Chapter 4

A General Framework for the Analysis of Sets of Constraints

Richard Caron
Department of Mathematics and Statistics
University of Windsor
Tim Traynor*
Department of Mathematics and Statistics
University of Windsor
Windsor, Ontario, N9B3P4, Canada

Abstract
This paper is about the analysis of sets of constraints, with no explicit assumptions. We explore the relationship between the minimal representation problem and a certain set covering problem of [Boneh (1984)]. This provides a framework that shows the connection between minimal representations, irreducible infeasible systems, minimal infeasibility sets, as well as other attributes of the preprocessing of mathematical programs. The framework facilitates the development of preprocessing algorithms for a variety of mathematical programs. As some such algorithms require random sampling, we present results to identify those sets of constraints for which all information can be sampled with nonzero probability.

Key Words: optimization, preprocessing, redundancy, irreducible infeasible systems, set covering, minimal infeasibility sets

4.1. Introduction

We consider an indexed family $\{A_i, B_i\}$ of partitions of an abstract space X. We think of A_i as the set of points satisfied by the i^{th} constraint of an optimization problem, and $B_i = A_i^c$, the set of those that violate it. For example we could be given a family of constraint functions g_i on X and $A_i = \{x \in X : g_i(x) \leq 0\}$ and $B_i = \{x \in X : g_i(x) > 0\}$. In general, the

*Corresponding author:`tt@uwindsor.ca`

function $g_i = \delta_i$, defined by

$$\delta_i(x) = \begin{cases} 0, & x \in A_i \\ 1, & x \in B_i; \end{cases} \tag{4.1}$$

namely, the indicator (or characteristic) function of B_i, could serve as such a constraint function. Often, in applications, the A_i will be closed sets.

The **feasible set** for the family $\{A_i, B_i\}_{i \in I}$ is given by $Z(I) = \bigcap_{i \in I} A_i$. This may, or may not, be empty. If it is empty, we will say that the family is **infeasible**; otherwise, **feasible**. In either case, we are interested in subsets J of I such that $Z(J) = Z(I)$. In this situation, we call J a **reduction** of I and the family $\{A_i, B_i\}_{i \in J}$, a reduction of $\{A_i, B_i\}_{i \in I}$. The family $\{A_i, B_i\}_{i \in J}$ is **irreducible** if $J \subseteq I$ and there is no proper reduction J' of J.

In the feasible case ($Z(I) \neq \emptyset$), the search for such subsets J is equivalent to the detection of redundancy, one aspect of preprocessing. For linear programs, the importance of preprocessing has been established, for example, by [Karwan *et al.* (1983)], [Andersen and Andersen (1995)], [Bixby (1994)], and [Lustig *et al.* (1994)]. Descriptions of deterministic methods to detect redundancy can be found in [Brearly *et al.* (1975)], [Tomlin and Welch (1986)], [Karwan *et al.* (1983)], [Greenberg (1996)], and [Caron *et al.* (1989)]. Probabilistic methods are described in [Boneh (1983)], [Berbee *et al.* (1987)] and [Caron *et al.* (1990)].

In the case of infeasible ($Z(I) = \emptyset$) linear programs, the search for reductions J is equivalent to the search for irreducible infeasible systems (IIS's). The paper by [Chinneck and Dravnieks (1991)] describes the powerful IIS algorithms that are available in many professional linear programming codes. Related to the problem of finding an IIS is that of finding a minimal infeasibility set (MIS), that is, a set J of minimum cardinality such that $Z(I \setminus J) \neq \emptyset$. [Chakravarti (1994)] showed that finding an MIS is NP-hard.

According to [Caron (2001)], very little attention has been paid to nonlinear constraint sets. Some exceptions are in [Obuchowska and Caron (1995)], [Jibrin (2002)], and [Obuchowska (2000)], for quadratic, positive semidefinite, and convex programming problems, respectively. The growing importance of more general mathematical programs to very large scale engineering design problems, among others, together with strong evidence of the importance of presolve and infeasibility analysis for linear programmes, indicates the need for this deficiency to be corrected. This paper is such a contribution.

In 1984, [Boneh (1984)] introduced and exploited an equivalence between the minimal representation problem and a set covering problem to develop a tool for detecting and removing redundant constraints. His implementation involved a random sampling of points x in X each of which created a row $\delta(x)$ in a set covering matrix. With the introduction of an objective function this becomes a set covering problem. Boneh showed that while any feasible solution gives a reduction, an optimal solution produces an irreducible reduction. As the set covering problem is NP-hard, polynomial time heuristics were suggested to find near-optimal reductions.

This paper develops Boneh's equivalence further. Our initial contribution, in Section 4.2, is a new presentation of the concepts, making, we believe, the correspondence between sets of constraints and set covering problems, and the proof of key results, more transparent and shorter. It also becomes clear that the equivalence holds regardless of feasibility, yielding the first theoretical framework to address minimal representations, IIS's and MIS's without a priori knowledge of feasibility.

4.2. The Set Covering Formulation

Consider the feasible set $Z(I)$. Since, for each i, $A_i = B_i^c$, the complement of $Z(I)$, the set of **infeasible points**, is

$$Z(I)^c = \bigcup_{i \in I} B_i.$$

Thus, $\{B_i : i \in I\}$ is a cover of $Z(I)^c$ and a reduction J of I defining $Z(I)$ amounts to a reduction of the cover, in that $\{B_i : i \in J\}$ also covers $Z(I)^c$:

$$\bigcup_{i \in J} B_i \supset Z(I)^c \tag{4.2}$$

This inclusion characterizing reduction tells us that each infeasible point is in some $B_i,\ i \in J$ and thus violates some constraint in the reduction J. Applied to irreducible reductions, this is the content of "The Main Theorem" of [Boneh (1984)]. In informal speech, when an irreducible reduction J is found, the constraints indexed by J are called **necessary** or **non-redundant**, and the others **redundant**. An irreducible reduction is not unique, but:

Theorem 1. *[Boneh (1984)] For each infeasible point x, some constraint violated by x must be necessary in each irreducible set of constraints.*

We emphasize that, as noted about, this is actually true for every *reduction*, even if it is not irreducible, since in practice it is usually not possible to obtain a truly irreducible one.

Corollary 2. *If x violates only one constraint, that constraint is necessary in each reduction.*

Let's gather the indicator functions δ_i defined above, into one "binary word valued" indicator function $\delta = \delta_I = (\delta_i)_{i\in I}$, mapping X to $\{0,1\}^I$. Its sets of constancy, in other words, the equivalence classes

$$[x] = \delta^{-1}(\delta(x)) = \{y : \delta(y) = \delta(x)\},$$

partition X. The resulting partition $\mathcal{P} = \mathcal{P}_I$ is the coarsest partition finer than each of the $\{A_i, B_i\}$ and

$$[x] = \left(\bigcap_{i:\delta_i(x)=0} A_i\right) \bigcap \left(\bigcap_{i:\delta_i(x)=1} B_i\right).$$

Since each class in $\mathcal{P}$ is determined by an element $\delta(x) \in \{0,1\}^I$, an upper bound on the cardinality of $\mathcal{P}$ is $2^{|I|}$. At times we will refer to $\delta(x)$ as the word or observation associated with the point x. Since $Z(I) = \bigcap_{i\in I} A_i = \delta^{-1}(0)$, the zero set of δ, it is also one of the classes in $\mathcal{P}$.

We extend these notions to subfamilies $\{A_i, B_i\}_{i\in J}$ of the original family $\{A_i, B_i\}_{i\in I}$. Thus, δ_J is the function on X with values in $\{0,1\}^J$ whose i^{th} component is δ_i, for each $i \in J$; in other words, δ_J is the composition of δ with the projection of $\{0,1\}^I$ onto $\{0,1\}^J$. We see that the partition $\mathcal{P}_J$ induced by δ_J is coarser than that of δ_I.

Theorem 3. *Let $y = 1_J \in \{0,1\}^I$. Then, J is a reduction of I if and only if $\delta(x) \cdot y \geq 1$ for all x with $\delta(x) \neq 0$.*

Proof. We have done all the work in setting up the notation. In terms of the indicator functions, the inclusion (4.2) says J is a reduction of I if and only if $\delta_I(x) \neq 0$ implies $\delta_J(x) \neq 0$ and this latter holds if and only if $\delta(x)\cdot 1_J \geq 1$. □

Thus, one can find an irreducible reduction of I by solving the set covering problem

$$\begin{array}{ll} \text{minimize} & |y| = \sum_{i\in I} y_i \\ \text{subject to} & \delta(x)\cdot y \geq 1, \text{ for all } x \text{ with } \delta(x) \neq 0. \end{array}$$

It is convenient to let E be the set of all possible words $\delta(x)$ other than 0, and think of it as a matrix whose rows are indexed by the infeasible equivalence classes (elements of the partition $\mathcal{P}$). Then this becomes a standard set-covering (SC) problem

$$\begin{array}{ll} \text{minimize} & |y| = \mathbf{1}^T y \\ \text{subject to } Ey \geq \mathbf{1}, \text{ y binary,} \end{array} \tag{4.3}$$

where $\mathbf{1}$ is a vector of ones of appropriate dimension.

In the case of linear programs, the corresponding SC problem can be solved in linear time. (This can be achieved by applying carefully applying Corollary 2.) This is not the case for more general problems. Fortunately, since any *feasible* y in the SC problem (4.3) corresponds to a reduction, one needn't actually find an *optimal* solution to obtain a benefit, and heuristics, such as the greedy algorithm in [Chvatal (1979)], can produce excellent results.

The partition $\mathcal{P}$, represented by the complete matrix E, provides a common framework for the concepts of minimal representation, irreducible infeasible system, and minimal infeasibility set. Suppose that we have a optimal solution to the set covering problem with corresponding irreducible reduction J. If the family is feasible, J provides a minimal representation. If the family is infeasible, J provides an irreducible infeasible system and the word with smallest row sum indicates a minimal infeasibility set. Thus, the concepts need no longer be treated separately.

Concerning the matrix E, we notice that:

(1) If columns k and ℓ in E are identical, then constraints k and ℓ are duplicate.
(2) If columns k and ℓ in E are complementary, then constraints k and ℓ are opposite, i.e., $A_k = B_\ell$.
(3) If column k is a column of zeros, then constraint k is everywhere satisfied.
(4) If column k is a column of ones, then constraint k is everywhere violated.

This next observation was suggested by A. Boneh in private conversation with the first author, and appeared in the master's major paper [Krishnamurthy (2001)] supervised by the authors.

Lemma 4. *If E contains $\binom{|I|}{m}$ rows with row sum m, then any reduction J of I has at least $(|I| - m + 1)$ elements.*

Proof. Note that $\binom{|I|}{m}$ is the number of possible rows of row sum m. If J is a subset of I with fewer than $|I| - m + 1$ elements, its complement in I contains a set K of m elements, which provides a row $e = 1_K^\top$ of E with $e \cdot 1_J = 0$, so that $E1_J \geq \mathbf{1}$ is not satisfied. □

The same argument can be applied to subsets I_0 of I. Thus, if I_0 has k elements and there are $\binom{k}{m}$ rows with exactly m non-zero entries in I_0, then $k + m - 1$ of the constraints in I_0 are necessary. In practice, it may be difficult to use this version, because it would require searching through too many submatrices. If $m = 1$, it would be easy for we could simply take I_0 to be the set of all i corresponding to row-sum 1, but this is already covered by Corollary 2. If m is 2, then this version would say that all but 1 of the members of I_0 are necessary.

Reducing the Set Covering Matrix. By a **reduction of the set covering matrix** E, we mean a subset F of E such that, for "column" binary words y, $Fy \geq \mathbf{1}$ implies $Ey \geq \mathbf{1}$. Clearly, the set covering problem obtained from the original by replacing E by F has the same feasible solutions, hence the same optimal solutions. The **bitwise partial ordering** on $\{0,1\}^I$, $e \leq f$, is given by $e \leq f$ if $e_i \leq f_i$, for all $i \in I$.

Lemma 5. *For $F \subseteq E \subseteq \{0,1\}^I$, F is a reduction of E (as a set covering matrix) iff for every $e \in E$, there exists $f \in F$, with $f \leq e$.*

Proof. If the condition holds, then for each $e \in E$, we can choose f with $e \geq f$, and then for each y, $ey \geq fy \geq 1$.

Conversely, suppose F is a reduction of E, but the condition is not satisfied; say, $e \in E$, but there is no $f \in F$ with $e \geq f$. Then, for each $f \in F$, we may choose $j = j_f$ with $1 = f_j > e_j = 0$. Let $J = \{j_f : f \in F\}$ and $y = 1_J$. Then $Fy \geq \mathbf{1}$, but $ey = 0$, a contradiction. □

Thus, E is irreducible (that is, has no proper reduction) if and only if no two elements are comparable. This is not to say that, if E corresponds to the family of constraints $\{A_i : i \in I\}$, the latter is irreducible.

4.3. Random Sampling

One way of collecting the elements of E is by sampling points $x \in X$ and calculating the corresponding $\delta(x)$.

Boneh's example. In [Boneh (1984)] the author presented a seemingly straightforward example to demonstrate his SC method. In a 1985 private communication, McDonald and Caron pointed out that a failure to sample on classes (members of the partion $\mathcal{P}$) of measure zero caused rows of E to be overlooked and led to incorrect conclusions. The 1999 Master's thesis by Feng [Feng (1999)], co-supervised by the authors, presented the first theoretical results aimed at the identification of a class of problems for which all classes can be sampled with nonzero probability. In the present paper, we provide a refined theorem and proof.

The **support** of a Borel measure is the complement of largest open set of zero measure. (See for example [Rao (2004)].) For a probability distribution on the Borel sets of a metric subspace of $\mathbf{R}^n$, this amounts to the smallest closed set of probability 1 (equivalently the set of points, each of whose neighbourhoods have positive probability — see [Chung (1974)], page 31.) We say the distribution P is **supported on** X if the support of P is X. Thus, if the distribution P is supported on X and $a \in X$, then every neighbourhood of a will intersect X in a set of positive probability. For example, X could be an interval (box) of $\mathbf{R}^n$ with non-empty interior and the distribution could be uniform on X or (the restriction to X of) a multivariate normal distribution. More generally, if X is a metric subspace of $\mathbf{R}^n$ and P has a continuous density f, with $(f > 0)$ dense in X, then P is supported on X.

Theorem 6. *Suppose that each A_i is given by $(g_i \leq 0) = \{x \in X : g_i(x) \leq 0\}$ where the g_i are continuous functions. For each $J \subseteq I$, put $g_J(x) = \max_{j \in J} g_j(x)$. If 0 is not a local minimum of any g_J, then each non-zero value of δ will be sampled with non-zero probability under any distribution supported on X.*

Proof. For a given $x \in X$, let J be the set of indices j with $g_j(x) = 0$. Then the equivalence class $[x]$ is $Z(J) \cap U(J^c)$, where $Z(J) = \bigcap_{j \in J} A_j = (g_J \leq 0)$ and $U(J^c) = \bigcap_{j \in J^c} B_j$, is an open set. If 0 is not a local minimizer, then the open set $(g_J < 0)$ contains a point a of $[x]$. Thus, the open set $(g_J < 0) \cap U(J^c)$ is a neighbourhood of a, hence intersects X in a set of positive probability. □

In Boneh's example, mentioned above, the hypotheses of Theorem 6 are not satisfied, since there is a local minimum of 0 for some g_J. The next result gives conditions under which the hypotheses of the theorem are satisfied.

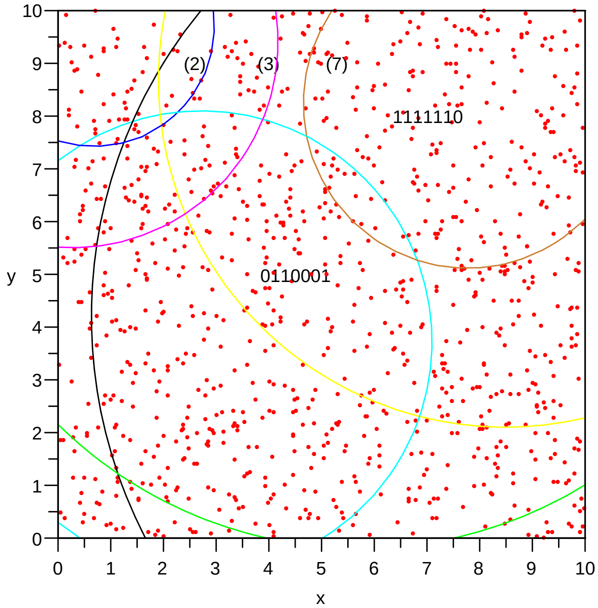
A constraint ($g_i \leq 0$) is said to be an **implicit equality** if there is a subset J of I with $Z(J) \neq \emptyset$ such that $g_i = 0$ on $Z(J)$. (The definition here is modified from that in [Obuchowska and Caron (1995)] to take into account the possibility of an infeasible family of constraints. The original concept was introduced by [Telgen (1983)].)

Corollary 7. *If the constraint functions are convex and there are no implicit equalities, all non-zero values of δ are chosen with positive probability under any distribution supported on X.*

Proof. In case all the functions g_j are convex, so are the g_J. The existence of a local minimum 0 would give a global minimum 0 and hence, $g_J = 0$ on $Z(J)$, which means the constraints g_j, $j \in J$ induce implicit equalities: on $Z(J)$, all g_i are 0. Thus, if there are no implicit equalities, 0 is not a local minimum for any g_J, and the theorem applies. □

We illustrate these ideas with families of (non-linear) convex quadratic constraints in 2 variables. Here X is an interval $[0, 10] \times [0, 10]$, the A_i are the intersections with X of elliptical regions with non-empty interior. Points are selected uniformly in X and the corresponding observations $\delta(x)$ are calculated. The distinct values of $\delta(x)$ are put into a set E and treated as the set covering matrix, although some rows may be missing. (Since the constraint functions are strictly convex, there can be no implicit equalities; hence, according to Corollary 7, each region will be sampled with positive probability.) Figure 1 shows an infeasible family of constraints, a plot of 1000 random points, the corresponding matrix E, and beside it an irreducible reduction, from which we see that irreducible reductions of the original problem are given by $\{2, 7\}$ and $\{3, 7\}$. Since the family is infeasible, these are Irreducible Infeasiblity Sets. The figure itself indicates that these results are probably correct. Chvatal's algorithm applied to this E yielded the reduction $\{2, 7\}$. Figure 2 shows a feasible family, its corresponding E and its (unique) corresponding irreducible reduction, which consists of only words with a single bit 1. This determines the minimal representation $\{1, 2, 3, 5, 6\}$. Note, by the way, that constraint 7 here turned out to be the entire interval X, so was always satisfied. This is reflected in the column of 0's in the matrix E.

Hit-and-Run variations. In [Boneh and Golan (1979)] a "hit-and-run" algorithm was introduced for the identification of redundancy and feasible region boundedness. This led to the development of a family of variations

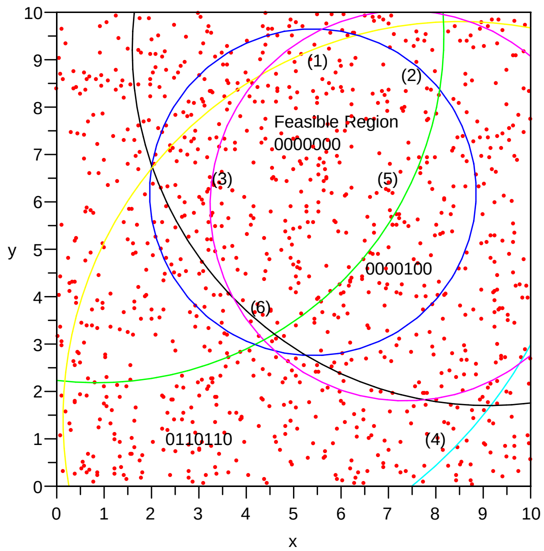


E	reduction of E
0101001	
0111000	
1111101	
1110101	
1110001	
1001011	
1111001	
1001001	
0110000	
0100001	
1110111	
0110001	
0111001	
1100001	
1100011	
1110011	
0001001	0110000
1000001	0100001
1000011	0001001
0001011	1000001

Figure 1: Infeasible family: IIS given by $\{2, 7\}$ or $\{3, 7\}$.

of the method ([Boneh (1983); Smith (1984); Telgen (1979); Berbee *et al.* (1987); Bélisle *et al.* (1998, 1993)]). Consider an absolutely continuous distribution π on an open set G of $\mathbf{R}^n$, with an almost-everywhere continuous Lebesgue density f, positive on G. Let ν be a distribution on the unit sphere of $\mathbf{R}^n$. Define a discrete time Markov chain by taking as transition kernel $P(x, B)$ the result of first choosing a direction s according to the distribution ν and then choosing a point according to the distribution π conditioned† on the line $L(x, s)$ through x in the direction s. In [Bélisle *et al.* (1993)] it is shown that if the support of ν spans $\mathbf{R}^n$ and if connected components ν-communicate (in particular if G is connected), then the chain converges in total variation to π. In particular, if the support of ν spans $\mathbf{R}^n$ and $P(x, B)$ comes from choosing a direction s according to ν, then a point from the uniform distibution on the intersection of $L(x, s)$ with a

†Since $L = L(x, s)$ has measure 0, this requires interpretation. One takes $f\mathbf{1}_L$ (normalized) as density with respect to 1 dimensional measure on L.



0110110	
0100100	
0010000	
0000100	
1110010	
0100000	
0110010	
1010000	
1100000	
1000000	
0100110	
1110000	
1110110	
0010010	
0000110	
0111100	1000000
0111110	0100000
1110100	0010000
0110100	0000100
0000010	0000010
E	reduction of E

Figure 2: Feasible family: constraints $1, 2, 3, 5, 6$ are necessary.

bounded open connected set G, specifically,

$$P(x, B) = \int \frac{\boldsymbol{\lambda}^1(B \cap L(x, G))}{\boldsymbol{\lambda}^1(G \cap L(x, s))} \, \nu(ds),$$

then the Markov chain converges in total variation to uniform distribution on G. Here $\boldsymbol{\lambda}^1$ denotes 1 dimensional Lebesgue (actually Hausdorff) measure in $\mathbf{R}^n$.

In application, if the space X is a subset of $\mathbf{R}^n$ whose interior is connected and whose boundary has measure 0, the analogous results would hold for X. Note that one only needs the support of ν to span $\mathbf{R}^n$, so one can simply use the uniform distribution on the n positive coordinate directions. This is the CD version of the Hit-and-run methods.

Once one decides to generate points along straight lines, $L(x, s)$, one can introduce modifications that have further advantages. For example, instead of just looking at the observation $\delta(x)$, one can collect many - in special cases all possible - observations along that line and easily retain an equivalent irreducible set of them. To illustrate, suppose X is convex and the A_j are of the form $(g_j \leq 0)$ with the g_j strictly convex, so that the $A_j \cap L(x, s)$ are completely determined by the solutions of the form $x + \sigma s$ to $g_j = 0$. Say these solutions are $a_i = x + \sigma_i s$, with $\sigma_i \leq \sigma_{i+1}$, for $1 \leq i < N$, $g_{j_i}(a_i) = 0$. Each index j_i will appear at most twice because of the convexity and as the parameter σ crosses σ_i, the j_i^{th} bit of $\delta(x+\sigma)$ will change from 1 to 0 or from 0 to 1. Thus, we can determine all the possible observations along that line.

This sets us up to use the following result, which enables one to select an equivalent irreducible set from the set of all observations collected along the line.

Theorem 8. *Let E be a set $\{e_1, \ldots, e_N\}$ of binary words in $\{0,1\}^J$. Suppose*

(1) for each $i < N$, either $e_i < e_{i+1}$ or $e_i > e_{i+1}$ and
(2) for each j there do not exist $i < k < \ell$ with $e_{ij} = 0$, $e_{kj} = 1$, and $e_{\ell j} = 0$.

Let $E' = \{e_i : i \in I'\}$ be the set of local minima of E in the partial ordering $\leq$. Thus, if $1 < i < N$, $i \in I'$ if and only if $e_{i-1} > e_i < e_{i+1}$, with the obvious modification for the cases $i = 1$ and $i = N$. Then, E' is a reduction of E for the set covering problem, and E' consists of incomparable words.

Condition (1) here is satisfied if there exists a unique j with $e_{i+1,j} \neq e_{ij}$,

Proof. To prove no two elements of E' are comparable, let $i, k \in I'$, with $i < k$. By the local minimality, $e_i < e_{i+1}$ and $e_{k-1} > e_k$. Choose j so that $e_{ij} < e_{i+1,j}$ and j' so that $e_{k-1,j'} > e_{k,j'}$. By condition (2) $e_{ij} < e_{kj}$ and $e_{ij'} > e_{kj'}$. Thus, neither $e_i \geq e_k$, nor $e_i \leq e_k$.

To show E' is a reduction of E, suppose $e_\ell \in E \setminus E'$. Then, either $e_\ell > e_{\ell+1}$ or $e_\ell > e_{\ell-1}$. In the first case, let i be the largest index $\geq \ell + 1$ with $e_{i-1} > e_i$. Then, $e_\ell > e_i$ and $e_i \in E'$. The case $e_\ell > e_{\ell-1}$ is similar. □

In conclusion, we would like to remind the reader, that although our illustrations here emphasized convex constraints, the framework is completely general: the sets A_i, B_i need not have any special geometric or topological properties.

Acknowledgement

This work has been supported by NSERC, the Natural Sciences and Engineering Research Council of Canada.

References

Andersen, E. and Andersen, K. (1995). Presolving in linear programming, *Mathematical Programming* **71**, pp. 221–245.

Bélisle, C., Boneh, A. and Caron, R. (1998). Convergence properties of hit-and-run samples, *Stochastic Models* **14**, pp. 767–800.

Bélisle, C., Romeijn, H. and Smith, R. (1993). Hit-and-run algorithms for generating multivariate distributions, *Mathematics of Operations Research* **18**, pp. 255–266.

Berbee, H., Boender, C., Rinnooy Kan, A. H., Romeijn, H., Scheffer, C., Smith, R. L. and Telgen, J. (1987). Hit-and-run algorithms for the identification of nonredundant linear equalities, *Mathematical Programming* **37**, pp. 184–207.

Bixby, R. (1994). Progress in linear programming, *ORSA Journal on Computing* **6**, pp. 15–22.

Boneh, A. (1983). PREDUCE - a probabilistic algorithm for identifying redundancy by a random feasible point generator (RFPG), in [Karwan *et al.* (1983)], pp. 108–134.

Boneh, A. (1984). Identification of redundancy by a set-covering equivalence, in J. Brans (ed.), *Operational Research '84* (Elsevier Science Publishers B.V.(North Holland), Amsterdam), pp. 407–422.

Boneh, A. and Golan, A. (1979). Constraints redundancy and feasible region boundedness by random feasible point generator (RFPG), Presented at the EURO III Conference, Amsterdam, April 9-11(22pp).

Brearly, A., Mitra, G. and Williams, H. (1975). Analysis of mathematical programming problems prior to applying the SIMPLEX method, *Mathematical Programming* **8**, pp. 54–83.

Caron, R. (2001). Redundancy in nonlinear programs, in C. Floudas and P. Pardalos (eds.), *Encyclopedia of Optimization* (Kluwer Academic Publishers), pp. 12–15.

Caron, R. J., Hlynka, M. and McDonald, J. F. (1990). On the best case performance of hit and run methods for detecting necessary constraints, *Mathematical Programming* **54**, pp. 233–249.

Caron, R. J., McDonald, J. F. and Ponic, C. M. (1989). A degenerate extreme point strategy for the classification of linear inequalities as redundant or necessary, *Journal of Optimization Theory and Applications* **62**, pp. 225–237.

Chakravarti, N. (1994). Some results concerning post-infeasibility analysis, *European Journal of Operational Research* **73**, pp. 139–143.

Chinneck, J. and Dravnieks, E. (1991). Locating minimal infeasible constraint sets in linear programs, *ORSA Journal on Computing* **3**, pp. 157–167.

Chung, K. L. (1974). *A course in probability theory*, 2nd edn. (Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London), probability and Mathematical Statistics, Vol. 21.

Chvatal, V. (1979). A greedy heuristic for the set-covering problem, *Mathematics of Operations Research* **4**, pp. 233–235.

Feng, J. (1999). *Redundancy in Nonlinear Systems: a Set Covering Approach*, Master's thesis, University of Windsor.

Greenberg, H. (1996). Consistency, redundancy and implied equalities in linear systsems, *Annals of Mathematics and Artificial Intelligence* **17**, pp. 37–83.

Jibrin, S. (2002). Detecting redundancy in optimization problems over intersection of ellipsoids, *J. Interdiscip. Math.* **5**, 2, pp. 183–194.

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

Krishnamurthy, A. (2001). *Classification of constraints by set covering*, Master's thesis, University of Windsor.

Lustig, I., Marsten, R. and Shanno, D. (1994). Interior point methods for linear programming: Computational state of the art, *ORSA Journal on Computing* **6**, pp. 1–14.

Obuchowska, W. (2000). Minimal representation of convex region defined by analytic constraints, *Journal of Mathematical Analysis and its Applications* **246**, pp. 100–121.

Obuchowska, W. T. and Caron, R. J. (1995). Minimal representation of quadratically constrained convex feasible regions, *Mathematical Programming* **68**, pp. 169–186.

Rao, M. M. (2004). *Measure theory and integration*, *Monographs and Textbooks in Pure and Applied Mathematics*, Vol. 265, 2nd edn. (Marcel Dekker Inc., New York), ISBN 0-8247-5401-8.

Smith, R. L. (1984). Efficient monte carlo procedures for generating points uniformly distributed over bounded regions, *Operations Research* **32**, pp. 1296–1308.

Telgen, J. (1979). The CD algorithm, Private communication.

Telgen, J. (1983). Identifying redundant constraints and implicit equalities in systems of linear constraints, *Management Science* **29**, pp. 1209–1222.

Tomlin, J. A. and Welch, J. F. (1986). Finding duplicate rows in a linear programming model, *Operations Research Letters* **5**, pp. 7–11.