HIT-AND-RUN ALGORITHMS FOR THE IDENTIFICATION OF NONREDUNDANT LINEAR INEQUALITIES

H.C.P. BERBEE

Department of Mathematical Statistics, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands

C.G.E. BOENDER

Econometric Institute, Erasmus University Rotterdam, The Netherlands

A.H.G. RINNOOY KAN

Econometric Institute, Erasmus University Rotterdam, The Netherlands

C.L. SCHEFFER

Department of Mathematics and Informatics, Delft University of Technology, The Netherlands

R.L. SMITH

Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan, USA

J. TELGEN

Van Dien + Co., Utrecht, The Netherlands

Two probabilistic hit-and-run algorithms are presented to detect nonredundant constraints in a full dimensional system of linear inequalities. The algorithms proceed by generating a random sequence of interior points whose limiting distribution is uniform, and by searching for a nonredundant constraint in the direction of a random vector from each point in the sequence. In the hypersphere directions algorithm the direction vector is drawn from a uniform distribution on a hypersphere. In the computationally superior coordinate directions algorithm a search is carried out along one of the coordinate vectors. The algorithms are terminated through the use of a Bayesian stopping rule. Computational experience with the algorithms and the stopping rule will be reported.

Key words: System of linear inequalities, redundancy, probabilistic hit-and-run algorithms, uniform interior points, Bayesian stopping rule.

1. Introduction

The problem of recognizing *redundant linear inequalities* (i.e., inequalities that can be deleted from a system without changing its set of feasible solutions) is of obvious computational importance. Methods to eliminate such constraints have been proposed by many researchers. Most of them are based on the *simplex method* (e.g. Thompson et al., 1966; Lisy, 1971; Gal, 1975; Telgen, 1979); sometimes, the

simplex method is in fact invoked to enumerate all the extreme points of the feasible solution set (Balinski, 1961; Mattheis, 1973). A detailed computational comparison (Karwan et al., 1983) reveals that all these approaches suffer under the administrative burden of maintaining a complete updated simplex tableau. As a result, they require a computational effort that is hardly compensated for by the subsequent speed-up of calculations carried out on the reduced set of inequalities.

An alternative for these time consuming procedures is provided by *heuristic* methods. Acting as *preprocessors* on the original problem data (Brearly et al., 1975; Bradley et al., 1980), many of these heuristics have found their way into commercial mathematical programming packages, for example under the name of the *REDUCE* option. Of course, the big drawback of these fast procedures is that there is no guarantee that all redundancy present in the system will be identified.

An attractive compromise between the two above approaches is provided by the *probabilistic preprocessors* (cf. Rabin, 1976) that are studied in this paper. The aim of these preprocessors is to identify nonredundant constraints rather than redundant ones. Starting from some interior point (assumed given) of the polyhedron defined by the system (assumed to be bounded and full dimensional), the preprocessor generates a random sequence of interior points: in each successive point, a search is carried out in a random direction and the constraints first encountered in that direction and its negation are identified as being nonredundant (cf. Theorem 1). In the *hypersphere directions method*, the random direction is generated from a uniform distribution on a hypersphere (Boneh and Golan, 1979; Smith, 1980; Boneh, 1983); in the *coordinate directions method*, it is chosen with equal probability from the coordinate direction vectors and their negations (Telgen, 1980). In both cases, the next interior point is generated randomly from a uniform distribution over the line segment connecting the two points where the direction vector and its negation intersect with the polytope. Methods of this type, for obvious reasons, are also referred to as *hit-and-run algorithms*. (In Smith (1984) the term *symmetric mixing algorithms* is used.)

For the hypersphere directions method, it is shown in Smith (1984) that the sequence of interior points has a *limiting distribution* which is uniform over the interior of the polytope. We give a new proof for this result and extend it to the (much more difficult) case of the coordinate directions method in Section 3. These results imply that both methods are *asymptotically correct* in the sense that each nonredundant constraint will be identified with probability one as the number of iterations increases. Indeed, the computational experiments in Karwan et al. (1983) indicate that the great speed at which interior points can be generated turns such a method into a very attractive preprocessing device of great practical value.

This good practical performance provides part of our motivation to consider the important question of when to terminate a hit-and-run procedure; under the assumption that the true number of nonredundant constraints is unknown, the answer is not obvious. To cope with this problem we will prove in Section 4 that the theorems of Section 3 imply that asymptotically each *facet* of the boundary of the feasible

region (or, equivalently, each nonredundant constraint) has a fixed probability of being hit. In Boneh and Golan (1979) this result is applied to determine the expected number of iterations that the hypersphere directions algorithm will need to identify all nonredundant constraints, under the additional assumptions that (i) the successive interior points are statistically independent, (ii) the asymptotic hitting probabilities are all equal, and (iii) each constraint is nonredundant. In practice, however, the hitting probabilities are very different: in systems with less than 100 constraints in 20-dimensional space, for example, we frequently observed empirical probabilities smaller than 10^{-4}. Boneh and Golan regard their result as a lowerbound for the number of iterations, and provide an upperbound under the assumption that the (ordered) probabilities decrease arithmetically. However, under the assumption (i) we will show in Section 4 that the generalized multinomial model studied in Boender (1984, Chapter 3) applies, so that we can obtain a more accurate *Bayesian stopping criterion* for both hit-and-run algorithms, in which the user may incorporate *a priori information* on the equality of the hitting probabilities and on the true number of nonredundant constraints, instead of assuming (ii) and (iii). In particular, the hit-and-run algorithms can be terminated if the *optimal Bayesian estimate* of the number of yet undiscovered nonredundant constraints is 0.

Our experimental results are presented in Section 5; concluding remarks and possible extensions to optimization problems (e.g. linear programming) are the subject of Section 6.

2. The hit-and-run algorithms

Consider a feasible region S defined by a system of linear inequalities

$$a_i^T x \le b_i \quad (i = 1, \ldots, m) \tag{1}$$

with $x \in \mathbb{R}^d$ and $\|a_i\| = 1$ $(i = 1, \ldots, m)$. We will assume that S is bounded, nonempty and of full dimension, so that S is a *polytope* that contains *interior points* for which the inequalities (1) are all satisfied as *strict inequalities*. A *redundant constraint* is defined as an inequality which may be dropped from the system (1) without changing the feasible region S. A *facet* is defined as the interaction of a nonredundant constraint with the boundary of S of dimension $n - 1$.

Our probabilistic hit and run preprocessors to investigate redundancies within system (1) are based on a search from an interior point X in the direction of a vector v with $\|v\| = 1$. Let us denote the straight line passing through X in the direction v by $X + \lambda v$ $(\lambda \in \mathbb{R})$. Then it is immediate that the value of λ at the intersection point with the i-th hyperplane $a_i^T x = b_i$ is equal to $\lambda_i = (b_i - a_i^T X)/a_i^T v$. According to the following theorem, the constraints hit first in the positive $(\lambda > 0)$ and negative $(\lambda < 0)$ direction can be declared to be nonredundant.

Theorem 1. *If*

$$r \triangleq \operatorname*{argmin}_{1 \le i \le m} \{\lambda_i \mid \lambda_i > 0\} \tag{2}$$

and

$$s \stackrel{\Delta}{=} \underset{1 \le i \le m}{\operatorname{argmax}} \{\lambda_i \mid \lambda_i < 0\} \tag{3}$$

are unique, then the constraints $a_r^T x \le b_r$ and $a_s^T x \le b_s$ are nonredundant.

Proof. We only offer a proof for the case corresponding to (2). It suffices to show (Telgen, 1979) that under the conditions of the theorem, there exists a point X'' such that

$$a_r^T X'' > b_r, \tag{4}$$

$$a_i^T X'' \le b_i \quad (i = 1, \ldots, m, \, i \ne r). \tag{5}$$

Obviously $\lambda_i = (b_i - a_i^T X)/a_i^T v > 0$ for an interior point X iff $a_i^T v > 0$. Hence, for $X' = X + \lambda_r v$ we have, for $i \ne r$,

$$a_i^T X' = a_i^T X + \frac{b_r - a_r^T X}{a_r^T v} \, a_i^T v < a_i^T X + \frac{b_i - a_i^T X}{a_i^T v} \, a_i^T v = b_i. \tag{6}$$

Since $a_r^T X' = b_r$, there exists an $\varepsilon > 0$ such that $X'' = X' + \varepsilon v$ satisfies (4) and (5).

If (2) and (3) are not unique, then either identical constraints have been hit, or an intersection of constraints has been hit. The former possibility is assumed not to occur from now on. The latter possibility can only occur with probability 0, and therefore disregarded.

The algorithms to be presented exploit the above theorem by searching in the direction of a random vector v'' from each point X'' of a random *sequence* of interior points. For the *hypersphere directions algorithm* (Boneh and Golan, 1979; Smith, 1980; Boneh, 1983], a direction vector is drawn in each point X'' of the sequence from a uniform distribution on a unit hypersphere with centre X''. The *coordinate directions method* (Telgen, 1980) generates one of the unit coordinate vectors or their negation as direction vector. Both algorithms choose the $(n+1)$-th interior point *uniformly* on the linesegment connecting the previous two *hitpoints*, i.e. the points where the line $X'' + \lambda v''$ intersects the boundary of S. The choice of a stopping criterion is discussed in Section 4.

The algorithms consist of the following steps.

Step 0: Find an interior point X^0. Set $n := 0$.

Step 1: *Hypersphere directions algorithm:* Generate a direction vector v'' from a uniform distribution on a unit hypersphere with centre X''. *Coordinate directions algorithm:* Generate a direction vector v'' with equal probability from one of the d coordinate vectors and their negations.

Step 2: Determine

$$\lambda_i := \frac{b_i - a_i^T X^n}{a_i^T v^n} \quad (i = 1, \ldots, m), \tag{7}$$

$$\lambda^+ := \min_{1 \le i \le m} \{\lambda_i \mid \lambda_i > 0\}, \tag{8}$$

$$\lambda^- := \max_{1 \le i \le m} \{\lambda_i \mid \lambda_i < 0\}. \tag{9}$$

Declare the constraints corresponding to the indices for which the minimum (8) and the maximum (9) are attained to be nonredundant.

Step 3: Generate u from a uniform distribution on $[0, 1]$ and set

$$X^{n+1} = X^n + (\lambda^- + u(\lambda^+ - \lambda^-))v^n. \tag{10}$$

Step 4: Set $n := n + 1$ and go to Step 1, unless a stopping criterion is satisfied.

We conclude this section with a comparison of the computational efficiency of the two algorithms. It is easily verified that the determination of the intersection points of a direction vector with the m hyperplanes (cf. (7)) is the most time consuming part of both algorithms. Since $S \in \mathbb{R}^d$, the computation of one intersection point for the hypersphere directions algorithm requires $O(d)$ time, which implies that it requires $O(md)$ time to evaluate the intersection points of a straight line with all the m hyperplanes. The major advantage of the coordinate directions algorithm is that since only a single coordinate is changed when moving from one interior point to the next, no more than 2 multiplications are needed to update λ_i $(i = 1, \ldots, m)$ (cf. (7)). Hence, for the coordinate directions algorithm the computation of the intersection points requires $O(m)$ time, rather than $O(md)$.

3. The uniform limiting distribution of the interior points

In this section we will prove that for both hit-and-run algorithms the random sequence $\{X^i\}_{i=0}^\infty$ of interior points converges to the uniform distribution μ on S, for each possible starting point in its *interior* S^0. These results are an essential part of the justification of our application of the generalized multinomial model, which underlies the stopping criterion proposed in Section 4. We remarked earlier that for the hypersphere directions algorithm this result has already been proved in (Smith, 1984). However, our proof for this case is new, and it provides an introduction to the proof for the coordinate directions algorithm.

For the remainder of this section we fix a starting point $x^0 \in S^0$, and a Borel set $B^0 \subset S$ with $\mu(B^0) > 0$, and without loss of generality we assume that the Lebesgue measure of S is equal to 1.

From the description of the algorithms in the previous section it is immediate that for all $n \in \mathbb{N}^+$, and $x^1, \ldots, x^n \in S$

$$\Pr\{X^{n+1} \in B^0 \mid X^0 = x^0, \ldots, X^n = x^n\} = \Pr\{X^{n+1} \in B^0 \mid X^n = x^n\}, \tag{11}$$

and that

$$\Pr\{X^{n+1} \in B^0 \mid X^n = x^0\} = \Pr\{X^1 \in B^0 \mid X^0 = x^0\}. \tag{12}$$

Hence, for both algorithms the sequence of interior points defines a *Markov chain* with a *stationary transition probability function* and *continuous state space S*.

Theorem 2. *The hypersphere directions algorithm generates a sequence of interior points whose limiting distribution is uniform on S, i.e.:*

$$\lim_{n \to \infty} \Pr\{\boldsymbol{X}^n \in B^0 \,|\, \boldsymbol{X}^0 = x^0\} = \mu(B^0). \tag{13}$$

Proof. According to Theorem 7.1 in Orey (1971) it is sufficient to prove Propositions (i) and (ii).

(i) μ is *invariant*, i.e. if a current interior point is uniformly distributed, then the next interior point is uniformly distributed as well:

$$\int_S \Pr\{\boldsymbol{X}^1 \in B^0 \,|\, \boldsymbol{X}^0 = x\} \, \mathrm{d}x = \mu(B^0). \tag{14}$$

(ii) The Markov chain is μ-*recurrent*, i.e. B^0 is visited at least once with probability 1:

$$\Pr\{\exists i \in \mathbb{N}^+ : \boldsymbol{X}^i \in B^0 \,|\, \boldsymbol{X}^0 = x^0\} = 1. \tag{15}$$

To prove (i) and (ii), we first derive the expression for the *transition density function* $p : S^0 \times S^0 \to \mathbb{R}^+ \cup \{0\}$. For $x, y \in S^0$ $(x \neq y)$, let H_y denote a d-dimensional hypercube with centre y and volume h^d, oriented along the ray from x to y. Then the transition density of x, y is defined as

$$p(x, y) = \lim_{h \downarrow 0} \frac{\Pr\{\boldsymbol{X}^1 \in H_y \,|\, \boldsymbol{X}^0 = x\}}{h^d}. \tag{16}$$

Denote the surface area of a d-dimensional hypersphere with radius r by $c(r)$, and let $m(x, y)$ be the diameter of S measured along the ray from x to y. Then, since the hypersphere directions algorithm in x generates its direction vector uniformly on the (unit) hypersphere with centre x, and chooses the next interior point uniformly on the intersection of this direction vector with S we have (cf. Figure 1):

$$\lim_{h \downarrow 0} \frac{\Pr\{\boldsymbol{X}^1 \in H_y \,|\, \boldsymbol{X}^0 = x\}}{h^d} = \lim_{h \downarrow 0} \frac{1}{h^d} \frac{2h^{d-1}}{c(\|y - x\|)} \frac{h}{m(x, y)} = \frac{2}{c(\|y - x\|)m(x, y)}. \tag{17}$$

Hence,

$$\begin{aligned}
\int_S \Pr\{\boldsymbol{X}^1 \in B^0 \,|\, \boldsymbol{X}^0 = x\} \, \mathrm{d}x &= \int_S \int_{B^0} p(x, y) \, \mathrm{d}y \, \mathrm{d}x \\
&= \int_S \int_{B^0} \frac{2}{c(\|y - x\|) \cdot m(x, y)} \, \mathrm{d}y \, \mathrm{d}x \\
&= \int_{B^0} \int_S \frac{2}{c(\|x - y\|) \cdot m(y, x)} \, \mathrm{d}x \, \mathrm{d}y \\
&= \int_{B^0} \Pr\{\boldsymbol{X}^1 \in S \,|\, \boldsymbol{X}^0 = y\} \, \mathrm{d}y = \int_{B^0} \mathrm{d}y = \mu(B^0),
\end{aligned} \tag{18}$$

which proves (i). We observe that for this proof it is sufficient that a transition density function exists, and that it is symmetric in its arguments.
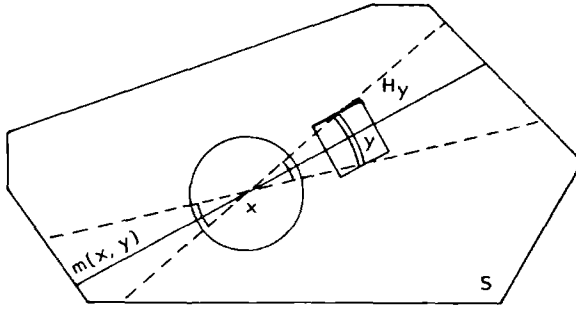
Fig. 1.

To prove (ii), let

$$r_S \triangleq \sup_{x, y \in S} \|y - x\|. \tag{19}$$

Since S is assumed to be bounded, r_S is finite. Hence

$$\Pr\{X^1 \in B^0 \mid X^0 = x^0\} = \int_{B^0} p(x^0, y) \, dy$$

$$= \int_{B^0} \frac{2}{c(\|y - x^0\|) m(x^0, y)} \, dy \geq \int_{B^0} \frac{2}{c(r_S) r_S} \, dy$$

$$= \frac{2\mu(B^0)}{c(r_S) r_S}. \tag{20}$$

Thus, since the lower bound (20) is independent of x^0 we have

$$\Pr\{\exists i \in \mathbb{N}^+ : X^i \in B^0 \mid X^0 = x^0\} = 1 - \Pr\{X^i \notin B^0 \; \forall i \in \mathbb{N}^+ \mid X^0 = x^0\}$$

$$= 1 - \lim_{j \to \infty} \Pr\{X^i \notin B^0 \; \forall i = 1, \dots, j \mid X^0 = x^0\}$$

$$\geq 1 - \lim_{j \to \infty} \left(1 - \frac{2\mu(B^0)}{c(r_S) r_S}\right)^j = 1. \tag{21}$$

In addition to x^0 and B^0, we define H^0 to be a hypercube with $\mu(H^0) > 0$, which is fully contained in S, and whose edges h_i $(i = 1, \dots, d)$ are oriented along the coordinate axes e_i $(i = 1, \dots, d)$.

Theorem 3. *The coordinate directions algorithm generates a sequence of interior points whose limiting distribution is uniform on S:*

$$\lim_{n \to \infty} \Pr\{X^n \in B^0 \mid X^0 = x^0\} = \mu(B^0). \tag{22}$$

Proof. Analogously to the proof of Theorem 2 we proceed by proving Propositions (i) and (ii).

We note, however, that given the location x of the n-th interior point, X^{n+1} will be contained in the set of coordinate axes through x. This set is of μ-probability 0,

which implies that for the coordinate directions algorithm the 1-step transition density function is not defined. Hence, to prove (i) we cannot apply the simple method of proof of Theorem 2.

To prove Proposition (i) it is sufficient to show that (cf. (14))

$$\int_S \Pr\{X^1 \in H^0 | X^0 = x\}\, dx = \mu(H^0). \tag{23}$$

Define $T_i \subset S$ as the set of points starting from which H^0 can be reached in one step if a move is made along e_i $(i = 1, \ldots, d)$. Then (cf. Figure 2):

$$\int_S \Pr\{X^1 \in H^0 | X^0 = x\}\, dx$$

$$= \sum_{i=1}^d \int_{T_i - H^0} \Pr\{X^1 \in H^0 | X^0 = x\}\, dx + \int_{H^0} \Pr\{X^1 \in H^0 | X^0 = x\}\, dx$$

$$= \sum_{i=1}^d \int_{T_i - H^0} \Pr\{X^1 \in H^0, v^0 = \pm e_i | X^0 = x\}\, dx$$

$$+ \sum_{i=1}^d \int_{H^0} \Pr\{X^1 \in H^0, v^0 = \pm e_i | X^0 = x\}\, dx$$

$$= \sum_{i=1}^d \int_{T_i} \Pr\{X^1 \in H^0, v^0 = \pm e_i | X^0 = x\}\, dx. \tag{24}$$

Choose $y \in T_i$, and define F_y^i as the intersection of S with the straight line through y along direction e_i. Then, for all $x \in F_y^i$,

$$\Pr\{X^1 \in H^0, v^0 = \pm e_i | X^0 = x\} = \frac{\|h_i\|}{d \|F_y^i\|}, \tag{25}$$

which does not depend on x. Hence, if x_i denotes the i-th component of x, then

$$\int_{F_y^i} \Pr\{X^1 \in H^0, v^0 = \pm e_i | X^0 = x\}\, dx_i = \int_{F_y^i} \frac{\|h_i\|}{d \|F_y^i\|}\, dx_i = \frac{\|h_i\|}{d \|F_y^i\|} \|F_y^i\| = \frac{\|h_i\|}{d}. \tag{26}$$
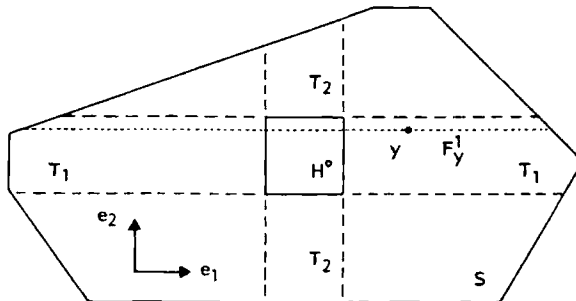

Fig. 2.

Thus, integration over all remaining components $j \neq i$ gives

$$\int_{T_i} \Pr\{X^1 \in H^0, v^0 = \pm e_i \mid X^0 = x\} \, dx = \frac{\sum_{j=1}^{d} \|h_j\|}{d}, \tag{27}$$

and, substitution of (27) in (24) yields

$$\int_S \Pr\{X^1 \in H^0 \mid X^0 = x\} \, dx = \prod_{i=1}^{d} \|h_i\| = \mu(H^0), \tag{28}$$

which proves (i).

From (21) it should be clear that the proof of Proposition (ii) follows easily if

$$\inf_{x \in S^0} \Pr\{X^1 \in B^0 \mid X^0 = x\} > 0. \tag{29}$$

However, as only moves along the direction of the coordinate vectors are allowed, (29) is obviously not true. Furthermore, one can easily devise polytopes S for which (29) does not hold if instead of one perpendicular step any finite number n of such steps may be taken, so that this problem cannot be remedied by considering n step transition probabilities either.

A different approach is therefore necessary to prove Proposition (ii). Our proof will be based on Proposition 5.1 in Orey (1971), in which the initial starting point is assumed to be a random variable with distribution η rather than being fixed at x^0. Let $\Pr_\eta\{C\}$ stand for the probability of an event C, given that X^0 is distributed according to η, and choose a real scalar $\alpha^0 > 0$. Then Orey's Proposition states that, if a set $A \subset S$ exists (possibly depending on B^0 and α^0) such that

$$\Pr_\eta\{\{X^i\}_{i=0}^{\infty} \in A \text{ infinitely often}\} > 1 - \alpha^0, \tag{30}$$

and

$$\inf_{x \in A} \Pr\{\exists i \in \mathbb{N}^+ : X^i \in B^0 \mid X^0 = x\} > 0, \tag{31}$$

then

$$\Pr_\eta\{\{X^i\}_{i=0}^{\infty} \in B^0 \text{ infinitely often}\} > 1 - \alpha^0. \tag{32}$$

This implies in turn that

$$\Pr_\eta\{\{X^i\}_{i=0}^{\infty} \in B^0 \text{ infinitely often}\} = 1, \tag{33}$$

and a fortiori

$$\Pr_\eta\{\exists i \in \mathbb{N}^+ : X^i \in B^0\} = 1. \tag{34}$$

Let the set A_ε be defined as the subset of those points in S whose distance to the boundary of S is greater than $\varepsilon > 0$. Then, with Orey's Proposition in mind, we will

prove the following statements I and II, under the additional assumption that η is bounded from above by μ, i.e. there exists a constant $c > 0$ such that $\eta(E) \leq c\mu(E)$ for all Borel sets E contained in S.

    I. There exists an $\varepsilon(B^0)$ such that for all $\varepsilon \leq \varepsilon(B^0)$, $A = A_\varepsilon$ satisfies (31).

    II. There exists an $\varepsilon(\alpha^0)$ such that for all $\varepsilon \leq \varepsilon(\alpha^0)$, $A = A_\varepsilon$ satisfies (30).

If I and II are true, then $A = A_\varepsilon$ with $\varepsilon = \min\{\varepsilon(B^0), \varepsilon(\alpha^0)\}$ simultaneously satisfies (30) and (31), which then proves (34). We then have proved (ii) for the case that the starting point follows an initial distribution which is bounded from above by μ. It then still remains necessary to extend (34) to the case that the process is started in x^0.

We will first make that final step. The proof consists of first performing n steps of the procedure, starting from x^0. The distribution of X^n then will serve as the initial distribution η, and the result desired is obtained by applying (33), assuming that I and II are satisfied.

By the Lebesgue Decomposition Theorem (Ash, 1972) the n step transition probability distribution can be decomposed into a *singular* part, v_n say, whose probability mass is concentrated on a set with μ-probability 0, and an *absolutely continuous* part, ω_n say, with an integrable function $p''(x^0, \cdot)$ so that for all Borel sets E in S

$$\omega_n(E) = \int_E p''(x^0, y) \, dy. \tag{35}$$

Clearly the Markov chain is concentrated on a set of 0 μ-probability if not all d distinct coordinate vectors have already been generated as random direction. If $n \geq d$ it is easily verified that the probability of this event does not exceed $d(1 - (1/d))^n$, so that also

$$v_n(S) \leq d\left(1 - \frac{1}{d}\right)^n \tag{36}$$

for $n \geq d$. Now define

$$\bar{\omega}_n \triangleq \frac{\omega_n}{1 - v_n(S)}. \tag{37}$$

Then $\bar{\omega}_n(S) = 1$, so that $\bar{\omega}_n$ is a probability distribution with a probability density function. From the description of the coordinate directions algorithm and from (35) it follows that $\bar{\omega}_n(E)$ is determined by n successive 1-dimensional integrations of constant functions over line segments through $x^0, X^1, \ldots, X^{n-1}$ whose function value is the inverse of the length of the line segments. Given any n and our starting point x^0, these line segments are bounded from below by a positive constant, so that $\bar{\omega}_n$ is bounded from above by μ. Hence, assuming I and II, (33) is satisfied for $\eta = \bar{\omega}_n$. Furthermore, for each Borel set E contained in S

$$\Pr\{X^n \in E \mid X^0 = x^0\} = \omega_n(E) + \upsilon_n(E) = \omega_n(E) \frac{1 - \upsilon_n(S)}{1 - \upsilon_n(S)} + \upsilon_n(E)$$

$$\geq \bar{\omega}_n(E)(1 - \upsilon_n(S)). \tag{38}$$

Hence, by (36) and (33) we have, for all $n \geq d$,

$$\Pr\{\{X^i\}_{i=0}^{\infty} \in B^0 \text{ infinitely often} \mid X^0 = x^0\}$$

$$= \int_S \Pr\{X^n \in dy \mid X^0 = x^0\} \Pr\{\{X^i\}_{i=0}^{\infty} \in B^0 \text{ infinitely often} \mid X^0 = y\}$$

$$\geq (1 - \upsilon_n(S)) \int_S \bar{\omega}_n(dy) \Pr\{\{X^i\}_{i=0}^{\infty} \in B^0 \text{ infinitely often} \mid X^0 = y\}$$

$$= (1 - \upsilon_n(S)) \Pr_{\bar{\omega}_n}\{\{X^i\}_{i=0}^{\infty} \in B^0 \text{ infinitely often}\}$$

$$\geq \left(1 - d\left(1 - \frac{1}{d}\right)^n\right) \Pr_{\bar{\omega}_n}\{\{X^i\}_{i=0}^{\infty} \in B^0 \text{ infinitely often}\}$$

$$= 1 - d\left(1 - \frac{1}{d}\right)^n, \tag{39}$$

so that

$$\Pr\{\{X^i\}_{i=0}^{\infty} \in B^0 \text{ infinitely often} \mid X^0 = x^0\} = 1. \tag{40}$$

and a fortiori

$$\Pr\{\exists i \in \mathbb{N}^+ : X^i \in B^0 \mid X^0 = x^0\} = 1. \tag{41}$$

This proves (ii) for the case that I and II are satisfied, so that it remains to establish the latter facts.

We recall that to prove I, we have to show that there is an $\varepsilon(B^0)$ such that for all $\varepsilon \leq \varepsilon(B^0)$

$$\inf_{x \in A_\varepsilon} \Pr\{\exists i \in \mathbb{N}^+ : X^i \in B^0 \mid X^0 = x\} > 0. \tag{42}$$

Choose any $\varepsilon > 0$, and $y, z \in A_\varepsilon$ with $\|y - z\| \leq \varepsilon$. Let H_ε be a hypercube with centre at the origin and edges of length $\varepsilon/\sqrt{d}$ oriented along the coordinate axes, and let $\bar{H} \subset H_\varepsilon$ be an arbitrary hypercube whose edges are parallel to the coordinate axes as well. Now construct the hyperrectangle R_{yz} oriented along the coordinate axes, which contains y and z as vertices. Then, since S is convex, and each point in A_ε is at distance at least ε from the boundary of S, it is easily verified that $s + H_\varepsilon \subset S$ for each vertex s (including y and z) of R_{yz}. The edges of R_{yz} are perpendicular. Thus, if the vertices s^1 and s^2 only differ in the j-th coordinate, we can reach each point in $s^2 + H_\varepsilon$ from each point in $s^1 + H_\varepsilon$ by a move along the direction e_j $(j = 1, \ldots, d)$. Hence, since the coordinate directions algorithm searches along one of the coordinate vectors with equal probability, and since a next interior point is chosen uniformly on the linesegment connecting the two previous hitpoints, we have that, for all $x \in y + H_\varepsilon$,

$$\Pr\{X^d \in z + \bar{H} \mid X^0 = x\} \geq \frac{\mu(\bar{H})}{(dr_S)^d} \tag{43}$$

where r_S is the maximal diameter of S (cf. definition (19)).

Next, let y and z be arbitrary points in A_ε, not necessarily satisfying $\|y - z\| \leq \varepsilon$. It is clear that we can choose $q \triangleq r_s / \varepsilon$ points $x^1, \ldots, x^q$ on the straight line l connecting y and z, such that each two successive points are at distance at most ε. Furthermore, since S is convex, A_ε is convex, and $l \in A_\varepsilon$. Thus, applying (43) $q + 1$ times, we find that for all $x \in y + H_\varepsilon$, and, for all Borel sets E such that $\mu(E \cap z + H_\varepsilon) > 0$,

$$\begin{aligned}
&\Pr\{X^{(q+1)d} \in E \mid X^0 = x\} \\
&\geq \Pr\{X^{(q+1)d} \in E \cap z + H_\varepsilon, \ X^{id} \in x^i + H_\varepsilon; \quad i = 1, \ldots, q \mid X^0 = x\} \\
&\geq \delta\mu(E \cap z + H_\varepsilon) > 0
\end{aligned} \tag{44}$$

with

$$\delta = \frac{1}{(dr_S)^d} \left(\frac{\mu(H_\varepsilon)}{(dr_S)^d} \right)^q. \tag{45}$$

Since the above lower bound does not depend on x, it follows that for all Borel sets E with $\mu(E \cap A_\varepsilon + H_\varepsilon) > 0$ that

$$\inf_{x \in A_\varepsilon} \Pr\{\exists i \in \mathbb{N}^+ \colon X^i \in E \mid X^0 = x\} > 0. \tag{46}$$

We now return to our fixed Borel set B^0 with $\mu(B^0) > 0$. It is immediate that there is an $\varepsilon(B^0)$ such that, for all $\varepsilon \leq \varepsilon(B^0)$, $\mu(B^0 \cap A_\varepsilon + H_\varepsilon) > 0$. Hence, (46) is satisfied for $E = B^0$, which proves I.

Finally, we will prove II, i.e. if η is bounded from above by μ, then for our fixed α^0, there exists an $\varepsilon(\alpha^0)$ such that for all $\varepsilon \leq \varepsilon(\alpha^0)$

$$\Pr_\eta\{\{X^i\}_{i=0}^\infty \in A_\varepsilon \text{ infinitely often}\} > 1 - \alpha^0. \tag{47}$$

Since $\mu(S) < \infty$, we can choose $\varepsilon(\alpha^0)$ such that $c\mu(S - A_{\varepsilon(\alpha^0)}) < \alpha^0$. Since μ is invariant this choice implies that, if $\varepsilon \leq \varepsilon(\alpha^0)$,

$$\Pr_\eta\{X^n \in S - A_\varepsilon\} \leq c\mu(S - A_\varepsilon) < \alpha^0 \tag{48}$$

for all n. Thus, if

$$N \triangleq \sup_{n \geq d}\{n \mid X^n \in A_\varepsilon\} \tag{49}$$

then, by (48),

$$\Pr_\eta\{N < n\} \leq \Pr_\eta\{X^n \in S - A_\varepsilon\} < \alpha^0 \tag{50}$$

for all $n \geq d$, so that

$$\Pr_\eta\{N = \infty\} > 1 - \alpha^0 \tag{51}$$

which proves II.

4. A Bayesian stopping criterion

Let the number of nonredundant constraints of a system of m linear inequalities be $k \le m$. We already observed that unless $k = m$ and m distinct nonredundant constraints have been found, it remains uncertain if all nonredundant constraints have been identified. Let $\bar{n}$ be the number of hitpoints, and define w as the number of *distinct* nonredundant constraints which have been found in the course of these trials. (Notice that after n iterations of an hit and run algorithm $\bar{n} = 2n$). In this section we will develop a *Bayesian stopping criterion* which determines for each pair $(\bar{n}, w)$ if the search procedure should be terminated, or not.

Assume that the k nonredundant constraints are labeled $1, \ldots, k$, and for each $x \in S^0$ define $\xi_i(x)$ as the subset of the hypersphere with centre x, such that a search from x in the direction of $\xi_i(x)$ will yield a hitpoint on the nonredundant constraint with index i ($i = 1, \ldots, k$). Denote the $(d-1)$-dimensional Lebesgue measure by $L_{d-1}\{\cdot\}$, and assume without loss of generality that the $(d-1)$-dimensional Lebesgue measure of the surface area of the above hypersphere is equal to 1. Then the probability that the $\bar{n}$-th search of the *hypersphere directions algorithm* will yield a hitpoint on the i-th nonredundant constraint is given by

$$\theta_i^H(\bar{n}) = \int_S L_{d-1}\{\xi_i(x)\} \Pr\{X^n \in dx \mid X^0 = x^0\} \quad (i = 1, \ldots, k). \tag{52}$$

Next, define $G_{ij} \subset S$ as the set of points in S from which the i-th nonredundant constraint can be found by a search along direction e_j ($i = 1, \ldots, k; j = 1, \ldots, d$). Then, analogously to the above reasoning, we obtain for the *coordinate directions algorithm* that the hitting probabilities are given by

$$\theta_i^C(\bar{n}) = \frac{1}{2d} \sum_{j=1}^{d} \Pr\{X^n \in G_{ij} \mid X^0 = x^0\} \quad (i = 1, \ldots, k). \tag{53}$$

Hence, applying Theorems 2 and 3 to respectively (52) and (53) we obtain that given $X^0 = x^0$

$$\lim_{\bar{n} \to \infty} \theta_i^H(\bar{n}) = \int_S L_{d-1}\{\xi_i(x)\} \mu(dx) \quad (i = 1, \ldots, k) \tag{54}$$

and

$$\lim_{\bar{n} \to \infty} \theta_i^C(\bar{n}) = \frac{1}{2d} \sum_{j=1}^{d} \mu(G_{ij}) \quad (i = 1, \ldots, k). \tag{55}$$

Hence, the hitting probabilities are asymptotically *fixed*; the convergence rates are addressed in Smith (1984). For the two versions of the hit-and-run algorithm we denote these limiting values θ_i^H and θ_i^C respectively; observe that θ_i^H is not necessarily equal to θ_i^C ($i = 1, \ldots, k$). In order to be able to proceed we will exploit this result by assuming that the actual hitting probability of nonredundant constraint i at each trial of the hypersphere and coordinate directions algorithm is equal to the

corresponding asymptotic probability θ_i^H and θ_i^C respectively $(i = 1, \ldots, k)$. Then, if the values of these probabilities would be given, several proper stopping rules are obvious. In that case, a procedure could be terminated, for example, if the total probability of the observed nonredundant constraints exceeds a prescribed value, or if the total probability of the observed nonredundant constraints is equal to 1, i.e. if all k nonredundant constraints have been found. The true number of nonredundant constraints of a system of linear inequalities, and a fortiori the corresponding hitting probabilities, are of course frequently unknown. Therefore we adopt a *statistical* approach in which the data produced by an hit-and-run algorithm are used to gain information about their values. Our starting point is the assumption that the actual hitting probabilities at each trial are equal to the asymptotic probabilities. Then the output of an algorithm is a sample from a *multinomial distribution*: each *cell* of the distribution corresponds to a nonredundant constraint, and the *cell probabilities* are equal to the corresponding hitting probabilities. Thus, if we make no further notational distinction between θ_i^H and θ_i^C $(i = 1, \ldots, k)$, the joint probability that in $\bar{n}$ searches the ith nonredundant constraint will be found n_i times $(i = 1, \ldots, k)$ is equal to

$$p(n_1, \ldots, n_k \,|\, k, \theta_1, \ldots, \theta_k) = \frac{\bar{n}!}{\prod_{i=1}^{k} n_i!} \prod_{i=1}^{k} \theta_i^{n_i} \tag{56}$$

$(\sum_{i=1}^{k} \theta_i = 1, \sum_{i=1}^{k} n_i = \bar{n})$. Given a sample $(n_1, \ldots, n_k)$ the multinomial formula (56) would enable us to learn about the values of the unknowns $(k, \theta_1, \ldots, \theta_k)$. However, since it is unknown in advance which of the m constraints are nonredundant, it is impossible to distinguish between different samples $(n_1, \ldots, n_k)$ up to a *relabeling* of the nonredundant constraints. For example, if in $\bar{n} = 5$ searches one nonredundant constraint has been found 4 times and another one once, it is unknown whether we observed $(n_1, n_2) = (4, 1)$, $(n_1, n_2) = (1, 4)$ or $(n_1, n_2, n_3, n_4, n_5, n_6, n_7) = (0, 4, 0, 1, 0, 0, 0)$ etc. Thus we have to restrict ourselves to *distinguishable aggregates* of the sample outcomes $(n_1, \ldots, n_k)$ that are independent of the labeling of the nonredundant constraints and which do not contain n_i's which are equal to 0. Given the result of $\bar{n}$ searches we denote the appropriate aggregates by $\{n_1, \ldots, n_w\}$ (recall that w is the number of observed nonredundant constraints). The required probability of these aggregates $\{n_1, \ldots, n_w\}$ is given by the *generalized multinomial distribution*.

Theorem 4 (Boender and Rinnooy Kan, 1983a; Boender, 1984). *Let a system of linear inequalities be given with k nonredundant constraints with probabilities $\theta_1, \ldots, \theta_k$. Then the probability that in $\bar{n}$ trials w different nonredundant constraints are found, of which one constraint is found n_1 times, another constraint n_2 times etc., is given by the generalized multinomial distribution*

$$p(\{n_1, \ldots, n_w\} \,|\, k, \theta_1, \ldots, \theta_k) = \frac{1}{\prod_{j=1}^{\bar{n}} c_j!} \frac{\bar{n}!}{\prod_{i=1}^{w} n_i!} \sum_{(g_1, \ldots, g_w) \in S_k[w]} \prod_{i=1}^{w} \theta_{g_i}^{n_i} \tag{57}$$

$(\sum_{i=1}^{k} \theta_i = 1, \sum_{i=1}^{w} n_i = \bar{n}, n_i > 0; \ i = 1, \ldots, w)$, where

$\quad c_j \triangleq$ the number of n_i's that are equal to j $(j = 1, \ldots, \bar{n})$, $\hspace{2cm}$ (58)

$\quad S_k[w] \triangleq$ the set of all permutations of w different elements of
$\qquad$ the set $\{1, \ldots, k\}$. $\hspace{6cm}$ (59)

Now, (57) can be used in a *Bayesian* approach in which the unknowns $k, \theta_1, \ldots, \theta_k$ are assumed to be themselves random variables $K, \theta_1, \ldots, \theta_K$ for which a *prior distribution* can be specified. Given the result of an hit-and-run algorithm, Bayes' rule is used to compute the *posterior distribution* of K, which incorporates both the prior beliefs and the sample information about the true number of nonredundant constraints.

Theorem 5 (Boender and Rinnooy Kan, 1983b; Boender, 1984). *Under the assumption of an arbitrary prior distribution $p(\cdot)$ for the number of nonredundant constraints K, and, conditional on $K = k$, a symmetric Dirichlet prior with hyperparameter α for the hitting probabilities $\theta_1, \ldots, \theta_k$, i.e.*

$$p(\theta_1, \ldots, \theta_k \mid k) = \frac{(\alpha k - 1)!}{((\alpha - 1)!)^k} \prod_{i=1}^{k} \theta_i^{\alpha - 1}, \hspace{3cm} (60)$$

the marginal posterior distribution of the true number of nonredundant constraints, conditional on an observed aggregate $\{n_1, \ldots, n_w\}$, is equal to

$$p(k \mid \{n_1, \ldots, n_w\}) \propto p(k) \frac{(\alpha k - 1)! \, k!}{(\bar{n} + \alpha k - 1)! (k - w)!}, \hspace{2cm} (61)$$

where $\propto$ denotes proportionality.

Our stopping rule will be based on the knowledge about the number of nonredundant constraints contained in the posterior distribution (61). Observe that (61) is independent of the number of times n_i that each of the observed nonredundant constraints has been found $(i = 1, \ldots, w)$, but only involves the total number of hitpoints $\bar{n}$ and the number of distinct observed nonredundant constraints w, so that also our stopping rule only depends on the pair $(\bar{n}, w)$.

Before we can apply (61) it remains to choose a proper prior distribution. Since a system of m linear inequalities in a d dimensional space consists of at least $d + 1$, and at most of m nonredundant constraints, the range of the prior for K is $d + 1, \ldots, m$. The prior probability that the true number of nonredundant constraints K is equal to k is assumed to grow linearly with k. Thus

$$p(k) \propto k \quad (k = d + 1, \ldots, m). \hspace{4cm} (62)$$

Note that in a Bayesian context a prior distribution is frequently chosen uniform. Since our analysis is based on the assumption that the actual hitting probabilities

are equal to the asymptotic hitting probabilities, we deliberately chose a prior that will result in longer running times of the algorithms than would be obtained for a uniform prior for K. However, the reader need not follow our suggestion and may choose any other version for (62), based on his own system of linear inequalities.

Given $K = k$, we are only free to choose the hyperparameter α of the symmetric Dirichlet prior distribution for the hitting probabilities $\theta_1, \ldots, \theta_k$ (cf. (60)). The value of α is a measure of the deviation of the prior (60) from its expectation $E(\theta_i) = 1/k$ $(i = 1, \ldots, k)$ (cf. Wilks, 1962). For $\alpha = 1$ the symmetric Dirichlet corresponds to the uniform distribution on the unit simplex ($\sum_{i=1}^{k} \theta_i = 1$; $0 < \theta_i < 1$; $i = 1, \ldots, k$). For $\alpha > 1$ a unique maximum is attained at the expected value, whose size increases if α gets larger: if $\alpha = \infty$ then all probabilities are a priori assumed to be equal to $1/k$ with probability 1: For $0 < \alpha < 1$ the distribution attains a unique minimum at the expected value which decreases as α approaches 0. Preliminary diagnostic experiments showed that the posterior (61), and a fortiori our stopping rule, are sensitive to the choice for α. Hence, a user who is uncertain about the correct value for α is in a difficult position. To cope with this problem we run a hit-and-run algorithm a certain number of iterations and estimate α by Good's formula (Good, 1965):

$$\hat{\alpha} = \frac{1 - \sum_{i=1}^{w} \left(\frac{n_i}{\bar{n}}\right)^2}{-1 + w \sum_{i=1}^{w} \left(\frac{n_i}{\bar{n}}\right)^2}.$$

(63)

Then this estimate is used as if it is the true value corresponding to the system of linear inequalities under investigation.

Now, given a choice of the prior distribution and an observed sample of hitpoints we can compute the posterior distribution of the true number of nonredundant constraints from (61). Then we can easily calculate the *posterior expected value* of the number of nonredundant constraints, which is well known to be the *optimal Bayesian estimate* with respect to a *quadratic loss function* (Lindley, 1978). For most $(\bar{n}, w)$ pairs, however, this may yield a real valued estimate, whereas the true number of nonredundant constraints is evidently an integer. Therefore, since it easily shown that the optimal integer Bayesian estimate under a quadratic loss function is the round-off of the real valued estimate, we will terminate our algorithms when this round-off of the real valued optimal Bayesian estimate of the number of nonredundant constraints is equal to the number of distinct nonredundant constraints observed. That is, the algorithms are stopped when the current $(\bar{n}, w)$ pair satisfies

$$E(K|(\bar{n}, w)) = \frac{\sum_{k=\max\{d+1,w\}}^{m} k^2 \dfrac{(\hat{\alpha}k - 1)! \, k!}{(\bar{n} + \hat{\alpha}k - 1)!(k - w)!}}{\sum_{k=\max\{d+1,w\}}^{m} k \dfrac{(\hat{\alpha}k - 1)! \, k!}{(\bar{n} + \hat{\alpha}k - 1)!(k - w)!}} < w + \tfrac{1}{2}.$$

(64)

5. Computational results

In this section we describe the performance of the two hit-and-run algorithms and the Bayesian stopping rule on 3 practical and 3 randomly generated test problems. Note that the algorithms require an initial interior point. If not readily available, this may be obtained from a phase I procedure. This starting problem is a common burden of all nonredundancy identification procedures (cf. Karwan et al., 1983).

The most important features of our test problems are displayed in Table 1. The experimental design for the randomly generated test problems is taken from Karwan et al. (1983). The practical problems A and B are from Tischer (1968), practical problem C is from Meyerman (1966). The practical problems incorporate $x_i \geq 0$ ($i = 1, \ldots, d$) as part of the set of constraints. The density of the problems (i.e., the fraction of non-zero coefficients of the constraints) is denoted by δ.

In Table 2 the number of seconds is shown that the hit-and-run algorithms required on a DEC 2060 computer to generate $\bar{n} = 100000$ hitpoints; the table shows the estimates of the hyperparameter α as well. The Figures 3 up to 8 depict the evolution of the fraction of observed nonredundant constraints in the course of sampling: stopping times (in seconds) are shown in Table 3. HD denotes the hypersphere directions algorithm and CD is the coordinate directions algorithm.

Figures 3–8 and Tables 1–3 reveal substantial differences between the hypersphere and coordinate directions algorithm. A final issue of theoretical and practical relevance is to what extent these algorithms are successful in approximating the uniform distribution over S in a limited number of experiments. Theoretically, exponential speed of convergence was established for the hypersphere directions method in Smith (1984), and we conjecture that a similar result is true for the coordinate directions method. Experimentally, we compared their performance on two 2-dimensional polytopes, of which the first one (Figure 9) was chosen to be very disadvantageous for the coordinate directions method and the other one (Figure 10) to be very advantageous. The results for 100 and 500 iterations of both methods are depicted in the figure. In spite of its simplicity, the coordinate directions method presents a very acceptable short run picture, providing additional practical confirmation of its superiority over its competitor.

6. Concluding remarks

Our experiments show that the hypersphere directions algorithm is inferior to its rival with respect to the required computer time to generate a given number of hitpoints as well as with respect to the number of identified nonredundant constraints per hitpoint. We will therefore in this section refer only to the coordinate directions algorithm. Although the best deterministic method does outperform this method on problems of small size (Karwan et al., 1983), it can be applied to (very) large

Table 1

Test problems

Problem	m	d	k	δ
A	91	26	52	0.33
B	29	5	11	0.78
C	59	23	54	0.15
D	20	10	19	1.00
E	20	10	18	0.50
F	30	10	18	0.50

Table 2

Computation times for 100 000 hitpoints, and estimates for α

Problem	Algorithm	Computation time	$\hat{\alpha}$
A	HD	1747	0.62
	CD	61	4.18
B	HD	315	1.55
	CD	49	8.65
C	HD	721	0.49
	CD	20	1.02
D	HD	577	0.71
	CD	24	0.71
E	HD	572	1.02
	CD	24	0.79
F	HD	813	1.01
	CD	45	0.79

Table 3

Stopping times

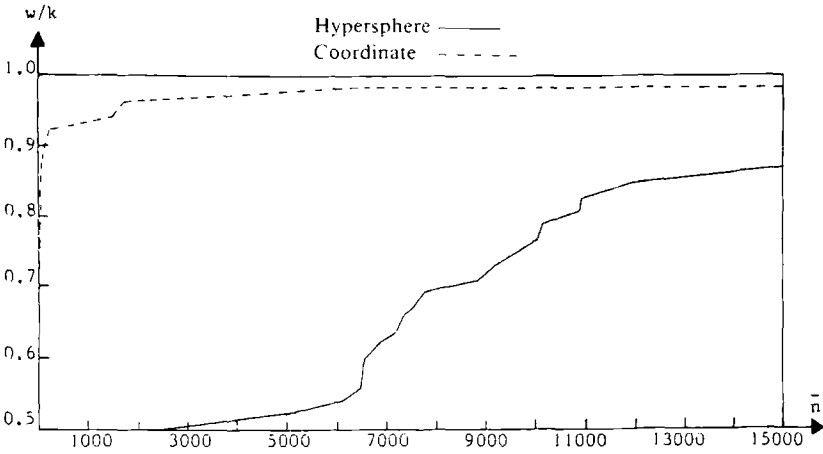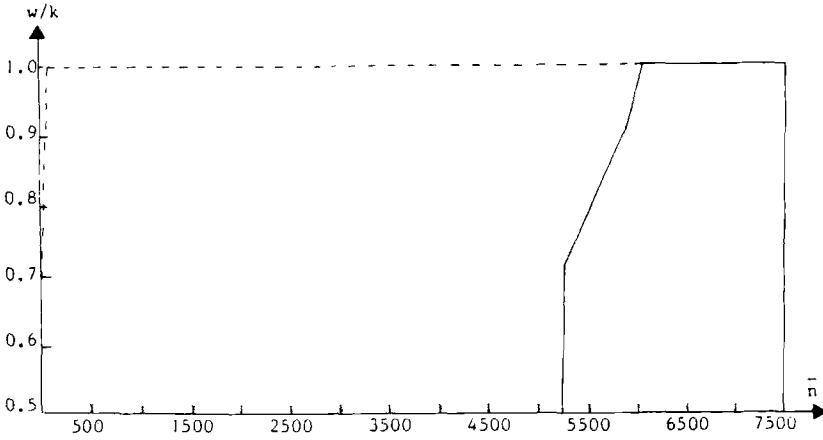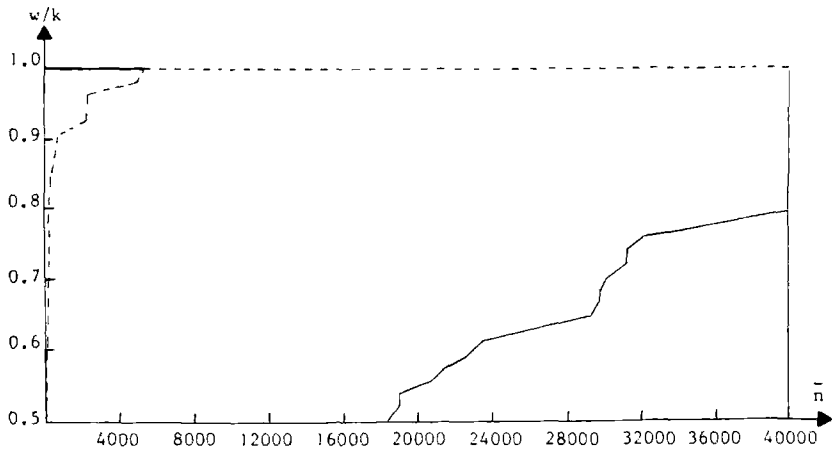| Problem	Algorithm	Time	$\bar{n}$	w
A	HD	1041	59 600	51
$k = 52$	CD	0.30	500	48
B	HD	16.7	5300	7
$k = 11$	CD	0.05	100	11
C	HD	721	100 000	49
$k = 54$	CD	1.1	5400	54
D	HD	5.2	900	11
$k = 19$	CD	0.45	2000	18
E	HD	1.7	300	11
$k = 18$	CD	0.26	1100	15
F	HD	2.4	300	11
$k = 18$	CD	0.5	1100	15

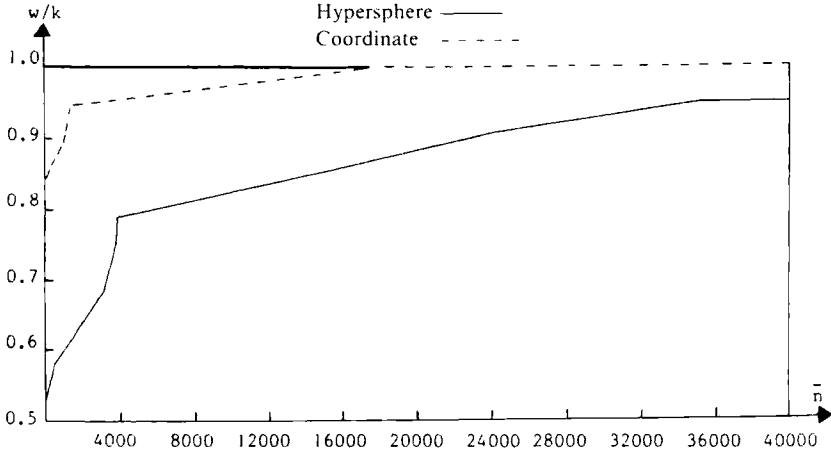Fig. 3. Problem A.

Fig. 4. Problem B.
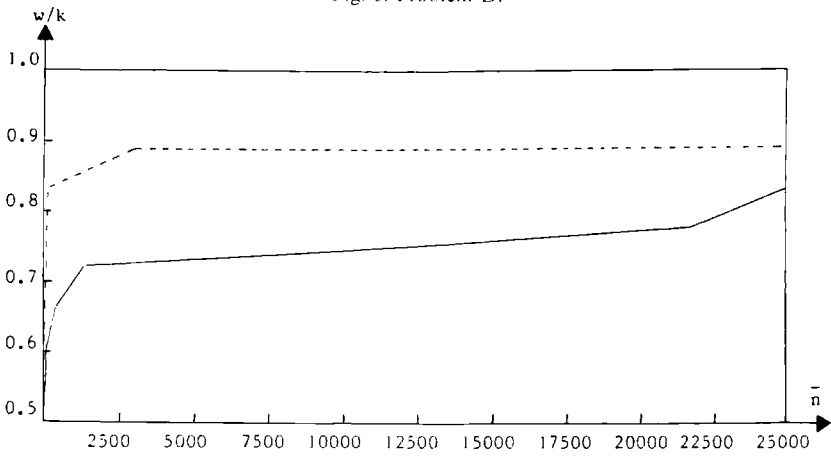

Fig. 5. Problem C.

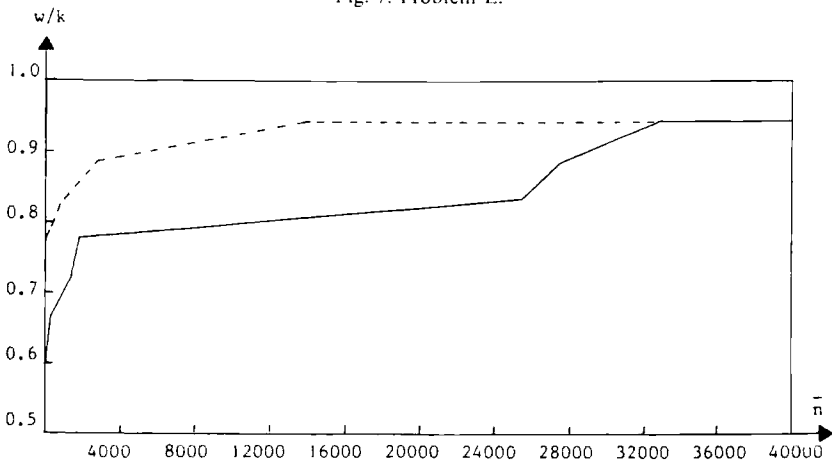Fig. 6. Problem D.
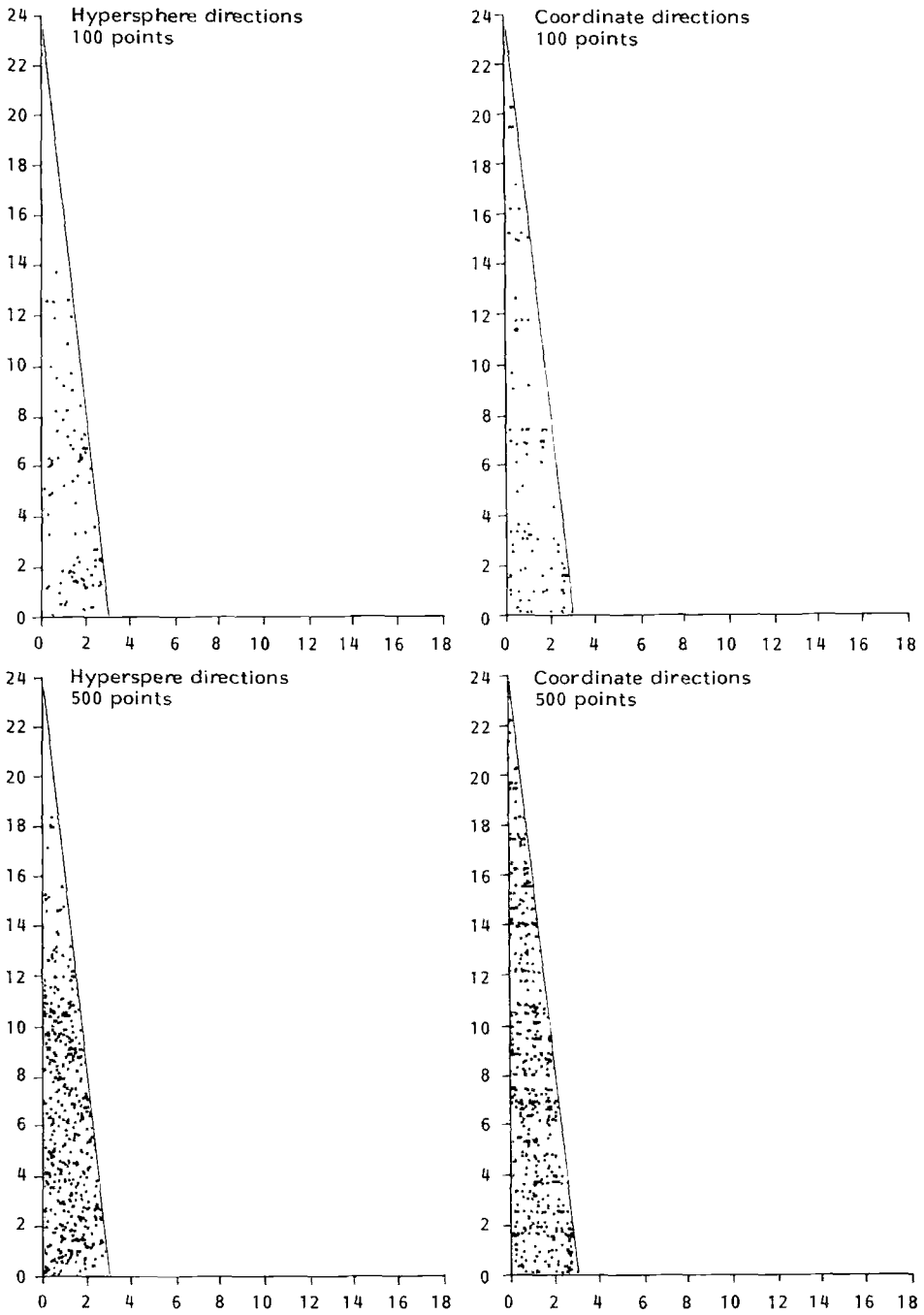

Fig. 7. Problem E.

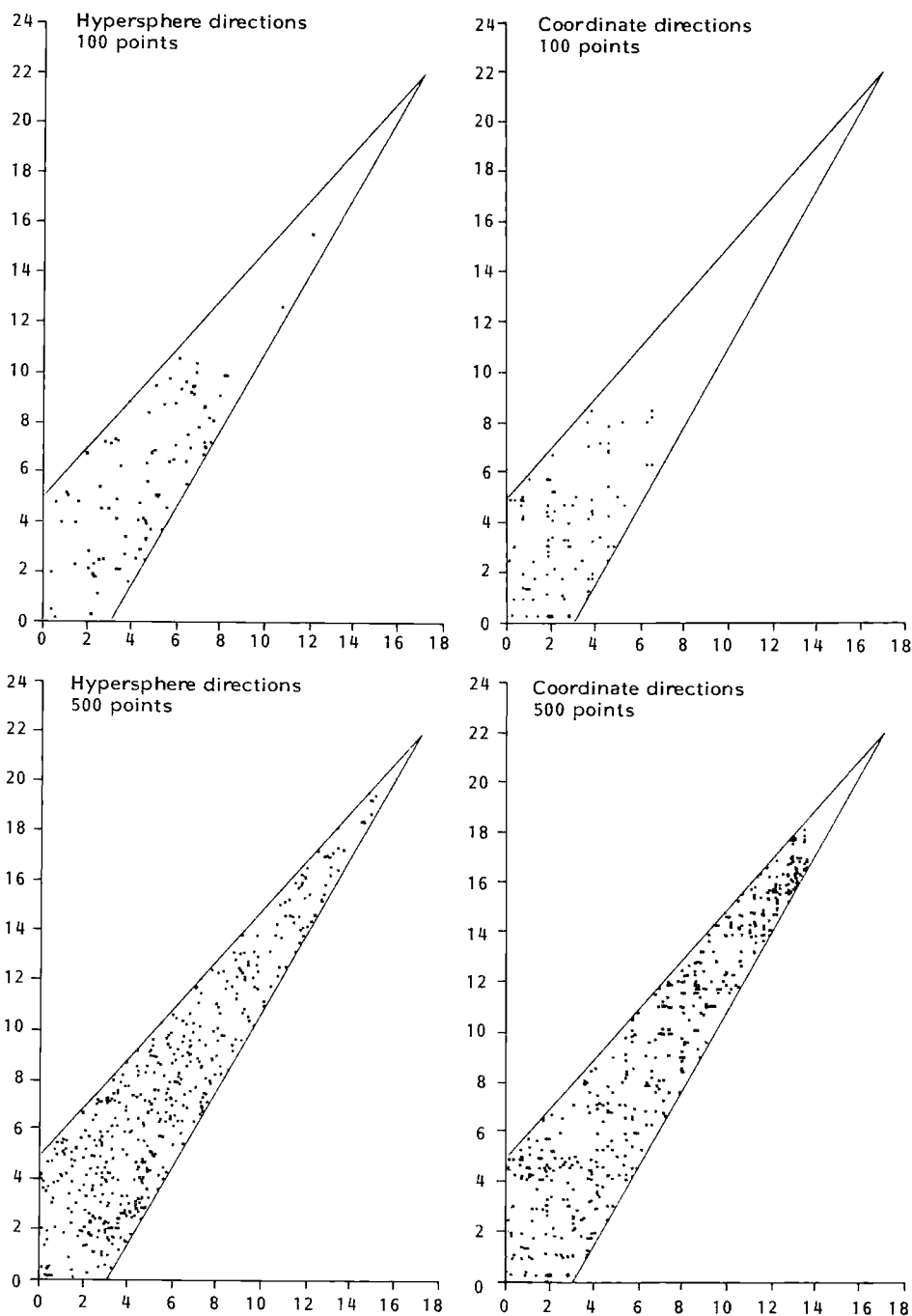Fig. 8. Problem F.

Fig. 9.

Fig. 10.

problems and therefore was judged in Karwan et al. (1983) to be one of the most attractive practical possibilities to test for redundancy.

So far the random method has been viewed as a preprocessor that is capable of identifying nonredundant constraints. In an optimization algorithm for (possibly nonlinear) objective functions with linear inequality constraints, the nonidentified constraints can be omitted. Obviously, the solution produced by the optimization algorithm has to be checked for feasibility afterwards, since the omission of some constraints may not be justified. In case of linear programming, feasibility can be easily restored by dual simplex steps.

It is well known that the optimal solution of a linear programming problem is attained in one of the vertices of the feasible region. Therefore for linear programming applications one may investigate the possibility of modifying the coordinate directions algorithm to move in the direction of the optimal vertex. The modified method may even converge rapidly to this optimal vertex. If not, then at least a better preprocessor will be obtained, since the nonredundant constraints which are active in the optimal vertex will have relatively greater hitting probabilities.

Finally, for nonlinear objective functions with linear inequality constraints the optimal solution may be situated anywhere in the feasible region. Since the coordinate directions algorithm involves an extremely fast method to generate (asymptotically) uniform points over such a feasible region, the method offers a starting point for various constrained *global optimization* procedures in which the generation of such points would be a first step (Timmer, 1984).

References

R.B. Ash, *Real Analysis and Probability* (Academic Press, New York, 1972).

M.L. Balinski, "An algorithm for finding all vertices of convex polyhedral sets," *Journal of the Society of Industrial Applied Mathematics* 9 (1961) 72-88.

C.G.E. Boender and A.H.G. Rinnooy Kan, "A Bayesian analysis of the number of cells of a multinomial distribution," *The Statistician* 32 (1983a) 240-248.

C.G.E. Boender and A.H.G. Rinnooy Kan, "Bayesian estimation of animal population size," Report 8322/0 Econometric Institute, Erasmus University Rotterdam (Rotterdam, 1983b).

C.G.E. Boender, "The generalized multinomial distribution: A Bayesian analysis and applications," Ph.D. Thesis, Erasmus University Rotterdam (Rotterdam, 1984).

A. Boneh and A. Golan, "Constraints redundancy and feasible region boundedness by random feasible points generated," paper presented at EURO III (1979).

A. Boneh, "A probabilistic algorithm identifying redundancy by a random feasible point generator (RFPG)," in: Karwan et al. (1983).

G.H. Bradley, G.G. Brown and G.W. Graves, "Structural redundancy in large scale optimization models," Technical Report CA 93940, Naval Postgraduate School (Monterey, 1980); also in: Karwan et al. (1983).

A.L. Brearly, G. Mitra and H.P. Williams, "Analysis of mathematical programming problems prior to applying the simplex algorithm," *Mathematical Programming* 8 (1975) 54-83.

T. Gal, "Zur Identifikation redundanter Nebenbedingungen in linearen Programmen," *Zeitschrift für Operations Research* 19 (1975) 19-28.

I.J. Good, *The Estimation of Probabilities*, Research Monograph no. 30, (The M.I.T. Press, Cambridge, MA, 1965).

M.H. Karwan, V. Lotfi, J. Telgen and S. Zionts, eds., *Redundancy in Mathematical Programming* (Springer-Verlag, Berlin, 1983).

D.V. Lindley, *Bayesian Statistics, A Review* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1978).

J. Lisy, "Metody pro nalezini redundant omezini v ulohach linearniho programovani," *Economicko Mathematicky Obzor* 7 (1971) 285–198.

T.H. Mattheis, "An algorithm for determining irrelevant constraints and all vertices in systems of linear inequalities," *Operations Research* 21 (1973) 247–260.

G.L. Meyerman, "Betekenis van een aantal cultuurtechnische faktoren voor de ontwikkelings-mogelijkheden van veenkoloniale akkerbouwbedrijven," Dissertation, Agricultural University Wageningen, The Netherlands (Wageningen, 1966).

S. Orey, *Limit Theorems for Markov Chain Transition Probabilities* (Van Nostrand, New York, 1971).

M.O. Rabin, "Probabilistic algorithms," in: J.F. Traub, ed., *Algorithms and Complexity* (Academic Press, London, 1976).

R.L. Smith, "Monte Carlo procedures for generating random feasible solutions to mathematical programs," ORSA/TIMS Conference Washington, DC (1980).

R.L. Smith, "Efficient Monte Carlo procedures for generating points uniformly over bounded regions," *Operations Research* 32 (1984) 1296–1308.

J. Telgen, *Redundancy and Linear Programs* (Mathematical Centre, Amsterdam, 1979).

J. Telgen, Private Communication with A. Boneh (1980).

G.L. Thompson, F.M. Tonge and S. Zionts, "Techniques for removing non-binding constraints and extraneous variables from linear programming problems," *Management Science* 12 (1966) 588–608.

G.T. Timmer, "Global optimization: A stochastic approach," Ph.D. Thesis, Erasmus University Rotterdam (Rotterdam, 1984).

H.J. Tischer, "Mathematische Verfahren zur Reduzierung der Zeilen und Spaltenzahl linearer Optimierungsaufgaben," Dissertation, Zentral-Institut für Fertigungstechnik des Maschinenbaues, Karl Marx Stadt, DDR (1968).

S.S. Wilks, *Mathematical Statistics* (Wiley, New York, 1962).