^*$.

Proposition 1 [1]**.** *The above defined level sequence* $\{\alpha_k\}$ *converges monotonically to the value* α^* *and if* f *is continuous on* D *then*

$$\arg\max f = \bigcap_{k=0}^{\infty} [f \geq \alpha_k].$$

The following definitions (cf. [1] and [10]) are useful to decide whenever $\operatorname{ess\,sup} f = \sup f$.

Definition 1. A measurable set $S \subset \mathbb{R}^n$ is called **robust** if $\operatorname{cl} \operatorname{int} S = \operatorname{cl} S$.

Definition 2. A function f from a measurable set $D \subset \mathbb{R}^n$ into $\mathbb{R}$ is called **robust** if the lower level set $[f < \alpha]$ is robust for arbitrary real α.

Definition 3. A function f from a measurable set $D \subset \mathbb{R}^n$ into $\mathbb{R}$ is called **dense** if $\mu[y - \varepsilon < f < y + \varepsilon] > 0$ for each $\varepsilon > 0$ and each $y \in f(D)$.

Proposition 2 [10]**.**

1. *Continuous functions over robust sets are robust and dense.*
2. *If D is a union of robust sets D_j and f is continuous on each such set D_j then f is dense on D.*
3. *If f is dense or lower semi continuous or if $-f$ is robust then* $\operatorname{ess\,sup} f = \sup f$.

Remark 1. In [13] and subsequent publications the above method is outlined for robust functions, robust analysis included. Applications of this method are considered e.g. in [14].

2.2 Numerical realization of ZM with Monte-Carlo strategy

The theoretically proposed method causes tremendous trouble since the integral $\int_{[f \geq \alpha_k]} f(x)\, d\mu(x)$ and the measure $\mu[f \geq \alpha_k]$ of the upper level set $[f \geq \alpha_k]$ must be numerically evaluated. This is impossible with reasonable efforts by direct integration methods already for simple functions f. However, there is a very rough but simple and applicable Monte-Carlo approximation of the quotient in Algorithm 1. For a deeper study of the numerical realization and theoretical results connected with it, we refer to [1, Part IV] or [13, Chapter 6].

Let a starting level $\alpha_0 < \alpha^*$ be given, e.g. by $\alpha_0 = f(x_0)$ for some $x_0 \in D_0 := D$. We evaluate the function f for a sufficiently large number of points $x_k \in D_0$ uniformly distributed in D_0 such that t function values of them are not smaller than α_0. After ordering with respect to the function values and renumbering the x_k points, we have

$$f(x_t) \geq f(x_{t-1}) \geq \cdots \geq f(x_1) \geq \alpha_0 > f(x_0) \geq \cdots$$

The mean value $M_f(\alpha_0)$ over the upper level set $[f \geq \alpha_0]$ can now be roughly approximated by

$$\hat{M}_f(\alpha_0) := \frac{1}{t} \sum_{k=1}^{t} f(x_k)$$

and following the theoretical Algorithm 1

$$\alpha_1 = \hat{M}_f(\alpha_0)$$

is the next level. Now we look for a box $D_1 := D(x_1, ..., x_t) \supset \{x_1, ..., x_t\}$ which can be expected w.r.t. the used uniform distribution of the points evaluated above. This box is a rough approximation of the level set $[f \geq \alpha_0]$. We give a short description of this procedure in the detail [1]. For $x_1, ..., x_t \in \mathbb{R}^n$ we define (the upper index is the index of the coordinate) $x_{\min} := \min\{x_1, ..., x_t\} := (\min\{x_1^1, ..., x_t^1\}, ..., \min\{x_1^n, ..., x_t^n\})$ and similarly $x_{\max} := \max\{x_1, ..., x_t\}$. Let $x_L := x_{\min} - (x_{\max} - x_{\min})/(t-1)$ and $x_R := x_{\max} + (x_{\max} - x_{\min})/(t-1)$ then $D_1 = \Pi_{k=1}^n \left[x_L^k, x_R^k\right]$ with its measure $\mu(D_1) = \Pi_{k=1}^n \left(x_R^k - x_L^k\right)$. The box D_0 and the level α_0 are replaced by D_1 and α_1 in the next iteration, respectively. We can decide to compute entirely new points in D_1 or to use again some or all points with $f(x_k) \geq \alpha_1$ from the previous iteration. The iteration can be stopped, e.g. if the statistical variance $\hat{V}_f(\alpha_k) = \frac{1}{t-1}\sum_{k=1}^t \left(f(x_k) - \hat{M}_f(\alpha_k)\right)^2 \approx \frac{1}{t-1}\sum_{k=1}^t (f(x_k) - \alpha_k)^2$ is smaller than a given tolerance $tol > 0$.

3 The volume function F of Phú and ZM

Phú made the remark in [10, p. 168]: *the water volume F in (almost) all natural lakes is convex with respect to the height α of the water surface.* This property yields new insight in the integral global optimization. We give a new interpretation of **ZM** which allows improvements of the numerical strategy. In the following two subsections we use results of the papers [5] and [10]. We list them here without proofs.

3.1 Definition and properties of F

Let $D \subset \mathbb{R}^n$ be given with $0 < \mu(D) < \infty$ and assume $f : D \to \mathbb{R}$ is summable over D.

Definition 4. $F : \mathbb{R} \to \mathbb{R}$ with $F(\alpha) := \int_{[f \geq \alpha]} (f(x) - \alpha)\, d\mu(x)$ is called **volume function** of f and $m, m_+ : \mathbb{R} \to \overline{\mathbb{R}}$, with $m(\alpha) := \mu[f \geq \alpha]$, $m_+(\alpha) := \mu[f > \alpha]$ are called **upper level set measure functions** of f.

α^* denotes the smallest zero of F, i.e. $\alpha^* = \sup\{\alpha \,|F(\alpha) > 0\} = \inf\{\alpha \,|F(\alpha) = 0\} = \sup\{\alpha \,|m(\alpha) > 0\}$. The volume function F has very

interesting topological properties and is closely connected with the upper level set measure functions.

Proposition 3. *The function F is convex, Lipschitzian on $\mathbb{R}$ with the constant $\mu(D)$, decreasing, nonnegative, almost everywhere differentiable on $\mathbb{R}$ with the derivative $F'(\alpha) = -m(\alpha)$, it has the subdifferential $\partial F(\alpha) = [-m(\alpha), -m_+(\alpha)]$ for all $\alpha \in \mathbb{R}$ and it is linear below α_* with slope $-\mu(D)$ and zero above α^*.*

It is well-known that the set-valued mapping $\alpha \to \partial F(\alpha)$ is maximally monotone and upper semicontinuous as set-valued mapping between the reals. This implies immediately the left-hand continuity of m and the right-hand continuity of m_+. Especially $\lim_{\alpha \downarrow \bar{\alpha}} m(\alpha) = m_+(\bar{\alpha})$ holds.

The following statements contain relations between the behavior of f and the continuous differentiability of F.

Proposition 4.

1. *F is continuously differentiable on the open set U if and only if $\mu[f = \alpha] = 0$ for all $\alpha \in U$.*
2. *If $W \supset D$ is an open set, f is continuously differentiable on W and $\nabla f(x) \neq 0$ almost everywhere (a.e.) on D then $\mu[f = \alpha] = 0$ for all $\alpha \in \mathbb{R}$.*

Remark 2. Be aware that there are a continuously differentiable strictly increasing function f on $[0,1]$ and a subset $A \subset [0,1]$ such that $\mu(A) > 0$, $f'(x) = 0$ on A and $\mu[f = \alpha] = 0$ for all $\alpha \in \mathbb{R}$ ([5, Example 3.1]).

Now we investigate connections between the multiplicity of the zero α^* of F and the smoothness of f in some neighbourhood of $x^* \in D$ with $f(x^*) = \alpha^*$.

Definition 5. Let $Q = (q_{ij})_{nn}$ be a positive definite matrix with the positive eigenvalues $\lambda_k, k = 1, 2, ..., n$. A **modified $\mathbf{l_p}$-norm** for some $x \in \mathbb{R}^n$ and $p \in [1, \infty)$ is defined by

$$\|x\|_{Q,p} = \left[\sum_{j=1}^{n} \sum_{i=1}^{n} |x_i|^{\frac{p}{2}} q_{ij} |x_j|^{\frac{p}{2}} \right]^{\frac{1}{p}}.$$

We start with a simple but principal example.

Example 1. Let $f(x) = \alpha^* - \left(\|x - x^*\|_{Q,p}\right)^p$, $x \in \mathbb{R}^n$ be given. We calculate the measure $m(\alpha)$ for some $\alpha \leq \alpha^*$. Using the known volume formula

$$c_{n,p} = \left(\frac{2}{p}\Gamma\left(\frac{1}{p}\right)\right)^n / \ \Gamma\left(\frac{n+p}{p}\right)$$

for the n-dimensional l_p-unit ball, we get by elementary calculations

$$m(\alpha) = c_{n,p} \prod_{i=1}^{n} \lambda_i^{-\frac{1}{p}} (\alpha^* - \alpha)^{\frac{n}{p}} \text{ for all } \alpha \leq \alpha^*. \tag{2}$$

Since m is continuous, we find by integration

$$F(\alpha) = \frac{p}{p+n} c_{n,p} \prod_{i=1}^{n} \lambda_i^{-\frac{1}{p}} (\alpha^* - \alpha)^{\frac{p+n}{p}} \text{ for all } \alpha \leq \alpha^*.$$

If f is defined on a cube D and x^* belongs to the interior of D then the above formulas are valid for all $\alpha \in [\alpha_0, \alpha^*]$, where α_0 is sufficiently close to α^*.

If we can find lower and upper estimations of f by hyperparaboloids of the above kind then a similar behavior can be derived in the case of a unique global maximizer (cf. [5, Proposition 4.2]).

Definition 6. Assume that $x^* \in \operatorname{int} D$ is a global maximizer of f and there are two positive definite matrices Q and $\bar{Q}$ with positive eigenvalues λ_k and $\bar{\lambda}_k$, respectively and numbers $p, \bar{p} \in [1, \infty)$ such that

$$\bar{g}(x) := f(x^*) - \left(\|x - x^*\|_{\bar{Q},\bar{p}}\right)^{\bar{p}} \leq f(x) \leq f(x^*) - \left(\|x - x^*\|_{Q,p}\right)^p =: g(x)$$

for all $x \in [g \geq \alpha]$, where $\alpha < f(x^*)$ is sufficiently close to $f(x^*)$. Then x^* is called a **strict global maximizer of f of order $\{\bar{p}, p\}$**. If $\bar{p} = p$, we say that x^* is a **strict global maximizer of f of order p**.

In the following we use the abbreviation $R_Q := \prod_{k=1}^{n} \lambda_k^{\frac{1}{n}}$.

Proposition 5. *If x^* is a strict global maximizer of f of order $\{\bar{p}, p\}$ then $p \geq \bar{p}$ and there is some $\alpha_0 < \alpha^*$ such that for all $\alpha \in [\alpha_0, \alpha^*]$ the estimations*

$$c_{n,\bar{p}} R_{\bar{Q}}^{-\frac{n}{\bar{p}}} (\alpha^* - \alpha)^{\frac{n}{\bar{p}}} \leq m(\alpha) \leq c_{n,p} R_Q^{-\frac{n}{p}} (\alpha^* - \alpha)^{\frac{n}{p}} \tag{3}$$

and

$$\frac{\bar{p}}{\bar{p}+n} c_{n,\bar{p}} R_{\bar{Q}}^{-\frac{n}{\bar{p}}} (\alpha^* - \alpha)^{\frac{\bar{p}+n}{\bar{p}}} \leq F(\alpha) \leq \frac{p}{p+n} c_{n,p} R_Q^{-\frac{n}{p}} (\alpha^* - \alpha)^{\frac{p+n}{p}} \tag{4}$$

are valid.

Proof. By construction, we have $[\bar{g} \geq \alpha] \subset [f \geq \alpha] \subset [g \geq \alpha] \subset D$, whenever $\alpha < \alpha^*$ is sufficiently close to $\alpha^* = f(x^*)$. Hence $\mu[\bar{g} \geq \alpha] \leq \mu[f \geq \alpha] \leq [g \geq \alpha]$ which gives together with (2) the estimation (3). The integration over α implies (4). ∎

Remark 3. F has in α^* a root of order 1 only if $m(\alpha^*) > 0$, i.e. there is a set of global maximizers with positive Lebesgue measure – a very exceptional case. Usually, the number of global maximizers is one or at least finite. The last case can similarly be treated as in Proposition 5. The maximizer with the greatest order determines the order of the root at α^* from the left-hand side. This means that the order of the root is in almost all practical cases greater than one.

3.2 Newton algorithm and integral global optimization

Using the volume function F of Phú, the essential supremum α^* can be found by looking for the smallest zero of F. Since F is convex, decreasing, Lipschitzian and a.e. differentiable, the zero can be determined by the following Newton-algorithm of [10, Algorithm 2].

Algorithm 2. (**Primal method**)

Initialization: $\alpha_0 = f(x_0)$ for some $x_0 \in D$

Iteration:

$$\alpha_{k+1} = \alpha_k + \frac{F(\alpha_k)}{\mu[f \geq \alpha_k]}. \tag{5}$$

Because of $\max \partial F(\alpha_k) = -m_+(\alpha_k)$ this algorithm can be accelerated when F is nondifferentiable in α_k by using $\mu[f > \alpha_k]$ instead of $\mu[f \geq \alpha_k]$. The global convergence of the sequence $\{\alpha_k\}$ generated by Algorithm 2 is ensured (cf. [10, Proposition 2.3]) and for arbitrary starting point already the first iteration belongs to the interval $[\alpha_*, \alpha^*]$ (cf. [10, Proposition 2.2]). It is obvious that Algorithm 2 is only a new notation of **ZM** (Algorithm 1), indeed

$$M_f(\alpha_k) = \alpha_k + \frac{F(\alpha_k)}{\mu[f \geq \alpha_k]}.$$

However, the above notation as a Newton type method gives new insights in the real behavior of this algorithm and yields some approvement.

Proposition 6. *If* $\mu[f \geq \alpha^*] > 0$, *i.e.* $\partial F(\alpha^*) = [-\mu[f \geq \alpha^*], 0]$ *is not a singleton, then the above defined sequence* $\{\alpha_k\}$ *is superlinearly convergent.*

Proof. For a smooth F in some pointed neighbourhood on the left of α^* the proof is given in [5, Proposition 3.1]. If F is not differentiable at some point in each pointed neighbourhood on the left of α^* then, because of the strict monotonicity of the sequences $\{\alpha_k\}$ and $\{\mu[f \geq \alpha_k]\}$, there exists a convex and differentiable function G such that $F(\alpha_k) = G(\alpha_k)$ and $G'(\alpha_k) = -\mu[f \geq \alpha_k]$. Choose, e.g. G as circular arc spline between $(\alpha_k, G(\alpha_k))$ and $(\alpha_{k+1}, G(\alpha_{k+1}))$ being tangential at these points, where the slopes of the tangents are $G'(\alpha_k)$ and $G'(\alpha_{k+1})$, respectively. $\lim_{k\to\infty} G'(\alpha_k) \leq \min \partial F(\alpha^*)$ and $\lim_{k\to\infty} G'(\alpha_k) \in \partial F(\alpha^*)$ ([6, Lemma 6.3.4]) show $G'(\alpha^* - 0) = \min \partial F(\alpha^*)$. ∎

Using Proposition 5 and Remark 3, we have the estimation $F(\alpha) = C(\alpha^* - \alpha)^\gamma + o((\alpha^* - \alpha)^\gamma)$ for all $\alpha \in [\alpha^* - \varepsilon, \alpha^*]$ and some $\gamma > 1$ and $\varepsilon > 0$ in almost all practical cases. Hence, $F'(\alpha^*) = 0$ and the sequence $\{\alpha_k\}$ is only linearly convergent with the convergence rate $(\gamma - 1)/\gamma$. It is well-known that the modified Newton iteration

$$\alpha_{k+1} = \alpha_k + \gamma \frac{F(\alpha_k)}{\mu[f \geq \alpha_k]}$$

with the step size γ is superlinearly convergent if for all iterations the inequality $a_k \leq \alpha^*$ can be ensured. Unfortunately, with some changes of the step size, this can only be ensured in a small neighbourhood of α^* (cf. [10, Proposition 4.2]). Geometrically, the use of such step size $\gamma > 1$ means that instead of the tangent a secant is used or the tangential support is replaced by a support with a parabola of order γ. Hence, γ has to be modified during the running algorithm by using suitable estimations. One can use the acceleration by Steffensen and Aitken (cf. [5, Algorithm 2], [4, Chapter 2.1]) or some generalized convexity properties ([4, Chapter 2.2]).

In this paper, we propose another acceleration, which ensures the inequality $a_k \leq \alpha^*$ for all k. The conjugate F^c of the volume function plays an essential role for the construction of such an algorithm.

4 Properties of the Fenchel conjugate F^c

The conjugate function F^c is defined by $F^c(u) = \sup_{\alpha\in\mathbb{R}}(u\alpha - F(\alpha))$. Let $\partial F^c(u) = [s(u), s_+(u)]$ be its subdifferential. We use the well-known facts (cf. e.g. [6, 7]) that $F^c : \mathbb{R} \to \mathbb{R}\cup\{+\infty\}$ is convex and lower semicontinuous (l.s.c.), dom $F^c = \bigcup_{\alpha\in\mathbb{R}} \partial F(\alpha) = [-\mu(D), 0]$, $\alpha \in \partial F^c(u) \Longleftrightarrow u \in \partial F(\alpha) \Longleftrightarrow F(\alpha) + F^c(u) = \alpha u$, $\alpha \to \partial F(\alpha)$ and $u \to \partial F^c(u)$ are upper semicontinuous (u.s.c.) and maximal monotone set-valued mappings. One can easily show that $\alpha^* = \operatorname{ess\,sup} f = \min \partial F^c(0) = s(0)$, $\alpha_* = \operatorname{ess\,inf} f = \max \partial F^c(-\mu(D)) = s_+(-\mu(D))$ and that F^c is Lipschitzian on dom F^c with the Lipschitz constant $L = \max\{|\alpha_*|, |\alpha^*|\}$. The dual variable u can be interpreted as the negative measure $-m(\alpha)$ of the upper level set $[f \geq \alpha]$ whenever $u = \min \partial F(\alpha)$.

Lemma 7. *Let $S = (-\mu(D), 0]$ and let $M = [\alpha_*, \alpha^*]$. Then it holds:*

1. *s is continuous on S if and only if m is strictly decreasing on M.*
2. *m is continuous on M if and only if s is strictly decreasing on S.*

Proof. 1. m is strictly decreasing if and only if ∂F^c is single valued.
a) sufficiency: The upper semicontinuity of ∂F^c implies the continuity of s.
b) necessity: The continuity of s implies that ∂F^c is single valued.
The proof of the second statement is similar. ∎

We show in the following that relatively weak properties of f imply continuous differentiability of the conjugate F^c.

Lemma 8. *If $f : D \to R$ is dense, then m and m_+ are strictly decreasing.*

Proof. Let $\varepsilon > 0$ be given. Then $0 < \mu[y-\varepsilon < f < y+\varepsilon] = \mu[f > y-\varepsilon] - \mu[f \geq y+\varepsilon] \leq \mu[f \geq y-\varepsilon] - \mu[f \geq y+\varepsilon] = m(y-\varepsilon) - m(y+\varepsilon)$. For m_+ the proof is similar. ∎

Proposition 9. *If f is dense on D then F^c is continuously differentiable on $(-\mu(D), 0)$ with $(F^c)'(u) = s(u)$.*

Proof. The density of f implies that m is strictly increasing which yields the continuity of s on $(-\mu(D), 0]$ ∎

If we know an upper bound b of the essential supremum of f then the transformed function $f - b$ is everywhere non positive. In this case the conjugate function F^c has similar properties as F.

Proposition 10. *The following statements are equivalent.*

(a) $\alpha^* \leq 0$,

(b) $F^c(u) \geq 0$ *on* dom F^c,

(c) F^c *is decreasing on* dom F^c.

Proof. (a)⇒(c): Let $-\mu(D) < u < 0$. Because of the monotonicity of ∂F^c, the inequality $(\alpha - \beta)u \geq 0$ holds for each $\alpha \in \partial F^c(u)$ and $\beta \in \partial F^c(0)$. With $\alpha = s(u), s_+(u)$ and $\beta = s(0)$ it follows $s(u) \leq s_+(u) \leq s(0) = \alpha^* \leq 0$. Hence, F^c is decreasing.

(c)⇒(b): $F^c(0) = 0$, $u \in \text{dom}\, F^c$ and F^c decreasing imply $F^c(u) \geq 0$.

(b)⇒(a): F^c is Lipschitzian and convex, $F^c(u) = 0$, $u < 0$ and $F^c(0) = 0$. Hence, $u \to \frac{F^c(u)-F^c(0)}{u}$ is increasing and nonnegative. This implies $\alpha^* = \min \partial F^c(0) = \lim_{u \to -0} \frac{F^c(u)-F^c(0)}{u} \leq 0$. ■

The smallest zero u^* of the conjugate function F^c plays a similar role as α^* for F provided $\alpha^* \leq 0$. Obviously, we have $u^* := \sup\{u \in \text{dom}\, F^c \,|F^c(u) > 0\}$. The following two properties are essential for the computation of u^* with a Newton type algorithm.

Lemma 11. *If* $\alpha^* \leq 0$ *then* $\alpha^* u^* = 0$.

Proof. $u^* < 0$ implies $F^c(u) = 0$ in $[u^*, 0]$. Hence $\partial F^c(0) = [0, +\infty)$ and therefore, $\alpha^* = \min \partial F^c(0) = 0$. ■

Lemma 12. *If* $\alpha^* \leq 0$ *and* $F^c(u) > 0$ *then* $s(u) < 0$.

Proof. The equivalence $[s(u) \in \partial F^c(u) \iff F(s(u)) + F^c(u) = u \cdot s(u)]$, and $u < 0$, $F^c(u) > 0$, $F(\alpha) \geq 0$ yield $s(u) < 0$. ■

If $\alpha^* \leq 0$ then the smallest zero $u^* = \min\{u \in \text{dom}\, F^c \,|F^c(u) = 0\}$ can be calculated by the Newton-algorithm.

Algorithm 3. (**Dual method**)

Initialization: $u_0 = -m(f(x_0))$ for some $x_0 \in D$,

Iteration:

$$u_{k+1} = u_k - \frac{F^c(u_k)}{s(u_k)}. \tag{6}$$

It follows that $u_k \leq u^*$ for each k. Because of Lemma 12, first, the denominator $s(u_k)$ is negative as long as the value u^* is not reached, and second,

the sequence $\{u_k\}$ is strictly increasing. Together with Lemma 11 we immediately obtain that either $u^* < 0$ which implies $\alpha^* = 0$ or $u^* = 0$ and $\alpha^* = \min \partial F^c(0)$. Now we investigate the convergence properties of the sequence $\{u_k\}$ generated by Algorithm 3 for several constellations between $u^* \leq 0, \alpha^* \leq 0$. Be aware that $u^*\alpha^* \neq 0$ is impossible.

Case 1. $u^* < 0$ and $\alpha^* = 0$
We have $\partial F^c(u) = \{0\}$ for all $u \in (u^*, 0) \neq \emptyset$. Hence $(u^*, 0) \subset \partial F(\alpha^*)$, i.e. the zero $\alpha^* = 0$ is single. By Proposition 6, **ZM** generates a superlinearly convergent sequence $\{\alpha_k\}$.

Case 2. $u^* = 0$ and $\alpha^* < 0$

1. We have $\partial F^c(u^*) = \partial F^c(0) = [s(0), +\infty)$ and $s(0) = \alpha^* < 0$. Therefore, $u^* = 0$ is a single zero of F^c and the sequence $\{u_k\}$ generated by Algorithm 3 is superlinearly convergent according to similar arguments now used in Proposition 6 for F^c. Below we quantify this behavior.
2. If, additionally, $\partial F(\alpha^*) = [-m(\alpha^*), 0]$ is a nontrivial interval then F^c is linear affine in the interval $[-m(\alpha^*), 0]$, and the algorithm stops after finite steps.

Cases 1 and 2.1 imply $\mu[f \geq \alpha^*] > 0$ which is not often satisfied.

Case 3. $u^* = 0$ and $\alpha^* = 0$ implies linear convergence as the following proposition shows.

Proposition 13. *Let $\alpha^* = u^* = 0$ and assume that p is the order of the zero α^* of F. Then $p > 1$ and u^* is a zero of F^c with the order $q > 1$, where $1/p + 1/q = 1$.*

Proof. Because of $s(0) = \alpha^* = 0$ and the bounded variation of F, m is strictly decreasing in some left-hand neighbourhood of $\alpha^* = 0$. Hence, by Proposition 7.2, s is continuous in some left-hand neighbourhood of $u^* = 0$. Therefore, F^c has in $u^* = 0$ a zero of higher order which implies the strict monotonicity of s, and in similar way the continuity of m in a neighbourhood of $\alpha^* = 0$. Thus, if p, q are the order of the associated zeros, there are some $\tau > 0$ and continuously differentiable functions $\sigma, \rho : (-\tau, 0] \to \mathbb{R}$, in 0 from the left-hand side, such that $F(\alpha) = |\alpha|^p \sigma(\alpha), \sigma(0) > 0$ and $F^c(u) = |u|^q \rho(u), \rho(0) > 0$. The use of the Young-Fenchel transformation $F(\alpha) + F^c(F'(\alpha)) = \alpha F'(\alpha)$ on the one hand and the elementary direct limit calculation on the

other hand yields $0 > (1-p)\,\sigma(0) = \lim_{\alpha\to+0}(F(\alpha) - \alpha F'(\alpha))\,/\,|\alpha|^p = -\lim_{\alpha\to+0} F^c(F'(\alpha))\,/\,|\alpha|^p = -\rho(0)\,(p\sigma(0))^q \lim_{\alpha\to+0} |\alpha|^{(p-1)q}\,/\,|\alpha|^p$. The existence of the last limit implies $p = (p-1)q$. ∎

The order p and q of the zeros implies the convergence rates $n\,/\,(n+p)$ and $n\,(p-1)\,/\,(n\,(p-1)+p)$ for Algorithm 2 and Algorithm 3, respectively. Hence Algorithm 2 is faster for $p \in (1,2)$ and Algorithm 3 for $p \in (2,\infty)$.

If we have $\alpha^* < 0$ then superlinear convergence can be achieved by Case 2. This situation can be arranged in the many practical problems. By using some geometrical considerations, the Algorithm 3 can be modified to a primal-dual method which can be implemented on base of Algorithm 1. Its theoretical development and the discussion of the test are the contents of the two next sections.

5 Acceleration by a primal-dual method (PDM)

We start with a geometrical interpretation of formula (6) in the Case 2, i.e. when $\alpha^* < 0$. With the notation $\bar{\alpha}_k := s\,(u_k)$ we get $u_{k+1} = u_k - F^c\,(u_k)\,/\bar{\alpha}_k$. Since $\bar{\alpha}_k = \min \partial F^c\,(u_k) \in \partial F^c\,(u_k)$ we obtain with Young-Fenchel's equality $F\,(\bar{\alpha}_k) + F^c\,(u_k) = \bar{\alpha}_k u_k$ immediately

$$u_{k+1} = \frac{F\,(\bar{\alpha}_k)}{\bar{\alpha}_k} = \frac{F\,(\bar{\alpha}_k) - F\,(0)}{\bar{\alpha}_k - 0}. \tag{7}$$

By the mean value theorem [6, Theorem 4.2.4], there is some $\alpha \in (\bar{\alpha}_k, 0)$ such that $u_{k+1} \in \partial F\,(\alpha)$. Since $u_{k+1} > u_k$, we can ensure by the convexity and monotonicity of F that our choice $\bar{\alpha}_{k+1} := \min\{\alpha\,|u_{k+1} \in \partial F\,(\alpha)\} = \min \partial F^c\,(u_{k+1}) = s\,(u_{k+1})$ belongs to $(\bar{\alpha}_k, 0)$. Let

$$\tilde{\alpha}_{k+1} := \bar{\alpha}_k + F\,(\bar{\alpha}_k)\,/m\,(\bar{\alpha}_k)$$

be the Newton iteration of $\bar{\alpha}_k$, called the primal step. Using $u_k \in \partial F\,(\bar{\alpha}_k)$ and the Ray Theorem, we get $|u_k|\,/1 = F\,(\bar{\alpha}_k)\,/\,(\tilde{\alpha}_{k+1} - \bar{\alpha}_k)$. From (7) it follows with $u_k < 0$ the update formula $u_{k+1} = (\tilde{\alpha}_{k+1} - \bar{\alpha}_k)\,|u_k|\,/\bar{\alpha}_k$. Hence,

$$u_{k+1} = (1 - \tilde{\alpha}_{k+1}/\bar{\alpha}_k)\,u_k$$

for the dual variables u_k. Beside the exceptional case when Algorithm 3 finishes after finite steps (i.e. $F\,(\bar{\alpha}_k) = 0$), we can reformulate the Algorithm 3 in the following way.

Algorithm 4. (**PDM**)

0. **Initialization:** Choose the tolerance $\varepsilon > 0, \alpha_0 < \alpha^*$ and put $u_0 = -m_+(\alpha_0)$.

k. **Iteration**

1. primal step: $\tilde{\alpha}_{k+1} = \bar{\alpha}_k + F(\bar{\alpha}_k)/m_+(\bar{\alpha}_k)$,

2. dual update: $\hat{u}_{k+1} = (1 - \tilde{\alpha}_{k+1}/\bar{\alpha}_k)\, u_k$,

3. primal update:

$$\begin{aligned}\hat{\alpha}_{k+1} &= \max\{\alpha \,|\hat{u}_{k+1} \in \partial F(\alpha) = [-m(\alpha), -m_+(\alpha)]\} \\ &= s_+(\hat{u}_{k+1}).\end{aligned}$$

If $\hat{\alpha}_{k+1} < \tilde{\alpha}_{k+1}$
then (a) $\bar{\alpha}_{k+1} = \tilde{\alpha}_{k+1}$ and $u_{k+1} = -m_+(\bar{\alpha}_{k+1})$
else (b) $\bar{\alpha}_{k+1} = \hat{\alpha}_{k+1}$ and $u_{k+1} = \hat{u}_{k+1}$.

4. Stopping rule: (see also next section)

If $|u_{k+1}| < \varepsilon$
then α_{k+1} is an approximation of $\operatorname{ess\,sup} f$.

We consider at first a simpler Algorithm 5 dropping the case $\hat{\alpha}_{k+1} < \tilde{\alpha}_{k+1}$ in Step k.3 from Algorithm 4, that means we put $\bar{\alpha}_{k+1} := \hat{\alpha}_{k+1}$ for all k.

Theorem 14. *Let sequences* $\{\bar{\alpha}_k\}$ *and* $\{u_k\}$ *be generated by Algorithm* 5. *Further assume that* f *is Lipschitzian over the convex compact set* $D \subset \mathbb{R}^n$ *with the Lipschitz-constant* L, $\alpha^* < 0$ *and* $\mu[f = \alpha] = 0$ *for all* $\alpha \in [\alpha_*, \alpha^*]$. *Then* $(\bar{\alpha}_k)$ *is superlinearly convergent to* α^* *satisfying the estimation*

$$m(\bar{\alpha}_{k+1}) < C\, m(\bar{\alpha}_k)^{\frac{n+1}{n}} \tag{8}$$

with $C = L/\left(|\alpha^*|\,\mu(D)^{\frac{1}{n}}\right) \sup_{x,y\in D} \|x - y\|$.

Remark 4. If f has a finite number of global maxima x^k and if f is continuously differentiable in some neighbourhood of these points and if we replace D by the shrinking sets $D_k \supset [f \geq \alpha_k]$ during the algorithm, a union of finite $n-$dimensional boxes around the points x_k, then the quotient $\frac{\sup_{x,y\in D_k}\|x-y\|}{\mu(D_k)^{\frac{1}{n}}}$ is asymptotically constant and $L_k = \sup_{x\in D_k} |\nabla f(x)|$ tends to zero for $k \to \infty$.

Remark 5. If f has a single global maximizer of order p then we obtain by using (3) the asymptotic estimation

$$(\alpha_* - \bar{\alpha}_{k+1}) < C^{\frac{p}{n}} \sqrt[n]{\frac{\sqrt[n]{c_{n,p}}}{R}} (\alpha_* - \bar{\alpha}_{k+1})^{1+\frac{1}{n}}$$

for $k \to \infty$.

Lemma 15 [12, Lemma 5.2]**.** *Let $D \subset \mathbb{R}^n$ be a convex compact set with the positive diameter $d := \max_{x,y \in D} \|x - y\|$. If $f : D \to \mathbb{R}$ is Lipschitzian with the constant L then*

$$m(\alpha) \geq \frac{\mu(D)}{(dL)^n} (\alpha^* - \alpha)^n . \tag{9}$$

Now we prove of Theorem 14 by using Lemma 15.

Proof. Instead of $\bar{\alpha}_k$ we use here α_k for simpler notation. The assumptions of Theorem 14 imply that F is strictly convex and continuously differentiable on $\mathbb{R}$. We obtain together with Lemma 7 that F^c is strictly convex and continuously differentiable on dom F^c, where the continuous differentiability at the boundary points of dom F^c is only one sided. Further, we have $-s = m^{-1}$ on $[-\mu(D), 0]$, i.e. $u = F'(\alpha)$ is equivalent to $\alpha = F^{c\prime}(u)$ for all $\alpha \in (\alpha_*, \alpha^*)$, all $u \in (-\mu(D), 0)$, $\alpha_* = F^{c\prime}(-\mu(D) + 0), \alpha^* = F^{c\prime}(u^* - 0), F'(\alpha_*) = -\mu(D)$ and $F'(\alpha^*) = u^* = 0$. Hence, Algorithm 3 is equivalent to Algorithm 5. With $-u = F'(\alpha)$ the inequality (9) can be transformed in dom F^c equivalently to $(F^{c\prime}(u^* - 0) - F^{c\prime}(u)) \leq dL\mu(D)^{-1/n} (u - u^*)^{1/n}$. Observing $F^{c'}(u) < F^{c\prime}(u^* - 0) = \alpha^* < 0$, the main step in Algorithm 3 can be reformulated to $u_{k+1} - u^* = u_k - F^c(u_k)/F^{c\prime}(u_k) - u_* = -(F^{c\prime}(u_k))^{-1} (F^c(u_k) + F^{c\prime}(u_k)(u_* - u_k))$. Using the mean-value theorem with integral remainder, we finally obtain

$$\begin{aligned}
&m(\alpha_{k+1}) = |F'(\alpha_{k+1}) - F'(\alpha^*)| = |u_{k+1} - u^*| \\
&\leq |F^{c\prime}(u_k)|^{-1} |F^c(u_k) + F^{c\prime}(u_k)(u_* - u_k)| \\
&< |\alpha^*|^{-1} \left| \int_0^1 [-F^{c\prime}(u_k + t(u^* - u_k)) + F^{c\prime}(u_k)] (u_* - u_k)\, dt \right| \\
&\leq |\alpha^*|^{-1} |u_* - u_k| \int_0^1 |F^{c\prime}(u^*) - F^{c\prime}(u_k)|\, dt \\
&\leq |\alpha^*|^{-1} dL\mu(D)^{-1/n} |u_k - u^*|^{1+1/n} \\
&= |\alpha^*|^{-1} dL\mu(D)^{-1/n} m(\alpha_k)^{1+1/n} .
\end{aligned}$$

■

Remark 6. If f has a unique global maximizer of order p like (3) then we get the asymptotic estimation

$$\left(F^{c\prime}\left(u^*-0\right)-F^{c\prime}\left(u\right)\right)\leq\left(c_{n,p}\right)^{-p/n}R_Q\left(u-u^*\right)^{p/n}$$

under the same assumptions for f, which analogously yields

$$m\left(\alpha_{k+1}\right)\leq\left|\alpha^*\right|^{-1}R_Q\left(c_{n,p}\right)^{-p/n}m\left(\alpha_k\right)^{1+p/n}.$$

We have here a quantified description of the behavior of the Lipschitz constant L_k mentioned before.

Remark 7. Algorithm 4 is faster than Algorithm 5 since the sequences $\{\bar{\alpha}_k\}$ and $\{u_k\}$ are strictly increasing. If Algorithm 3 is not superlinearly convergent then there is a finite k_0 in Algorithm 4 such that $\bar{\alpha}_{k+1}=\tilde{\alpha}_{k+1}$ at most for $k\leq k_0$ and $\bar{\alpha}_{k+1}=\hat{\alpha}_{k+1}$ for all $k>k_0$, i.e. after k_0 steps Algorithm 4 and Algorithm 5 run identically by using the starting level $\bar{\alpha}_{k_0}$. Algorithm 4 should be preferred if the iteration number k can be expected to be small.

6 Implementation of PDM and test results

The **PDM** has three properties:

1. The number of function evaluations in each iteration is proportional to the ratio $q_k:=|\bar{\alpha}_k-\alpha^*|\ /\ |\bar{\alpha}_{k+1}-\alpha^*|$. That means, speeding up of the level sequence increases the number of function evaluations at each iteration.

2. Because of the superlinear convergence of the level set sequence, we only need a relatively small number of iterations to reach α^* within a prescribed tolerance.

3. **PDM** runs different from **ZM** only if $\tilde{\alpha}_{k+1}<\bar{\alpha}_{k+1}$. Then we have in each iteration the old level $\bar{\alpha}_k$ and two new levels $\tilde{\alpha}_{k+1}$ and $\bar{\alpha}_{k+1}$.

In **ZM** the ratio $|\alpha_k-\alpha^*|\ /\ |\alpha_{k+1}-\alpha^*|$ is nearly constant and we have the old level α_k and the new level α_{k+1} at each iteration.

In order to benefit from property 2 as much as possible, and to minimize the disadvantage of property 1, we choose the following strategy.

First: we do not take the level box with respect to $\bar{\alpha}_k$ but the level box with respect to $\tilde{\alpha}_{k+1}$ for determining points x with $f\left(x\right)\geq\bar{\alpha}_{k+1}$.

Second: the first level $\tilde{\alpha}_{k+1}$ is theoretically obtained by the Zheng-step using the mean value m_k of the t (already ordered) function values $\bar{\alpha}_k \leq f\left(x_1^k\right) \leq f\left(x_2^k\right) \leq \cdots \leq f\left(x_t^k\right)$. In the implemented algorithm we require a uniform lower bound of the number of points x_k belonging to the box $D\left(x_s^k, ..., x_t^k\right), s < t, \tilde{\alpha}_{k+1} \leq f(x_s)$. Hence,

$$\tilde{\alpha}_{k+1} = \min\left\{f\left(x_{\tilde{t}}\right), m_k\right\}, \tilde{t} = \left\lfloor \frac{t}{3} \right\rfloor.$$

Third: the second level $\hat{\alpha}_{k+1}$ is theoretically given by the dual and primal update. Again we require a uniform lower bound of the number of old points x_k belonging to the associated level box. We determine $\hat{\alpha}'_{k+1} = f\left(x_\rho^k\right)$ (theoretical level) looking for a ρ with $-\mu\left(D\left(x_{\rho+1}^k, ..., x_t^k\right)\right) \leq u_{k+1} \leq -\mu\left(D\left(x_\rho^k, ..., x_t^k\right)\right)$ by using a bisection with respect to the index ρ. Hence,

$$\hat{\alpha}_{k+1} = \min\left\{f\left(x_{\hat{t}}\right), f\left(x_\rho^k\right)\right\}, \hat{t} = \left\lfloor \frac{4t}{5} \right\rfloor.$$

Naturally, the last uniform bound destroys the superlinear behavior of $\{\hat{\alpha}_k\}$, i.e. there is some $q > 0$ with $q_k > q$ for all k. However, both uniform bounds become active in the last two or three iterations before the termination. So far we do not know the optimal choice of these bounds. Without these bounds the **PDM** can run much worse than **ZM**, often in this case, the last iteration before the termination costs too much function evaluations since the quotient of measures of the new and old level box becomes too small.

We use (see appendix) the standard test functions for global minimization (see e.g. [1] for discontinuous functions, [2, 8, 11]) of dimensions 2, 3, 4 and 5 which have only one global minimizer in some chosen box.

Each of the following 16 test problems is computed 20 times with both methods for $t = 20, 30, 40$ and 60. The pseudo random number generator is initialized by the clock. We evaluate the quality of each test by the numbers 1 to 6. $k = 1, 2, ...5$ stands for an absolute distance of the calculated approximation to the exact optimal solution of at most 10^{-5+k} and number 6 in any other case. Further, we distinguished for each run the termination criteria like "variance $\leq tol$", "number of function evaluations $\geq M$", "Number of iterations $\geq m$", "$\bar{\alpha}_{k+1} + tol \geq f(x_t)$" and "level set measure $\leq tol^{\dim/2}$". We choose $tol = 10^{-5}$, $M = 2000 \cdot \dim \cdot t/20$ and $m = 10 * \dim$. We use the variant that at most $1/5$ of the arguments with the best values can survive for the next iteration whenever they belong to the new level set.

The procedure is organized in a way that termination occurs nearly immediately when the function evaluation number M is reached or after the iteration whenever one of the above criteria is satisfied. If the last iteration is finished by a break then dates of the previous iteration are used combined with values of the interrupted iteration which belong to the new level.

The following table contains for $t = 30$ the values of $fval$, $\sigma fval$ and *quality* which denote the mean value over 20 runs for the number of function evaluations, the statistical standard deviation of the function evaluations, and the quality of the approximation of the optimal solution (best $= 1, \ldots, 6 =$ worst, see above), respectively.

The tests show that the number of function evaluations is in the average one and a half times to twice larger for the **ZM** as in the proposed **PDM**. However, some essential loss of probability for finding the global minimizer could not be noticed. Beside the number of function evaluations both algorithms principally show the same behavior.

Case 1. If the *global minimizer belongs to the interior of the first box far enough from the boundary,* then both algorithms find the global minimizer in all tests, however **PDM** needs in the average significantly less ($\frac{1}{3}$ till $\frac{1}{2}$) numbers of function evaluations (acceleration effect). The additional bisection in **PDM** can be neglected with respect to the time for evaluating the function values. The natural conjecture, that **PDM** is less sure in finding the global minimizer, could not be stated in all tests. The termination was mostly caused in the **PDM** by "variance $\leq tol$" or "level set measure $\leq tol^{\dim/2}$", whereas the **ZM** terminates more often by "number of function evaluations $\geq M$". Thus **PDM** yields some time with smaller efforts more exact results. A lot of local minimizers does not influence the convergence if Case 3 (see below) can be avoided.

If we compare the best run of **ZM** and the worst run of **PDM** for the same example then the number of function evaluations was also here smaller for the primal-dual modification. The variance with respect to the function evaluations is for the **ZM** significantly smaller. Both properties together imply that we are faster in each case with **PDM** but we can be much more faster with large probability. However, with the original **ZM** we are in each case slow and the probability to be a little bit faster is small.

Case 2. If the *global minimizer is on the boundary of the first box* then both algorithms cut the global minimum and do not reach it, since the box sequence is nearly nested. The acceleration effect remains true.

Test $t = 30$		$fval$		$\sigma fval$ / $fval$		Quality(1..6)		Term. ZM/PDM		
Ex.	dim	$\frac{\text{ZM}}{\text{PDM}}$	$\frac{PDM}{\text{dim}}$	ZM	PDM	ZM	Z–PD	Var	$fval$	$meas$
good										
1	2	1.7	240	0.11	0.17	2.90	0.05	20/20		
4	4	2.0	1397	0.18	0.65	4.60	0.75	03/16	17/04	
5	3	1.8	195	0.11	0.15	3.55	0.05	20/20		
10	3	1.8	191	0.06	0.10	3.90	–0.05	20/20		
11	2	2.1	1217	0.12	0.23	3.80	0.10	19/20	01/00	
13	2	1.9	330	0.07	0.11	2.50	–0.85	19/20		01/00
14	2	1.8	207	0.08	0.15	2.55	–0.50			20/20
15	2	1.6	1077	0.10	0.21	3.40	0.35	19/20	01/00	
mean										
3	2	1.3	509	0.09	0.12	2.00	0.00	20/20		
8	2	1.5	295	0.06	0.09	2.95	–0.15	20/20		
9	5	1.4	1905	0.16	0.50	4.65	0.90	01/00	09/07	10/13
12	2	1.5	266	0.08	0.14	3.60	0.05	20/20		
16	2	1.4	405	0.07	0.12	2.05	–0.25			20/20
ill										
2	2	1.0	3073	0.01	0.02	5.50	–0.05		20/20	
6	2	1.1	2724	0.01	0.22	3.00	0.35	00/11	20/09	
7	2	1.1	2588	0.09	0.21	3.80	0.35	04/14	16/06	

Case 3. If the *global minimizer is in a small neighbourhood of the boundary of the first box and some local minimizers with nearly the same function value are far away* then both algorithms work very slowly. By chance, one of the local (global) minimizers is approximated since the box cannot shrink to zero within a reasonable number of function evaluations, i.e. the termination is almost always given by "number of function evaluations $\geq M$". Thus the exactness is more or less stochastic. An acceleration effect cannot be noticed.

7 Conclusions

The **ZM** and **PDM** algorithms are suitable for solving global optimization problems also in those cases when the problem functions are discontinuous. Both methods produce approximations of upper level sets with increasing level during the running of the algorithm. Thus, they are suitable for determining all local maxima within a certain small distance to the global

maximum. This property is essential, e.g. for the use of these methods in semi-infinite programming. The speeding up of **ZM** by **PDM** without essential loss of information makes integral methods more attractively.

Higher dimensional problems ($6 \leq \dim \leq 100$) and the test Cases 2 and 3 can be successfully solved whenever the above methods are combined with branch and bound strategies similar to interval methods (see, e.g. the software BARLO [3]).

References

[1] Soo Hong Chew and Quan Zheng, Integral Global Optimization, Springer, Berlin 1988.

[2] C.A. Floudas and P.M. Pardalos, A Collection of Test Problems for Constrained Global Optimization Algorithms, Springer, Berlin 1990.

[3] J. Hichert, BARLO — Eine Software zur Wesentlichen Globalen Optimierung, Technical Report 17/98, Technische Universität Ilmenau 1998.

[4] J. Hichert, Methoden zur Bestimmung des wesentlichen Supremums mit Anwendung in der globalen Optimierung, PhD thesis, Technische Universität Ilmenau 1999.

[5] J. Hichert, A. Hoffmann, and H.X. Phu, *Convergence speed of an integral method for computing the essential supremum*, in: I.M. Bomze, T. Csendes, R. Horst and P.M. Pardalos (eds.): Developments in Global Optimization, pp. 153–170, Kluwer Academic Publishers, Dordrecht 1997.

[6] J.-B. Hiriart-Urruty and C. Lemaréchal, Convex Analysis and Minimization Algorithms, Vol. I (Fundamentals), Springer, Berlin 1993.

[7] J.-B. Hiriart-Urruty and C. Lemaréchal, Convex Analysis and Minimization Algorithms, Vol. II (Advanced Theory and Bundle Methods), Springer, Berlin 1993.

[8] http://solon.cma.univie.ac.at/∼neum/glopt.

[9] P.M. Pardalos, *An open global optimization problem on the unit sphere*, Journal of Global Optimization **6** (1995), 213.

[10] H.X. Phú and A. Hoffmann, *Essential supremum and supremum of summable functions*, Numerical Functional Analysis and Optimization **17** (1,2) (1996), 167–180.

[11] A. Törn and A. Žilinskas, Global Optimization, Springer, Berlin 1989.

[12] Z.B. Zabinsky and R.L. Smith, *Pure adaptive search in global optimization*, Mathematical Programming **53** (1992), 323–338.

[13] Quan Zheng, Integral Global Optimization of Robust Discontinuous Functions, PhD thesis, The Graduate School of Clemson University, Clemson 1992.

[14] Quan Zheng and Deming Zhuang, *Integral global minimization: Algorithms, implementations and numerical tests*, Journal of Global Optimization **7** (1995), 421–454.

Received 10 February 2000
Revised 19 May 2000

A Appendix

No.	Function or citation	First box / name	Solution
1	$4x^2 - 2x^4 + \frac{1}{3}x^6 + xy - 4y^2 + 4y^4$	$[-2,2] \times [-2,0.5]$ 6-hump-camel-back	$x \approx 0.0898$ $y \approx -0.7127$ $\alpha^* \approx -1.0316$
2	$\left(y - \frac{5.1}{4\pi^2}x^2 + \frac{5}{\pi}x - 6\right)^2 + \frac{40\pi-5}{4\pi}\cos(x) + 10$	$[-5,10] \times [2,15]$ Branin	$x \approx 3.1416$ $y \approx 12.1250$ $\alpha^* \approx 0.3979$
3	see, e.g. [11]	$[-2,2]^2$ Goldstein-Price	$(x\,,y) = (0,-1)$ $\alpha^* = -3$
4	see, e.g. [11]	$[0,10]^4$ Shekel	$\mathbf{x} \approx (4,4,4,4)$ $\alpha^* \approx -10.5364$
5	see, e.g. [11]	$[0,1]^3$ Hartmann	$w \approx 0.1146$ $x \approx 0.5556$ $y \approx 0.8525$ $\alpha^* \approx -3.863$
6	see, e.g. [11]	$[-1,1]^2$ Rastrigin	$\mathbf{x} = (0,0)$ $\alpha^* = -2$
7	$100\left(y - x^2\right)^2 + (x-1)^2$	$[-1.5,1.5] \times [-0.5,2.5]$ Rosenbrock - banana	$\mathbf{x} = (1,1)$ $\alpha^* = 0$
8	$2\left(x^2+y^2\right) - \left\lfloor x^2+y^2 \right\rfloor$, see, e.g. [1]	$[-10,10]^2$ piecewise continuous	$\mathbf{x} = (0,0)$ $\alpha^* = 0$
9	$z := \sqrt{\sum_{k=1}^{5} \lvert x_i \rvert}$, see, e.g. [1] $1 + z + \operatorname{sgn}\left(\sin\left(\frac{1}{z} - 0.5\right)\right)$	$[-1,1]^5$ unbounded variation	$\mathbf{x} = (0,0)$ $\alpha^* = 0$
10	see, e.g. [9]	$[0,\pi] \times [0,2\pi] \times [0,\pi]$ Fekete points	$w \approx 2.0944$ $x \approx 3.1416$ $y \approx 2.0944$ $\alpha^* \approx -3.862$
11	see, e.g. [11]	$[-100,100]^2$ Griewank function	$\mathbf{x} = (0,0)$ $\alpha^* = 0$
12	$x^2 - 12x + 11 + 10\cos\left(\frac{\pi}{2}x\right) + 8\sin(\pi x) - \frac{1}{\sqrt{2\pi}}e^{-(y-0.5)^2}$	$[0,7] \times [0,5]$ Chinchinadze	$x \approx 5.6235$ $y \approx 0.5000$ $\alpha^* \approx -40.96$
13	$x^2 + y^2$	$[1,2]^2$	$\mathbf{x} = (1,1)$ $\alpha^* = 2$
14	practical example (Ilmenau) dispersion function I	$[0.8,1.4] \times \left[10^{-10}, 10^{-1}\right]$ $\infty -$ pol at argmin	$x \approx 1.0166$ $y \approx 0.00465$ $\alpha^* \approx -40.964$
15	$\sin(x) + 0.2\cos(10y) + 0.4\cos(5x) + (y-5)^2\,10^{-2}$	$[0,10]^2$ Hichert	$x \approx 4.4265$ $y \approx 4.7127$ $\alpha^* \approx -1.5546$
16	practical example (Ilmenau) dispersion function II	$[1,5] \times [0,0.1]$ very thin pik	$x \approx 1.0296$ $y = 0$ $\alpha^* \approx 0.00170$