OrthoMADS: a deterministic MADS instance with orthogonal directions

Article *in* SIAM Journal on Optimization · January 2009

DOI: 10.1137/080716980 · Source: DBLP

4 authors:

Mark A. Abramson
Utah Valley University

Charles Audet
Polytechnique Montréal

J. E. Dennis
Rice University

Sébastien Le Digabel
Polytechnique Montréal

SIAM J. OPTIM.
Vol. 20, No. 2, pp. 948–966

ORTHOMADS: A DETERMINISTIC MADS INSTANCE WITH ORTHOGONAL DIRECTIONS*

MARK A. ABRAMSON†, CHARLES AUDET‡, J. E. DENNIS, JR.§, AND SÉBASTIEN LE DIGABEL‡

Abstract. The purpose of this paper is to introduce a new way of choosing directions for the mesh adaptive direct search (MADS) class of algorithms. The advantages of this new ORTHOMADS instantiation of MADS are that the polling directions are chosen deterministically, ensuring that the results of a given run are repeatable, and that they are orthogonal to each other, which yields convex cones of missed directions at each iteration that are minimal in a reasonable measure. Convergence results for ORTHOMADS follow directly from those already published for MADS, and they hold deterministically, rather than with probability one, as is the case for LTMADS, the first MADS instance. The initial numerical results are quite good for both smooth and nonsmooth and constrained and unconstrained problems considered here.

Key words. mesh adaptive direct search (MADS) algorithms, deterministic, orthogonal directions, constrained optimization, nonlinear programming

AMS subject classifications. 90C56, 90C30, 65K05

DOI. 10.1137/080716980

1. Introduction.

This paper considers optimization problems of the form

$$\min_{x \in \Omega} f(x),$$

where $f : \Omega \subset \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is typically evaluated through a black-box computer simulation with no derivatives available, and Ω is a set of feasible points also defined by black-box nonlinear, or even Boolean, constraint functions. Because no exploitable information on the nature of f or Ω exists, we consider direct search methods which only use functions evaluations to drive their search.

Mesh adaptive direct search (MADS) is introduced in [6] as a search/poll direct search class of methods with a hierarchical convergence analysis based on local smoothness of the functions defining the problem. It extends the generalized pattern search (GPS) method of [28]. The constraints are treated by the extreme barrier approach, which simply rejects points outside Ω by setting their objective function value to ∞. The first instance of this class of methods is called LTMADS [6]—the prefix LT stands for *lower triangular*, since a random lower triangular matrix is used to construct exploration directions.

*Received by the editors February 27, 2008; accepted for publication (in revised form) April 2, 2009; published electronically July 3, 2009. The first three authors were supported by AFOSR FA9550-07-1-0302, the Boeing Company, and ExxonMobil Upstream Research Company.
http://www.siam.org/journals/siopt/20-2/71698.html

†The Boeing Company, P.O. Box 3707 MC 7L-21, Seattle, WA 98124-2207 (Mark.A.Abramson@boeing.com, http://www.gerad.ca/NOMAD/Abramson/abramson.html).

‡GERAD and Département de Mathématiques et de Génie Industriel, École Polytechnique de Montréal, C.P. 6079, Succ. Centre-ville, Montréal, Québec H3C 3A7, Canada (Charles.Audet@gerad.ca, http://www.gerad.ca/Charles.Audet, Sebastien.Le.Digabel@gerad.ca, http://www.gerad.ca/Sebastien.Le.Digabel). The work of the second author was supported by FCAR grant NC72792 and NSERC grant 239436-05.

§Computational and Applied Mathematics Department, Rice University - MS 134, 6100 South Main Street, Houston, TX 77005-1892 (dennis@caam.rice.edu, http://www.caam.rice.edu/~dennis). This author was supported by LANL 94895-001-04 34.

The computational results in [3, 6, 9, 10] confirm that LtMads can solve difficult problems efficiently, but it has drawbacks that we wish to correct in this paper. First, there is a probabilistic component to the choice of polling directions. For each new mesh size, a direction is randomly generated. That direction is completed as described in [6] to a positive spanning set of directions from the current iterate to other current mesh points. The resulting algorithm is shown to have Clarke stationary point convergence with probability one. However, it has been observed [17] that this way of choosing polling directions can lead to undesirably large angles between some of the LtMads poll directions at a given iteration.

The purpose of this paper is to introduce a new variant of Mads, which we call OrthoMads, that uses an orthogonal positive spanning set of polling directions and thus avoids large angles between the $2n$ directions. In Figure 3, we show some experiments in which the normalized (i.e., rescaled to have unit norm) OrthoMads directions cover the surface of the unit sphere more densely and evenly than those of LtMads. The ideas of capturing a rich set of directions in a generalized pattern search context and avoiding large angles were studied by Alberto et al. [2]. They use $n + 1$ polling directions, such that the angle between any two of them is constant. The directions are rotated at every iteration to introduce new directions. The main difference between what they do and our work is that our construction also ensures that the union of normalized directions over all iterations is dense on the unit sphere.

We show that OrthoMads shares the same theoretical convergence results as LtMads, except that the convergence is not qualified by being of probability one. In the numerical tests given here, OrthoMads is able to solve more problems with fewer function evaluations than LtMads.

The paper is divided as follows. Section 2 briefly presents Mads and LtMads. OrthoMads is then detailed in section 3, where we propose a deterministic way to construct a polling set on the current mesh of orthogonal polling directions (the OrthoMads directions). Section 3 also gives the convergence results, based on those in [6]. Finally, we present numerical results in section 4 and some concluding remarks in section 5.

Notation. Throughout the text, $\|\cdot\|$ denotes the ℓ_2 norm, $e_i \in \mathbb{R}^n$ is the ith coordinate vector, $e \in \mathbb{R}^n$ is the vector whose components are all equal to 1, and $\mathbb{N} = \{0, 1, 2, \ldots\}$.

2. The Mads algorithm and its LtMads instantiation.

This section contains a brief description of the Mads class of algorithms as well as its LtMads practical instantiation. For a complete presentation, the reader is invited to consult [6]. OrthoMads, presented in section 3, is a second instantiation of Mads.

A Mads algorithm is an iterative method where each iteration k is separated into two steps, the *search* and the *poll*, in which the objective function f and the test for feasibility is evaluated at finitely many trial points. The goal of each iteration is to generate a point in Ω whose objective function value is smaller than $f(x_k)$, where x_k is the current best feasible solution, called the *incumbent*.

The trial points lie on the mesh M_k defined by

$$M_k \;=\; \{x + \Delta_k^m Dz \;:\; x \in V_k,\; z \in \mathbb{N}^{n_D}\} \subset \mathbb{R}^n,$$

where $V_k \subset \mathbb{R}^n$ is the set of all evaluated points by the start of iteration k, $\Delta_k^m \in \mathbb{R}_+$ is the *mesh size parameter* at iteration k, and D is a matrix in $\mathbb{R}^{n\times n_D}$ composed of n_D directions in $\mathbb{R}^n$. More precisely, D is constant throughout all iterations, and it is the product of a nonsingular $n \times n$ real matrix and an $n \times n_D$ integer matrix whose columns form a positive spanning set. For the LtMads and OrthoMads instantiations of Mads, the matrix D is simply defined as $D = [I_n \ -I_n]$, where I_n is the identity matrix in dimension n.

The search step generates finitely many trial points on the mesh M_k. To do so, it may apply user knowledge about the problem coupled with inexpensive surrogates for the objective function or constraints. The surrogate functions are typically evaluated at several mesh points, and the expensive functions are then evaluated only at the most promising trial points [11]. This paper focuses on the poll step, which is more rigidly defined and is characterized by the set of *poll trial points* around the incumbent x_k,

$$P_k \ = \ \{x_k + \Delta_k^m d : d \in D_k\} \ \subset \ M_k,$$

where D_k is the set of poll directions. Each column of D_k is formed by taking integer combinations of the columns of D so that the columns of D_k form a positive spanning set. A second scalar parameter, Δ_k^p, called the *poll size parameter*, is used to bound the distance from the incumbent x_k to the poll trial points.

In Gps, the directions in D_k are always chosen from the columns of D, and $\Delta_k^p = \Delta_k^m$ for all k. Therefore, in Gps, there is only the same finite set of possibilities for selecting the directions in every D_k.

Both the LtMads and OrthoMads instantiations of Mads use an integer ℓ_k, called the *mesh index*, to update the poll and mesh size parameters. They are linked in this paper as follows:

$$\Delta_k^p = 2^{-\ell_k} \quad \text{and} \quad \Delta_k^m = \min\{1, 4^{-\ell_k}\}. \tag{1}$$

At the initial iteration, $k = 0$, $\ell_0 = 0$, and $\Delta_0^m = \Delta_0^p = 1$. At each iteration k, the mesh and poll size parameters satisfy $\Delta_k^m \le \Delta_k^p$ and $\Delta_k^m 2^{|\ell_k|} = \Delta_k^p$. If no new incumbent has been found in the search or poll steps, the iteration is said to be unsuccessful and $\ell_{k+1} \leftarrow \ell_k + 1$ (Δ_k^m and Δ_k^p are reduced). Otherwise, the iteration is a success and $\ell_{k+1} \leftarrow \ell_k - 1$ (Δ_k^m and Δ_k^p are increased). With (1), after an unsuccessful iteration, Δ_k^m is reduced faster than Δ_k^p, so that the number of possible poll trial points increases, allowing more flexibility in the choice of the poll directions D_k.

In LtMads, however, the poll directions are randomly generated: A random integer nonsingular lower triangular matrix (thus the Lt) is constructed randomly, and then its rows are randomly permuted. The columns of the resulting matrix B_k form a basis of $\mathbb{R}^n$, and they are then completed to a minimal positive basis $D_k = [B_k \ b]$ (where b is the negative of the sum of the n columns of B_k) or a maximal positive basis $D_k = [B_k \ -B_k]$. The poll directions are composed of the directions of the positive basis. The set of normalized directions in the positive bases D_k becomes dense on the unit sphere with probability one as k goes to infinity.

We conclude this basic introduction to Mads by mentioning that the convergence analysis presented in [6], based on the Clarke calculus [12], gives hierarchical results based on the local differentiability of f and on a constraint qualification by Rockafellar [26]. The convergence results of the new instantiation OrthoMads will be identical, except that they hold without any probabilistic arguments. We refer the reader to [6] for details.

3. The OrthoMads instantiation of Mads. The difference between LtMads and OrthoMads lies in the way that directions in D_k are generated: With LtMads, D_k is randomly generated and directions are not necessarily orthogonal, possibly leading to large angles between directions and large unexplored convex cones of directions at a given step. However, the union of all normalized LtMads directions over all iterations k is dense on the unit sphere with probability one.

OrthoMads introduces a new deterministic way to generate the poll directions D_k. This new method generates an orthogonal basis. The maximal positive spanning set is $D_k = [H_k \ \ -H_k]$, where the columns H_k form an orthogonal basis of $\mathbb{R}^n$. The columns of D_k forming the OrthoMads directions are said to be orthogonal because any pair of noncolinear directions are orthogonal. Furthermore, the union of all normalized OrthoMads directions over all iterations is dense on the unit sphere. In addition, the directions of D_k are integer, so that poll points lie on the mesh defined with $D = [I_n \ \ -I_n]$.

The advantage of determinism is that numerical results are now reproducible. In [6] numerical experiments are performed on a series of several LtMads runs to illustrate the variations in the results due to the random component.

The orthogonality of the OrthoMads directions D_k gives a distribution of the poll trial points in the search space that is better in the sense of the cosine measure [13, 20]:

$$\mathrm{cm}(D_k) = \min_{0 \neq v \in \mathbb{R}^n} \max_{d \in D_k} \frac{v^T d}{\|v\|\|d\|}. \tag{2}$$

For a given positive basis D_k and nonzero vector v, the value $\max_{d \in D_k} \frac{v^T d}{\|v\|\|d\|}$ is the cosine of the smallest angle between v and the any column of D. The cosine measure returns the least of the values over all nonzero directions v. A large cosine measure is desired in order to reduce as much as possible the cones of unexplored directions. Because the OrthoMads poll directions are orthogonal, the cosine measure simplifies to $\mathrm{cm}(D_k) = 1/\sqrt{n}$, which is the largest possible value among all possible positive bases.

At each iteration of OrthoMads, the main steps for the construction of these directions are as follows. First, the quasi-random Halton sequence [18] produces a vector in $[0, 1]^n$ (section 3.1). Second, this vector is scaled and rounded to an appropriate length (section 3.2). The resulting direction is called the *adjusted Halton direction*. Third, the Householder transformation is applied to the adjusted Halton direction, producing n orthogonal and integer vectors, forming a basis for $\mathbb{R}^n$ (section 3.3). Finally, the basis is completed to a positive basis formed by $2n$ OrthoMads poll directions D_k, by including in D_k the basis and its negatives (section 3.4). Figure 1 summarizes these steps, and will be referred to throughout this section.

In this section, we also show that the OrthoMads directions meet all the conditions detailed in [4, 6], so that OrthoMads is a valid Mads instance and thus inherits all of its convergence properties.

3.1. The Halton sequence $\boldsymbol{u_t}$. Halton [18] introduced a deterministic family of sequences that grow dense in the hypercube $[0, 1]^n$. We consider the simplest sequence of this family, whose tth element is

$$u_t \ = \ (u_{t,p_1}, u_{t,p_2}, \ldots, u_{t,p_n})^T \in [0, 1]^n,$$

where $p_1 = 2, p_2 = 3, p_3 = 5$, p_j is the jth prime number, and $u_{t,p}$ is the radical-inverse function in base p. More precisely,

$$u_{t,p} \ = \ \sum_{r=0}^{\infty} \frac{a_{t,r,p}}{p^{1+r}},$$

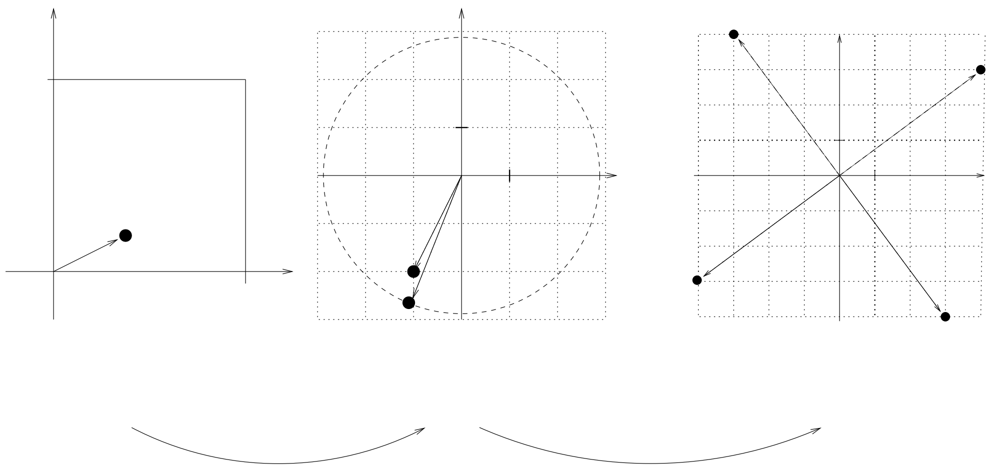

FIG. 1. *Example with $n = 2$ and $(t, \ell) = (6, 3)$. The Halton direction is $u_t = (3/8, 2/9)^T$, the adjusted Halton direction $q_{t,\ell} = (-1, -2)^T$ with $\alpha_{t,\ell} = 2$, and the set of poll directions $D_k = [H_{t,\ell}\ \ -H_{t,\ell}]$ with $H_{t,\ell}e_1 = (3, -4)^T$ and $H_{t,\ell}e_2 = (-4, -3)^T$. Every poll direction $d \in D_k$ satisfies $\Delta_k^m \|d\| = 5/64 < \Delta_k^p = 1/8$.*

where the $a_{t,r,p} \in \mathbb{Z}_+$ are the unique coefficients of the base p expansion of t:

$$t \ = \ \sum_{r=0}^{\infty} a_{t,r,p} p^r.$$

Table 1 describes the first five elements of u_t for $n = 4$. Our specific sequence of u_t vectors is from this point addressed as the sequence of *Halton directions*, and the integer t as the *Halton index*.

TABLE 1
The sequence of Halton directions for $n = 4$ and Halton indexes $t = 0, 1, \dots, 7$. For example, $u_{5,3}$ is obtained with $1 \times 3^{-2} + 2 \times 3^{-1} = 7/9$.

t	t in base				u_t			
	2	3	5	7	$u_{t,2}$	$u_{t,3}$	$u_{t,5}$	$u_{t,7}$
0	0	0	0	0	0	0	0	0
1	1	1	1	1	1/2	1/3	1/5	1/7
2	10	2	2	2	1/4	2/3	2/5	2/7
3	11	10	3	3	3/4	1/9	3/5	3/7
4	100	11	4	4	1/8	4/9	4/5	4/7
5	101	12	10	5	5/8	7/9	1/25	5/7
6	110	20	11	6	3/8	2/9	6/25	6/7
7	111	21	12	10	7/8	5/9	11/25	1/49

The linear correlation between the last columns of u_t for small values of t can be removed by excluding the initial points of the Halton sequence [23]. In the present work, we start the sequence at the nth prime number (p_n), which we denote t_0 and call the *Halton seed*.

The following properties will be used in section 3.2:

$$2u_t - e = 0 \Leftrightarrow n = t = 1, \tag{3}$$

$$|2u_{t,p_i} - 1| = |2u_{t,p_j} - 1| \Leftrightarrow t = 0. \tag{4}$$

Property (3) is trivial to show, and property (4) follows from the fact that u_{t,p_i} and u_{t,p_j} can be written as reduced fractions with denominators that are powers of different prime numbers p_i and p_j.

The next result shows that the union of all the directions in the sequence of Halton is dense in $[0,1]^n$, i.e., any direction $v \in [0,1]^n$ is an accumulation point of the sequence $\{u_t\}_{t=1}^{\infty}$.

PROPOSITION 3.1. *The Halton sequence* $\{u_t\}_{t=1}^{\infty}$ *is dense in the hypercube* $[0,1]^n$.

Proof. It suffices to show that for any vector $v \in [0,1]^n$ and any $\varepsilon > 0$, there exists an integer t such that $\|u_t - v\| < \varepsilon$. A construction of such an integer t involves solving a system of n Diophantine equations, and existence of a solution is ensured by the Chinese remainder theorem [15] and by the fact that prime numbers are used in the definition of u_t. We refer the reader to [18] for a detailed proof.

3.2. The adjusted Halton direction $q_{t,\ell}$. The directions in D_k used in the poll step of MADS cannot be chosen arbitrarily; they must satisfy precise requirements. The Halton directions u_t do not satisfy these requirements and the first steps toward generating a satisfactory set D_k are to translate, scale, and round u_t. These operations depend on the mesh index ℓ_k (or ℓ to simplify notation), which is related to the mesh size parameter Δ_k^m (1). The mesh index ℓ is used to transform the direction u_t into the *adjusted Halton direction* $q_{t,\ell} \in \mathbb{Z}^n$, a direction whose norm is close to $2^{|\ell|/2}$. Furthermore, the normalized direction $\frac{q_{t,\ell}}{\|q_{t,\ell}\|}$ will be constructed so that it is close to $\frac{2u_t - e}{\|2u_t - e\|}$. We already observed in (3) that $2u_t - e = 0$ is possible only if $n = 1$ and $t = 1$, and our algorithm never uses $t = 1$ (we begin our Halton sequence at the nth prime number, which is strictly larger than 1).

In order to define $q_{t,\ell}$, we first introduce the following sequence of functions based on the tth Halton direction u_t:

$$q_t(\alpha) \;=\; \operatorname{round}\left(\alpha \frac{2u_t - e}{\|2u_t - e\|}\right) \;\in\; \mathbb{Z}^n \cap \left[-\alpha - \frac{1}{2}, \alpha + \frac{1}{2}\right]^n,$$

where round refers to the rounding up operation (with round(0.5) = 1 and round(−0.5) = −1) and where $\alpha \in \mathbb{R}_+$ is a scaling factor. The function $\|q_t(\cdot)\|$ is a monotone nondecreasing step function on $\mathbb{R}_+$. Let $\alpha_{t,\ell}$ be a scalar such that $\|q_t(\alpha_{t,\ell})\|$ is as close as possible to $2^{|\ell|/2}$ without exceeding it:

$$\alpha_{t,\ell} \;\in\; \begin{array}{ll} \underset{\alpha \in \mathbb{R}_+}{\operatorname{argmax}} & \|q_t(\alpha)\| \\ \text{s.t.} & \|q_t(\alpha)\| \le 2^{|\ell|/2}. \end{array} \tag{5}$$

The following lemma gives a lower bound on an optimal solution of this optimization problem.

LEMMA 3.2. *There exists an optimal solution of problem* (5) *satisfying* $\alpha_{t,\ell} \ge \frac{2^{|\ell|/2}}{\sqrt{n}} - \frac{1}{2}$.

Proof. Every feasible solution α to problem (5) satisfies

$$\|q_t(\alpha)\|^2 \;=\; \sum_{i=1}^{n} \operatorname{round}\left(\frac{\alpha(2u_t^i - 1)}{\|2u_t - e\|}\right)^2 \;\le\; \sum_{i=1}^{n}\left(\alpha + \frac{1}{2}\right)^2 \;=\; n\left(\alpha + \frac{1}{2}\right)^2,$$

where u_t^i denotes the ith element of u_t. Define $\beta = \frac{2^{|\ell|/2}}{\sqrt{n}} - \frac{1}{2}$. Then β is feasible for problem (5), since $\|q_t(\beta)\|^2 \le n(\beta + \frac{1}{2})^2 = 2^{|\ell|}$. Combining this result with the fact that $\|q_t(\alpha)\|$ is a monotone nondecreasing step function with respect to $\alpha \in \mathbb{R}_+$ implies that there exists an optimal solution of problem (5) satisfying $\alpha_{t,\ell} \geq \beta$.

Problem (5) can easily be solved since the steps in the function $\|q_t(\alpha)\|$ occur at all α in the set

$$\left\{ \frac{(2j+1)\|2u_t - e\|}{2|2u_t^i - e|} : i = 1, 2, \ldots, n, j \in \mathbb{N} \right\}.$$

The lower bound proposed in the previous lemma is used as a starting value. This bound will also be used later to justify that $\alpha_{t,\ell}$ grows large with ℓ.

The adjusted Halton direction $q_{t,\ell}$ is defined to be equal to $q_t(\alpha_{t,\ell})$. The following Lemma ensures that $q_{t,\ell}$ is a nonzero integer vector.

LEMMA 3.3. *If $t \neq 0$, the adjusted Halton direction $q_{t,\ell}$ satisfies $\|q_{t,\ell}\| \geq 1$.*

Proof. From (4), if $t \neq 0$, then $\alpha = \frac{\|2u_t - e\|}{2\|2u_t - e\|_\infty}$ is feasible for problem (5) since $\|q_t(\alpha)\| = 1 \leq 2^{|\ell|/2}$ for all ℓ.

A consequence of the proof of Lemma 3.3 that will be used later is that when $\ell = 0$ and $t \neq 0$, then the vector $q_{t,\ell}$ contains a unique nonzero coordinate equal to ± 1.

Table 2 shows elements of the sequences u_t and $q_{t,\ell}$ for $n = 4$ and eight pairs (t, ℓ) whose values are compatible with the ORTHOMADS algorithm presented in section 3.4 (i.e., $t = \ell + 7$, where $t_0 = 7$ is the $n = 4$th prime number). The values of $\alpha_{t,\ell}$ and the square norm $\|q_{t,\ell}\|^2$ are also reported (this last value is used in section 3.3). One can also notice that $\alpha_{t,\ell}$ often differs from $2^{|\ell|/2}$. In the example illustrated in Figure 1, with $(t, \ell) = (6, 3)$ and $q_t(\alpha) = \text{round}\left(\frac{\alpha}{\sqrt{481}}(-9, -20)^T\right)$, an optimal solution of (5) is $\alpha_{t,\ell} = 2$, and it satisfies $\|q_{t,\ell}\| = \sqrt{5} < \sqrt{8} = 2^{|\ell|/2} < \|q_t(2^{|\ell|/2})\| = \|(-1, -3)^T\| = \sqrt{10}$.

TABLE 2

The sequence of Halton directions u_t and the adjusted Halton directions $q_{t,\ell}$ for $n = 4$ and eight pairs (t, ℓ).

(t, ℓ)	u_t				$\alpha_{t,\ell}$	$q_{t,\ell}$				$\|q_{t,\ell}\|^2$
	$u_{t,2}$	$u_{t,3}$	$u_{t,5}$	$u_{t,7}$						
$(7, 0)$	7/8	5/9	11/25	1/49	0.8	0	0	0	-1	1
$(8, 1)$	1/16	8/9	16/25	8/49	1.0	-1	1	0	0	2
$(9, 2)$	9/16	1/27	21/25	15/49	1.9	0	-1	1	-1	3
$(10, 3)$	5/16	10/27	2/25	22/49	2.8	-1	-1	-2	0	6
$(11, 4)$	13/16	19/27	7/25	29/49	3.5	2	2	-2	1	13
$(12, 5)$	3/16	4/27	12/25	36/49	5.6	-3	-4	0	2	29
$(13, 6)$	11/16	13/27	17/25	43/49	7.0	3	0	3	6	54
$(14, 7)$	7/16	22/27	22/25	2/49	11.5	-1	5	6	-8	126

The following proposition gives a property of the scaling and rounding operations that transform a vector v into $q = \text{round}(\alpha v/\|v\|)$. The property states that the directions $v/\|v\|$ and $q/\|q\|$ are arbitrarily close for sufficiently large values of α.

PROPOSITION 3.4. *Let $v \neq 0$ be a vector in $\mathbb{R}^n$. For any $\varepsilon > 0$,*

$$if\ \alpha > \frac{2\sqrt{n}}{\varepsilon} + \frac{\sqrt{n}}{2}\ and\ q = \text{round}\left(\alpha \frac{v}{\|v\|}\right) \neq 0,\ then\ \left\| \frac{q}{\|q\|} - \frac{v}{\|v\|} \right\| < \frac{\varepsilon}{2}.$$

Proof. Consider $\varepsilon > 0$ and $\alpha > 2\sqrt{n}/\varepsilon + \sqrt{n}/2$. The vector q may be expressed as $q = \alpha \frac{v}{\|v\|} + \delta$, where $\delta = (\delta_1, \delta_2, \ldots, \delta_n)^T$ and $|\delta_i| \leq 1/2$ for all $i = 1, 2, \ldots, n$.

It follows that

$$\begin{aligned}\left\| \frac{q}{\|q\|} - \frac{v}{\|v\|} \right\| &= \left\| \left(\frac{\alpha}{\|q\|} - 1 \right) \frac{v}{\|v\|} + \frac{\delta}{\|q\|} \right\| \\ &\le \left\| \left(\frac{\alpha}{\|q\|} - 1 \right) \frac{v}{\|v\|} \right\| + \left\| \frac{\delta}{\|q\|} \right\| \\ &= \frac{|\alpha - \|q\||}{\|q\|} + \frac{\|\delta\|}{\|q\|}.\end{aligned}$$

The norm of q can be bounded with

$$\alpha \frac{\|v\|}{\|v\|} - \|\delta\| \le \|q\| \le \alpha \frac{\|v\|}{\|v\|} + \|\delta\|,$$

and therefore $|\alpha - \|q\|| \le \|\delta\|$. Furthermore, $\alpha > 2\sqrt{n}/\varepsilon + \sqrt{n}/2 > \sqrt{n}/2$, and $\|\delta\| \le \sqrt{n}/2$ imply that α satisfies $0 < \alpha - \|\delta\|$. It follows that

$$\left\| \frac{q}{\|q\|} - \frac{v}{\|v\|} \right\| \;\le\; \frac{2\|\delta\|}{\|q\|} \;\le\; \frac{2\|\delta\|}{\alpha - \|\delta\|} \;\le\; \frac{\sqrt{n}}{\alpha - \sqrt{n}/2} \;<\; \frac{\varepsilon}{2}.$$

3.3. Construction of an orthogonal integer basis. Given an integer nonzero vector $q \in \mathbb{Z}^n$, we apply the (symmetric) scaled Householder transformation [19] to construct an orthogonal basis for $\mathbb{R}^n$ composed of integer vectors:

$$H = \|q\|^2 (I_n - 2vv^T), \quad \text{where} \quad v = \frac{q}{\|q\|}. \tag{6}$$

The use of Householder transformations in the context of GPS algorithms was first proposed in [2] to rotate a minimal positive basis.

PROPOSITION 3.5. *The columns of H form an integer orthogonal basis for $\mathbb{R}^n$.*

Proof. First, the columns of H are mutually orthogonal, since $v^T v = 1$ and

$$\begin{aligned}H^T H &= \|q\|^4 (I_n - 2vv^T)^T (I_n - 2vv^T) \\ &= \|q\|^4 (I_n - 2vv^T - 2vv^T + 4vv^T vv^T) \;=\; \|q\|^4 I_n.\end{aligned}$$

Second, by dividing the previous equation by $\|q\|^4$ and applying symmetry, we reveal the inverse of H as $H^{-1} = \frac{1}{\|q\|^4} H$. Since H^{-1} exists, the columns of H form a basis in $\mathbb{R}^n$. Finally, the entries of

$$H \;=\; \|q\|^2 I_n - 2\|q\|^2 \frac{q}{\|q\|} \frac{q^T}{\|q\|} \;=\; \|q\|^2 I_n - 2qq^T$$

are integer, since q and $\|q\|^2$ are integer.

The next proposition shows that the Householder transformation applied to a dense set of normalized directions produces a dense set of normalized directions.

PROPOSITION 3.6. *Let $\{q_t\}_{t=1}^{\infty}$ be a sequence in $\mathbb{R}^n$. For $t = 1, 2, \ldots,$ define $v_t = \frac{q_t}{\|q_t\|}$ and $H_t = \|q_t\|^2 (I_n - 2v_t v_t^T)$. If $\{v_t\}_{t=1}^{\infty}$ is dense on the unit sphere, then the normalized sequence composed of the ith columns of H_t, $\left\{\frac{H_t e_i}{\|H_t e_i\|}\right\}_{t=1}^{\infty}$ is dense on the unit sphere for any $i \in \{1, 2, \ldots, n\}$.*

Proof. Let $w \in \mathbb{R}^n$ with $\|w\| = 1$ be an arbitrary vector of norm 1, let $\varepsilon > 0$ be some small positive number, and let $i \in \{1, 2, \ldots, n\}$ be the index of a column. For

$n > 1$ ($n = 1$ is trivial), we need to show that there exists an index $t \in \mathbb{N}$ such that the ith column of H_t, $H_t e_i$, satisfies

$$\left\| \frac{H_t e_i}{\|H_t e_i\|} - w \right\| \;<\; \varepsilon.$$

First, observe that $\|H_t e_i\| = \sqrt{e_i^T H_t^T H_t e_i} = \|q_t\|^2$, and therefore $\frac{H_t e_i}{\|H_t e_i\|} = e_i - 2v_t v_t^T e_i$. Now, define the vector $d \in \mathbb{R}^n$, where

$$d = \begin{cases} \frac{1}{\sqrt{2(1-w_i)}}(e_i - w) & \text{if } w_i < 1, \\ e_{i+1 (\text{mod } n)} & \text{otherwise.} \end{cases}$$

Observe that if $w_i = 1$, then the vector d satisfies $\|d\| = 1$ and $2d_i d = 0 = e_i - w$, and if $w_i < 1$, then

$$\begin{aligned} \|d\| = \sqrt{d^T d} \;&=\; \sqrt{\frac{1}{2(1-w_i)}(e_i - w)^T (e_i - w)} \;=\; 1, \\ 2d_i d = \frac{1}{(1-w_i)}(e_i - w)_i (e_i - w) \;&=\; e_i - w. \end{aligned}$$

By assumption, $\{v_t\}_{t=1}^{\infty}$ is dense on the unit sphere, and therefore there exists some index t such that $v_t = d + \delta$, where $\delta \in \mathbb{R}^n$ is small enough to satisfy $\|\delta_i(d+\delta) + d_i\delta\| < \varepsilon/2$. The proof may be completed as follows:

$$\begin{aligned} \left\| \frac{H_t e_i}{\|H_t e_i\|} - w \right\| &= \left\| e_i - 2v_t v_t^T e_i - w \right\| \\ &= \left\| e_i - 2(d+\delta)(d+\delta)^T e_i - w \right\| \\ &= \| e_i - 2(d_i + \delta_i)(d+\delta) - w \| \\ &= \| e_i - 2d_i d - w - 2(\delta_i(d+\delta) + d_i\delta) \| \\ &= \| e_i - (e_i - w) - w - 2(\delta_i(d+\delta) + d_i\delta) \| \\ &= 2\,\| \delta_i(d+\delta) + d_i\delta \| \;<\; \varepsilon. \end{aligned}$$

In Figure 1, the Householder transformation is applied to $q_{t,\ell} = (-1, -2)^T$ and produces the integer orthogonal basis $H_{t,\ell} = \begin{bmatrix} 3 & -4 \\ -4 & -3 \end{bmatrix}$.

Observe that in the proof of the previous result, the norm of each column of H_t is shown to be equal to $\|q_t\|^2$. Now, consider the case where q_t is the adjusted Halton direction $q_{t,\ell}$. In light of the comment immediately following Lemma 3.3, when $\ell = 0$ and $t \neq 0$, then $\|q_{t,\ell}\|$ is equal to 1, which implies that the maximal positive basis $[H_t \;\; -H_t]$ is composed of the positive and negative coordinate directions, i.e., the same directions considered by the coordinate search variant of the GPS algorithm.

3.4. The ORTHOMADS algorithm. The new ORTHOMADS instance of MADS can now be defined by combining the components introduced in sections 3.1–3.3. The poll set P_k used by ORTHOMADS at iteration k is entirely determined by the values of the Halton and mesh indexes t_k and ℓ_k. The t_kth element of the Halton sequence u_{t_k} is used to create the adjusted Halton direction q_{t_k,ℓ_k} whose norm is as close as possible to $2^{|\ell_k|/2}$. The Householder transformation on q_{t_k,ℓ_k} produces an orthogonal integer basis H_{t_k,ℓ_k}, and the norm of each column is close to $2^{|\ell_k|}$.

The LtMads and OrthoMads algorithms are identical except for the construction of the set P_k and the poll directions D_k. The set of directions $D = [I_n \ \ -I_n]$ defining the mesh M_k, the link between the mesh index ℓ_k and the mesh and poll size parameters (1), and the mesh update rules detailed in section 2 are identical for both algorithms. Figure 2 describes the OrthoMads algorithm.

[0] Initialization

  $x_0 \in \Omega$, $\ell_0 \leftarrow 0$, $k \leftarrow 0$, $t_0 \leftarrow p_n$

[1] Iteration k

  SEARCH (optional)
    evaluate f on a finite set $S_k \subset M_k$
  POLL (optional if the SEARCH was successful)
    if the POLL size is the smallest one so far
    $\left(\text{i.e., if } \Delta_k^p = \min\{\Delta_j^p : j = 0, 1, \ldots, k\}\right)$
      $t_k \leftarrow \ell_k + t_0$
    else (i.e., smaller POLL sizes were considered)
      $t_k \leftarrow 1 + \max\{t_j : j = 0, 1, \ldots, k-1\}$
    compute u_{t_k}, q_{t_k,ℓ_k}, H_{t_k,ℓ_k}, and $D_k = [H_{t_k,\ell_k} \ \ -H_{t_k,\ell_k}]$
    evaluate f on $P_k \subset M_k$

[2] Updates

  if the iteration is successful (there exists $x_s \in S_k$ or $x_p \in P_k$ such that $f(x_s) < f(x_k)$ or $f(x_p) < f(x_k)$)
    $x_{k+1} \leftarrow x_s$ or x_p
    $\ell_{k+1} \leftarrow \ell_k - 1$
  else (iteration failure)
    $x_{k+1} \leftarrow x_k$
    $\ell_{k+1} \leftarrow \ell_k + 1$
  $k \leftarrow k + 1$
  goto [1] if no stopping condition is met

Fig. 2. *The* OrthoMads *algorithm.*

The poll directions D_k depend entirely on the Halton and mesh indexes t_k and ℓ_k. These integers are chosen to ensure that there will be a sequence of unsuccessful iterations for which the mesh size parameter goes to zero, and such that the directions used in that subsequence will be the tail of the entire Halton sequence. In order to accomplish that goal, we keep track of the value of the smallest poll size parameter visited so far. At every iteration where Δ_k^p is equal to that value, we set $t_k = \ell_k + t_0$. A consequence of this way of fixing t_k is that the set of ordered indices

$$U := \{k_1, k_2, \ldots\} = \{k \ : \ \text{ iteration } k \text{ is unsuccessful, and } \Delta_k^p \leq \Delta_j^p \ \forall j = 0, 1, \ldots k\}$$

satisfies $(t_{k_1}, \ell_{k_1}) = (t_0, 0)$, $(t_{k_2}, \ell_{k_2}) = (t_0 + 1, 1), \ldots, (t_{k_i}, \ell_{k_i}) = (t_0 + i - 1, i - 1)$, and the set of Halton directions $\{u_{t_k}\}_{k \in U}$ is precisely $\{u_t\}_{t=t_0}^{\infty}$.

At the other iterations, those for which smaller poll sizes were previously considered, we just keep increasing t_k so that a new Halton direction is used. Examples of pairs (t_k, ℓ_k) can be seen in Table 3. The boldface entries are those where the poll size parameter is the smallest one so far. In this example, the first three indices of U would be $\{4, 5, 8\}$.

As in LtMads, the basis H_{t_k,ℓ_k} is completed to a maximal positive basis composed

TABLE 3

Example of ORTHOMADS *iterations for $n = 4$. Iterations $k \in \{4, 5, 8\}$ correspond to failed iterations with consecutive Halton elements $t_k = 7, 8$, and 9 satisfying $t_k = \ell_k + t_0$ with $t_0 = 7$.*

k	$(t_k$	,	$\ell_k)$	Δ_k^m	Δ_k^p	$\|D_k e_i\|$	Succ/Fail
0	(**7**	,	**0**)	1	1	1	S
1	(8	,	−1)	1	2	2	S
2	(9	,	−2)	1	4	3	F
3	(10	,	−1)	1	2	2	F
4	(**7**	,	**0**)	1	1	1	F
5	(**8**	,	**1**)	1/4	1/2	2	F
6	(**9**	,	**2**)	1/16	1/4	3	S
7	(11	,	1)	1/4	1/2	2	F
8	(**9**	,	**2**)	1/16	1/4	3	F
9	(**10**	,	**3**)	1/64	1/8	6	...

TABLE 4

Example of a sequence of ORTHOMADS *bases corresponding to failed iterations. Pairs (t_k, ℓ_k) correspond to consecutive Halton elements $t = 7, 8, \ldots, 14$ with $t_k = \ell_k + t_0$, $\ell_0 = 0$, and $t_0 = 7$.*

k (t_k, ℓ_k) $\|H_{t_k,\ell_k} e_i\|$	H_{t_k,ℓ_k}	k (t_k, ℓ_k) $\|H_{t_k,\ell_k} e_i\|$	H_{t_k,ℓ_k}
4 $(7,0)$ 1	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$	14 $(11,4)$ 13	$\begin{bmatrix} 5 & -8 & 8 & -4 \\ -8 & 5 & 8 & -4 \\ 8 & 8 & 5 & 4 \\ -4 & -4 & 4 & 11 \end{bmatrix}$
5 $(8,1)$ 2	$\begin{bmatrix} 0 & 2 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$	16 $(12,5)$ 29	$\begin{bmatrix} 11 & -24 & 0 & 12 \\ -24 & -3 & 0 & 16 \\ 0 & 0 & 29 & 0 \\ 12 & 16 & 0 & 21 \end{bmatrix}$
8 $(9,2)$ 3	$\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 2 & -2 \\ 0 & 2 & 1 & 2 \\ 0 & -2 & 2 & 1 \end{bmatrix}$	17 $(13,6)$ 54	$\begin{bmatrix} 36 & 0 & -18 & -36 \\ 0 & 54 & 0 & 0 \\ -18 & 0 & 36 & -36 \\ -36 & 0 & -36 & -18 \end{bmatrix}$
11 $(10,3)$ 6	$\begin{bmatrix} 4 & -2 & -4 & 0 \\ -2 & 4 & -4 & 0 \\ -4 & -4 & -2 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix}$	19 $(14,7)$ 126	$\begin{bmatrix} 124 & 10 & 12 & -16 \\ 10 & 76 & -60 & 80 \\ 12 & -60 & 54 & 96 \\ -16 & 80 & 96 & -2 \end{bmatrix}$

of $2n$ directions,

$$D_k \;=\; [H_{t_k,\ell_k} \;\; -H_{t_k,\ell_k}],$$

the set of poll directions. A minimal positive basis with $n + 1$ directions is not considered here in order to minimize the cosine measure (2). Table 4 illustrates ORTHOMADS bases H_{t_k,ℓ_k}, with possible pairs (t_k, ℓ_k).

Notice that any direction $D_k e_i$ $(1 \le i \le 2n)$ satisfies $\|D_k e_i\| = \|q_{t,\ell}\|^2 \le (2^{|\ell|/2})^2 = 2^{|\ell|}$. Therefore, the poll trial point $x_k + \Delta_k^m D_k e_i$ is at a Euclidean distance of at most $\Delta_k^m 2^{|\ell|} = \Delta_k^p$ from the poll center. This distance is comparable to that used in LTMADS, where the poll trial points are exactly at a distance Δ_k^p (using the

ℓ_∞ norm) from the poll center. We also already remarked that directions generated using the mesh of index $\ell_k = 0$ are the coordinate directions.

We conclude this section with the following propositions showing that OrthoMads has the same convergence properties as in [6] with no need for a probabilistic argument.

Proposition 3.7. *The set of normalized directions* $\left\{\frac{q_{t,\ell}}{\|q_{t,\ell}\|}\right\}_{t=1}^{\infty}$ *with* $\ell = t - t_0$ *is dense on the unit sphere.*

Proof. Let $\varepsilon > 0$ and $d \in \mathbb{R}^n$ with $\|d\| = 1$. Proposition 3.1 states that the Halton sequence $\{u_t\}_{t=1}^{\infty}$ is dense in the unit cube $[0,1]^n$. Therefore, there exists an index t such that $\frac{2^{|t-t_0|/2}}{\sqrt{n}} - \frac{1}{2} > \frac{2\sqrt{n}}{\varepsilon} + \frac{\sqrt{n}}{2}$ and $\left\|\frac{2u_t - e}{\|2u_t - e\|} - d\right\| \leq \frac{\varepsilon}{2}$.

Lemma 3.2 ensures that $\alpha_{t,\ell} \geq \frac{2^{|\ell|/2}}{\sqrt{n}} - \frac{1}{2} > \frac{2\sqrt{n}}{\varepsilon} + \frac{\sqrt{n}}{2}$. Combining this last inequality with Proposition 3.4 gives

$$\begin{aligned}\left\|\frac{q_{t,\ell}}{\|q_{t,\ell}\|} - d\right\| &\leq \left\|\frac{q_{t,\ell}}{\|q_{t,\ell}\|} - \frac{2u_t - e}{\|2u_t - e\|}\right\| + \left\|\frac{2u_t - e}{\|2u_t - e\|} - d\right\| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \;=\; \varepsilon.\end{aligned}$$

This allows us to state our main result.

Theorem 3.8. OrthoMads *is a valid* Mads *instance.*

Proof. In order to show that OrthoMads is a valid Mads instance we need to show that the poll directions satisfy the following four properties [4, 6]:

- Any direction $D_k e_i$ $(1 \leq i \leq 2n)$ can be written as a nonnegative integer combination of the directions of D: This is the case by construction.
- The distance from the poll center x_k to a poll trial point (in ℓ_∞ norm) has to be bounded above by Δ_k^p: This is also the case by construction because we ensured that $\|D_k e_i\| \leq 2^{|\ell_k|}$ for all i in $\{1, 2, \ldots, 2n\}$ and $\|\Delta_k^m D_k e_i\|_\infty \leq \|\Delta_k^m D_k e_i\| \leq \Delta_k^m 2^{|\ell_k|} = \Delta_k^p$.
- Limits (as defined in [14]) of convergent subsequences of the normalized sets $\overline{D_k} = \{d/\|d\| : \; d \in D_k\}$ are positive spanning sets. This can be shown the same way as in [4] where the proof for LtMads is detailed, since, for OrthoMads and with $\overline{H_{t_k,\ell_k}} = \{d/\|d\| : d \in H_{t_k,\ell_k}\}$, $\det\left(\overline{H_{t_k,\ell_k}}\right) = -1$.
- The set of normalized directions used over all failed iterations is dense on the unit sphere: The strategy chosen for the values of t_k and ℓ_k ensures that there exists a sequence of failed iterations corresponding to consecutive values of t_k. These iterations $k \in U$ can be chosen to correspond to large values of ℓ_k because, from [6], $\lim_{k \in U_k \to \infty} \Delta_k^m = 0$, and $\Delta_k^m = 4^{-\ell_k}$ for $\ell_k \geq 0$. For $k \in U$, the sets of directions $\{D_k\}_{k \in U}$ are constructed from consecutive directions q_{t_k,ℓ_k}, which are dense on the unit sphere after normalization (Proposition 3.7). Then, from Proposition 3.6 and since $D_k = [H_{t_k,\ell_k} \; - H_{t_k,\ell_k}]$, the set of normalized directions $\left\{\frac{D_k e_i}{\|D_k e_i\|}\right\}_{k \in U}$ is also dense on the unit sphere for all $i = 1, 2, \ldots, 2n$. ∎

4. Numerical tests. In this section, OrthoMads is compared on 45 problems from the literature to its immediate predecessor, LtMads [6], as well as to the Gps method [28]. All tests were performed with Nomad 3.1 [1], our latest C++ implementation of Mads. In the Mads algorithms, the theory supports handling constraints by the extreme barrier approach: Points outside Ω are simply ignored and f is not evaluated. For Gps, the extreme barrier approach is supported by the theory only for a finite number of linear constraints [21]. Still, for comparison, we apply two different approaches: the extreme barrier (GPS-EB), and the filter method described in [5] (GPS-FILTER), which has stronger theoretical support.

Because of its random behavior, 30 instances of LTMADS are performed for each problem. GPS and ORTHOMADS are scored by comparing them against the 30 LTMADS instances. A score of s for GPS or ORTHOMADS means that this instance gave a value of f at least as good as s of the 30 LTMADS instances, with a relative precision of 1%. The worst score is 0 and a perfect score corresponds to 30. We classify scoring as follows: a poor instance has a score less than 10, an acceptable instance is between 10 and 19, and a good instance has a score greater than or equal to 20.

The mesh index $\ell_k \in \mathbb{Z}$ (see (1)) that defines the mesh and poll size parameters Δ_k^m and Δ_k^p at iteration k is allowed to be negative for both LTMADS and ORTHOMADS. Maximal positive bases ($2n$ directions) are used in the three methods, as is the opportunistic strategy (the poll is interrupted at the first success) and the optimistic strategy (after a successful point has been found, a search point is generated further along the same direction). No other search is performed. The stopping criteria is satisfied when the number of function evaluations reaches $1000n$ or when the poll size parameter Δ_k^p drops below 10^{-12}. The initial mesh size is always set to $\Delta_0^m = 1$, and when problems include bound constraints, a scaling is performed in order to consider bounds between 0 and 100.

The methods are tested on 45 problems divided into 4 groups: our choice of smooth and nonsmooth unconstrained problems is the same as in [17] and [16], respectively, with 21 smooth problems from the CUTEr test set [24][1] and 13 nonsmooth problems from [22], which is a compilation of nonsmooth problems from the literature. We also tested on 9 constrained problems from [7, 9, 22], and in addition, we added two problems from [3] that correspond to real applications.

The problem descriptions and numerical results are summarized in Tables 5–8. $f(x^*)$ denotes the best known minimal value of f, and *value* denotes the final value of f for each method. Note that *value* is displayed as 0.00 when its absolute value is less than 10^{-6}. Finally, *evals* represents the number of function evaluations that each method performed. Table 5 shows results on the 21 unconstrained smooth problems from CUTEr. ORTHOMADS has a perfect score on 13 of these problems. For both ARWHEAD problems, all runs shared the same behavior: Every run generated the optimal solution during the first iteration since, at that iteration, the poll directions of GPS, LTMADS, and ORTHOMADS are composed of the standard $2n$ coordinate directions. Table 6 displays results on the 13 unconstrained nonsmooth problems, where ORTHOMADS achieves good scores on 5 problems. Table 7 shows results for the 9 constrained problems. A total of 7 problems are considered good for ORTHOMADS. Finally, Table 8 presents results for the two real applications, and ORTHOMADS has perfect scores on one of them, and a good score on the other.

Table 9 summarizes the results. The first observation is that both MADS instances outperform GPS. For 20 problems out of 45, ORTHOMADS found the same solution as the best of 30 LTMADS runs. The new method solved 30 out of 45 problems efficiently enough that, for these problems, the single run of ORTHOMADS was better than two-thirds of the 30 LTMADS runs. For 8 problems, the two methods performed equally well, and for 7 problems, two-thirds of the LTMADS runs gave a better solution than the one produced by ORTHOMADS. The last four columns of Table 9 display the ratio

[1] Two different versions of the BDQRTIC problem can be found in the literature. We consider the one defined in the CUTEr set [24]: $f(x) = \sum_{i=1}^{n-4}(-4x_i+3)^2 + (x_i^2 + 2x_{i+1}^2 + 3x_{i+2}^2 + 4x_{i+3}^2 + 5x_n^2)^2$. In [17, 25], the term $(-4x_i + 3)$ is not squared.

Table 5
Results for CUTEr *unconstrained smooth problems. A score of s for a method indicates that the final f value is at least as good as s of the* 30 LTMADS *runs (with a relative error of* 1%*).*

Problem n $f(x^*)$	LTMADS×30 worst evals value	median evals value	best evals value	GPS evals value	score	ORTHOMADS evals value	score
ARWHEAD 10 0.00	791 0.00	791 0.00	791 0.00	910 0.00	30	891 0.00	30
ARWHEAD 20 0.00	1581 0.00	1581 0.00	1581 0.00	1820 0.00	30	1781 0.00	30
BDQRTIC 10 18.3	6554 18.3	5472 18.3	4381 18.3	3439 18.3	30	4723 18.3	30
BDQRTIC 20 58.3	20000 58.3	18953 58.3	16279 58.3	17328 58.3	30	20000 58.3	30
BIGGS6 6 0.00	3284 3.06E-1	6000 1.27E-3	6000 2.48E-5	6000 7.80E-3	12	788 4.74E-4	26
BROWNAL10 10 0.00	10000 1.84	10000 1.77E-4	10000 0.00	10000 1.00	8	10000 7.81E-3	10
BROWNAL20 20 0.00	20000 1.00	20000 1.00	20000 3.60E-2	20000 1.00	29	20000 9.52E-1	29
PENALTY1 10 7.09E-5	10000 7.14E-5	10000 7.09E-5	10000 7.09E-5	10000 8.82E-5	0	10000 7.09E-5	30
PENALTY1 20 1.58E-4	20000 1.58E-4	20000 1.58E-4	20000 1.58E-4	20000 1.87E-4	0	20000 1.58E-4	30
PENALTY2 10 2.94E-4	10000 3.06E-4	10000 2.99E-4	10000 2.95E-4	10000 3.72E-4	0	10000 2.97E-4	30
PENALTY2 20 6.39E-3	20000 6.40E-3	20000 6.39E-3	20000 6.39E-3	20000 8.86E-3	0	20000 6.39E-3	30
POWELLSG 12 0.00	12000 3.45E-4	12000 1.32E-4	1635 0.00	1498 0.00	30	1483 0.00	30
POWELLSG 20 0.00	20000 2.32E-3	20000 8.12E-4	20000 3.93E-4	3034 0.00	30	3011 0.00	30
SROSENBR 10 0.00	10000 6.14E-1	10000 1.68E-1	10000 4.24E-2	10000 1.69E-1	15	10000 1.45E-1	19
SROSENBR 20 0.00	20000 10.4	20000 2.96	20000 2.62E-1	20000 3.68E-1	26	20000 3.86E-1	26
TRIDIA 10 0.00	10000 6.30E-5	5407 0.00	5468 0.00	5506 0.00	30	5311 0.00	30
TRIDIA 20 0.00	20000 6.05E-3	20000 2.29E-4	20000 0.00	20000 0.00	30	20000 0.00	30
VARDIM 10 0.00	8520 0.00	7181 0.00	6123 0.00	10000 2.55	0	8861 0.00	30
VARDIM 20 0.00	20000 11.6	20000 3.73E-5	20000 0.00	20000 36.1	0	20000 3.08E-2	4
WOODS 12 0.00	12000 18.6	12000 9.52	12000 6.50E-1	12000 36.3	0	12000 12.7	5
WOODS 20 0.00	20000 28.8	20000 16.3	20000 5.71	20000 432	0	20000 35.8	0
				average scores	15.7		24.2

Table 6

Results for unconstrained nonsmooth problems from [22]*. A score of s for a method indicates that the final f value is at least as good as s of the* 30 LtMads *runs (with a relative error of* 1%*).*

Problem			LtMads×30			Gps		OrthoMads	
	n	$f(x^*)$	worst evals value	median evals value	best evals value	evals value	score	evals value	score
ElAttar			1002	1513	3992	4766	15	804	5
	6	5.6E-1	8.02	1.84	5.6E-1	2.51		7.18	
Evd61			548	3387	3609	897	6	1480	6
	6	3.49E-2	1.62	1.15E-1	3.69E-2	8.72E-1		6.99E-1	
Filter			870	1510	1085	1111	7	1648	12
	9	6.19E-3	9.61E-3	8.99E-3	8.27E-3	9.50E-3		9.31E-3	
Goffin			31907	31913	29201	31671	30	50000	30
	50	0.00	7.00E-1	4.50E-1	1.25E-2	0.00		6.05E-5	
Hs78			913	906	855	927	20	779	29
	5	−2.92	0.00	−2.15	−2.84	−2.37		−2.76	
L1Hilb			48499	50000	50000	50000	0	50000	0
	50	0.00	2.02	3.63E-1	3.75E-2	4.24		4.79	
MxHilb			12505	14967	18608	7789	0	31197	15
	50	0.00	1.49E-1	7.57E-2	1.26E-3	9.93E-1		7.90E-2	
Osborne2			11000	3159	11000	3054	21	2982	10
	11	4.80E-2	2.11E-1	1.19E-1	5.32E-2	8.88E-2		1.37E-1	
Pbc1			1047	1930	1618	1024	0	1383	25
	5	2.23E-2	3.96E-1	1.91E-1	4.48E-2	1.07		8.51E-2	
Polak2			1399	1132	1002	1126	30	1116	30
	10	54.6	54.6	54.6	54.6	54.6		54.6	
Shor			1305	1128	1030	791	0	1062	10
	5	22.6	22.9	22.7	22.6	23.8		23.0	
Wong1			1999	2138	1981	806	30	2631	30
	7	681	697	694	693	766		694	
Wong2			7685	4575	3964	1736	0	7454	12
	10	24.3	39.0	27.4	24.7	47.4		28.1	
						average scores	12.2		16.5

of the number of instances solved to the best known solution on the total number of instances, with a relative 1% tolerance when the best known solution is different than zero, and an absolute tolerance of 10^{-6} otherwise. For example, 54 LtMads runs on nonsmooth problems reached by the best known solution, over a total of 13×30 runs, which gives a solution quality ratio of $\simeq 14\%$. Both Mads implementations gave superior results compared to the Gps implementations.

Figure 3 illustrates the spread of the directions for both LtMads and OrthoMads. Rosenbrock's function [27] with $n = 2$ and $n = 3$ was used with 2000 and 3000 evaluations, respectively. In the two-dimensional case, all the normalized directions used to generate poll trial points are directly represented on the top two subfigures. It is clear that OrthoMads directions are well distributed on the unit circle. This is not the case with LtMads because half the directions correspond to either $\pm e_1$ or $\pm e_2$. For $n = 3$, the two plots on the bottom represent the standard angles of the normalized directions in spherical coordinates. There again it can be seen that OrthoMads directions have a better distribution than those of LtMads, since at least two-thirds of the LtMads directions possess some null coordinates. On the subfigure using LtMads with $n = 3$, the horizontal bar at $\Phi = \pi/2$ corresponds to the set of directions where $z = 0$. The vertical bars at $\theta = \pm\pi/2$ correspond to directions with $x = 0$, and the one at $\theta = 0$ and $\theta = \pi$ correspond to directions with

TABLE 7

Results for constrained problems. A score of s for a method indicates that the final f value is at least as good as s of the 30 LTMADS *runs (with a relative error of* 1%*).*

Problem n m $f(x^*)$	LTMADS×30 worst evals value	 median evals value	 best evals value	GPS-FILTER evals value	 score	GPS-EB evals value	 score	ORTHOMADS evals value	 score
CRESCENT10 [7] 10 2 −9.00	2107 −8.61	1602 −8.92	4934 −8.97	2416 −2.32	0	1106 −2.32	0	8361 −8.98	30
DISK10 [7] 10 1 −17.3	1835 −17.2	2076 −17.3	3346 −17.3	2610 −10.0	0	987 −10.0	0	2664 −17.3	30
B250 [9] 60 1 7.95	60000 21.9	60000 10.4	60000 8.13	25235 440	0	48216 570	0	55217 11.6	11
B500 [9] 60 1 80	60000 350	60000 139	60000 80	17907 750	0	19564 1084	0	60000 86	27
G2 [9] 10 2 −0.740	6111 −0.206	4753 −0.519	5433 −0.740	3387 −0.569	20	1825 −0.330	5	10000 −0.597	23
G2 [9] 20 2 −0.804	14249 −0.338	14328 −0.561	17212 −0.757	11110 −0.464	3	6578 −0.465	3	20000 −0.715	29
HS114 [22] 9 6 −1769	1906 −1003	1756 −1120	3721 −1369	2919 −1252	25	1795 −1038	4	1978 −1040	4
MAD6 [22] 5 7 0.102	1090 0.113	1197 0.104	1527 0.102	2117 0.102	30	1046 0.104	19	1809 0.103	30
PENTAGON [22] 6 15 −1.86	1136 −1.53	1003 −1.81	1564 −1.86	529 0.00	0	529 0.00	0	1055 −1.81	24
				average scores	8.7		3.4		23.1

TABLE 8

Results for real applications. A score of s for a method indicates that the final f value is at least as good as s of the 30 LTMADS *runs (with a relative error of* 1%*). Displayed z values for problem* STY *are divided by* 10^7.

Problem n m $f(x^*)$	LTMADS×30 worst evals value	 median evals value	 best evals value	GPS-FILTER evals value	 score	GPS-EB evals value	 score	ORTHOMADS evals value	 score
MDO [3] 10 10 −3964	3353 −3964	2442 −3964	1721 −3964	2195 −2820	0	1091 −1442	0	2720 −3964	30
STY [3] 8 11 −3.35	1802 −2.88	1913 −3.28	1521 −3.35	2269 −2.92	10	1580 −3.17	11	1634 −3.27	26
				average scores	5.0		5.5		28.0

TABLE 9

Summary of GPS *and* ORTHOMADS *performance. L, F, E, and O correspond, respectively, to* LTMADS, GPS-FILTER, GPS-EB, *and* ORTHOMADS. *A bad instance has a score between* 0 *and* 9*, an acceptable (acc.) instance has a score between* 10 *and* 19*, a good instance has a score higher than* 20 *and a perfect (perf.) instance has a score of* 30*. The last four columns represent the solution quality ratio, or the ratio of the number of instances solved to the best known solution on the total number of instances.*

Problems	Average scores (out of 30)			# of prob.	# of bad instances			# of acc. instances			# of good instances			# of perf. instances			Sol. quality ratio (%)			
	F	E	O		F	E	O	F	E	O	F	E	O	F	E	O	L	F	E	O
smooth	15.7	15.7	24.2	21	9	9	3	2	2	2	10	10	16	8	8	13	46	38	38	57
nonsmooth	12.2	12.2	16.5	13	7	7	3	1	1	5	5	5	5	3	3	3	14	15	15	8
constrained	8.7	3.4	23.1	9	6	8	1	0	1	1	3	0	7	3	0	3	25	11	0	33
real appli.	5.0	5.5	28.0	2	1	1	0	1	1	0	0	0	2	0	0	1	50	0	0	50
total or avg.	12.8	11.8	21.9	45	23	25	7	4	5	8	18	15	30	14	11	20	33	24	22	38

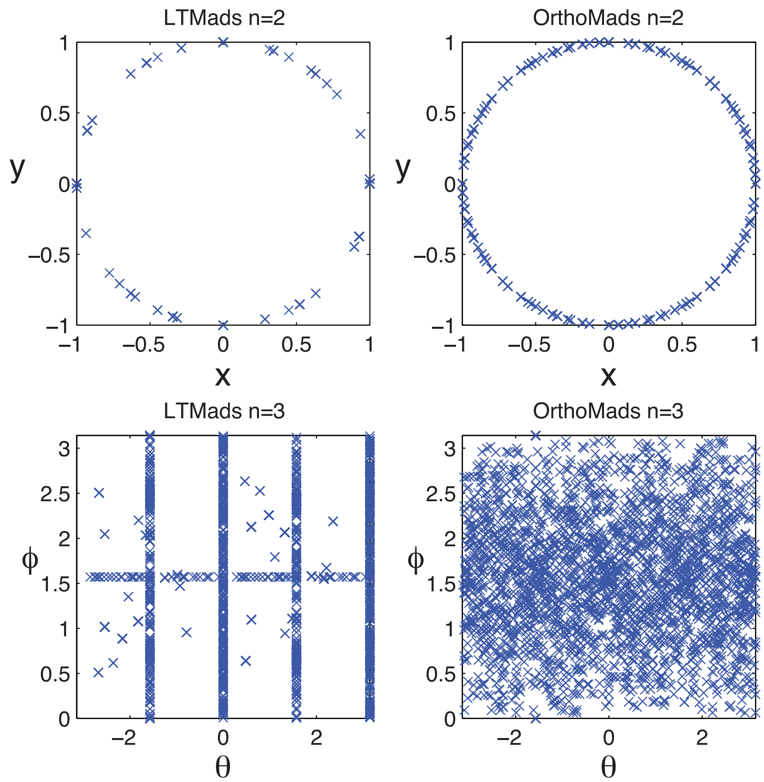

Fig. 3. LtMads *and* OrthoMads *normalized poll directions on the Rosenbrock function with* $n = 2$ *and* $n = 3$.

$y = 0$.

5. Discussion. This paper introduced OrthoMads, an alternative instantiation of the Mads class of algorithms. The advantages of OrthoMads over the original LtMads are that the Mads directions are chosen deterministically, and that those directions are orthogonal to each other, thus minimizing the size of the cones of unexplored directions. Moreover, OrthoMads inherits all of the Mads convergence properties, without probabilistic arguments, and without additional parameters.

Intensive tests on 45 problems from the literature showed that both Mads instances outperform the Gps algorithm, and that OrthoMads is competitive with multiple runs of LtMads, and it has a better distribution of the poll directions.

In order to compare the deterministic OrthoMads with the random LtMads, single runs of OrthoMads were compared to multiple runs of LtMads with different random seeds. We have mentioned the advantages of a deterministic method, but a stochastic method like LtMads has the possibility that by varying the random seed a better solution might be found. Of course, this assumes that the expense of additional runs is acceptable. Multiple runs of OrthoMads are reported in [8], using different values for the Halton t_0 for three engineering applications, different starting points, and three strategies to handle the constraints. However, we believe that our target class of problems, which involve costly black-box functions, rarely allows the cost of performing multiple runs. In this context, it appears that OrthoMads is the best method, and this is the reason why we decided to define it as the default Mads instantiation of our Mads software, Nomad [1].

Acknowledgments. We would like to thank Ana Custódio and Luis Vicente for helpful remarks about missed cones of directions in single iterations of LtMads. These remarks and similar observations of our own helped motivate us to discover OrthoMads. We also wish to thank Andrew Booker for suggesting quasi-Monte Carlo methods which pointed us in directions that led to our use of Halton sequences. We also thank Margaret Wright and an anonymous referee for their careful reading and helpful remarks.

REFERENCES

[1] M. A. Abramson, C. Audet, G. Couture, J. E. Dennis, Jr., and S. Le Digabel, *The* NOMAD *Project*, software available at http://www.gerad.ca/nomad.

[2] P. Alberto, F. Nogueira, H. Rocha, and L. N. Vicente, *Pattern search methods for user-provided points: Application to molecular geometry problems*, SIAM J. Optim., 14 (2004), pp. 1216–1236.

[3] C. Audet, V. Béchard, and S. Le Digabel, *Nonsmooth optimization through mesh adaptive direct search and variable neighborhood search*, J. Global Optim., 41 (2008), pp. 299–318.

[4] C. Audet, A. L. Custódio, and J. E. Dennis, Jr., *Erratum: Mesh adaptive direct search algorithms for constrained optimization*, SIAM J. Optim., 18 (2008), pp. 1501–1503.

[5] C. Audet and J. E. Dennis, Jr., *A pattern search filter method for nonlinear programming without derivatives*, SIAM J. Optim., 14 (2004), pp. 980–1010.

[6] C. Audet and J. E. Dennis, Jr., *Mesh adaptive direct search algorithms for constrained optimization*, SIAM J. Optim., 17 (2006), pp. 188–217.

[7] C. Audet and J. E. Dennis, Jr., *A MADS algorithm with a progressive barrier for derivative-free nonlinear programming*, SIAM J. Optim., 20 (2009), pp. 445–472.

[8] C. Audet, J. E. Dennis, Jr., and S. Le Digabel, *Globalization Strategies for Mesh Adaptive Direct Search*, Comput. Optim. Appl., to appear.

[9] C. Audet, J. E. Dennis, Jr., and S. Le Digabel, *Parallel space decomposition of the mesh adaptive direct search algorithm*, SIAM J. Optim., 19 (2008), pp. 1150–1170.

[10] C. Audet and D. Orban, *Finding optimal algorithmic parameters using derivative-free optimization*, SIAM J. Optim., 17 (2006), pp. 642–664.

[11] A. J. Booker, J. E. Dennis, Jr., P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset, *A rigorous framework for optimization of expensive functions by surrogates*, Structural Optimization, 17 (1999), pp. 1–13.

[12] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983. Reissued in 1990 by SIAM Publications, Philadelphia, in Classics Appl. Math. 5.

[13] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to Derivative-Free Optimization*, MPS/SIAM Ser. Optim. 8, SIAM, Philadelphia, 2009.

[14] I. D. Coope and C. J. Price, *Frame-based methods for unconstrained optimization*, J. Optim. Theory Appl., 107 (2000), pp. 261–274.

[15] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed., MIT Press and McGraw-Hill, 2001, pp. 873–876.

[16] A. L. Custódio, J. E. Dennis, Jr., and L. N. Vicente, *Using simplex gradients of nonsmooth functions in direct search methods*, IMA J. Numer. Anal., 28 (2008), pp. 770–784.

[17] A. L. Custódio and L. N. Vicente, *Using sampling and simplex derivatives in pattern search methods*, SIAM J. Optim., 18 (2007), pp. 537–555.

[18] J. H. Halton, *On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals*, Numer. Math., 2 (1960), pp. 84–90.

[19] A. S. Householder, *Unitary triangularization of a nonsymmetric matrix*, J. Assoc. Comput. Mach., 5 (1958), pp. 339–342.

[20] T. G. Kolda, R. M. Lewis, and V. Torczon, *Optimization by direct search: New perspectives on some classical and modern methods*, SIAM Rev., 45 (2003), pp. 385–482.

[21] R. M. Lewis and V. Torczon, *Pattern search methods for linearly constrained minimization*, SIAM J. Optim., 10 (2000), pp. 917–941.

[22] L. Lukšan and J. Vlček, *Test Problems for Nonsmooth Unconstrained and Linearly Constrained Optimization*, Technical report V-798, ICS AS CR, Prague, 2000.
[23] J. Matoušek, *On the L_2-discrepancy for anchored boxes*, J. Complexity, 14 (1998), pp. 527–556.
[24] N. I. M. Gould, D. Orban, and P. L. Toint, *CUTEr and SifDec: A constrained and unconstrained testing environment, revisited*, ACM Trans. Math. Software, 29 (2003), pp. 373–394.
[25] M. J. D. Powell, *On trust region methods for unconstrained minimization without derivatives*, Math. Program., 97 (2003), pp. 605–623.
[26] R. T. Rockafellar, *Generalized directional derivatives and subgradients of nonconvex functions*, Canad. J. Math., 32 (1980), pp. 257–280.
[27] H. H. Rosenbrock, *An automatic method for finding the greatest or least value of a function*, Comput. J., 3 (1960), pp. 175–184.
[28] V. Torczon, *On the convergence of pattern search algorithms*, SIAM J. Optim., 7 (1997), pp. 1–25.