A Robust Formulation of the Uncertain Set Covering Problem

Dirk Degel Pascal Lutter

Chair of Management, especially Operations Research—Ruhr-University Bochum
Universitaetsstrasse 150, 44801 Bochum, Germany
{dirk.degel,pascal.lutter}@rub.de

Abstract—This work introduces a robust formulation of the uncertain set covering problem combining the concepts of robust and probabilistic optimization and defines 'Γ-robust α-covers'. It is shown that the proposed robust uncertain set covering problem can be stated as a compact mixed-integer linear programming model which can be solved with modern computer software. This model is a natural extension of the classical set covering problem in order to cope with uncertainty in covering constraints.

Keywords—Set covering problem, Robust optimization, Robust uncertain set covering problem.

1 Introduction

The Set Covering Problem (SCP) minimizes the column costs that are necessary to ensure a full coverage of all rows. Let $I = \{1, \ldots, m\} \subset \mathbb{N}$ denote the index set of rows (indexed with i), $J = \{1, \ldots, n\} \subset \mathbb{N}$ denotes the index set of columns (indexed with j) and let $\mathcal{N}_i = \{j \in J \mid i \text{ can be covered by } j\}$ denote the neighborhood of a given column i. Costs associated with a column j are denoted by c_j for all $j \in J$. A $\{0, 1\}$-linear formulation of the classical set covering problem is given by (see [9, 5])

$$\min \quad \sum_{j \in J} c_j y_j \tag{1}$$

$$\text{s. t.} \quad \sum_{j \in J} a_{ij} y_j \geq 1 \qquad \forall i \in I \tag{2}$$

$$y_j \in \{0, 1\} \qquad \forall j \in J \tag{3}$$

where

$$a_{ij} = \begin{cases} 1, & \text{if } j \in \mathcal{N}_i \\ 0, & \text{else.} \end{cases}$$

The set covering problem is a well known *NP*-hard combinatorial optimization problem with many applications, particularly in emergency medical service facility location (see i. e. Degel et al. [4]). In order to cope with different aspects of uncertainty regarding the input parameters various models can be found in literature. Beraldi and Ruszczyński [1] introduce the probabilistic set covering problem, a chance-constraint formulation, where the right-hand side of constraint (2) is replaced by a binary random variable. Pereira and Averbakh [8] present a robust version of the set covering problem with interval uncertainty in cost-coefficients c_j. Hwang, Chiang and Liu [7, 3] as well as Fischetti and Monaci [6] focus on uncertainty in the coefficients a_{ij}. Hwang et al. [7] develop a fuzzy set covering problem and provide a binary linear reformulation of the problem [3]. Fischetti and Monaci define an uncertain set covering problem where constraint (2) is also replaced by a chance-constraint which deals with columnwise coefficient uncertainty. For each column $j \in J$ the entries in the coefficient vector $a_{\cdot j}$ flip from 1 to 0 with a known probability $p_j \in [0, 1]$. That means a complete column $j \in J$ which is assumed to be able to cover row $i \in I$ ($a_{ij} = 1$) may disappear ($a_{ij} = 0$) with a probability of p_j.

We extend this concept by including individual and independent coefficient disappearing probabilities $p_{ij} \in [0, 1]$ for each row $i \in I$ and column $j \in J$. Generalizing the approach of [6] and assuming the coefficients a_{ij} to be independent binary random variables leads to

$$a_{ij} = \begin{cases} 1, & \text{with probability } 1 - p_{ij} \\ 0, & \text{with probability } p_{ij}\,. \end{cases}$$

Let P be a probability measure, then we define an α-covering constraint, $\alpha \in (0, 1]$, of row i by

$$P(\sum_{j \in J} a_{ij} y_j \geq 1) \geq \alpha\,.$$

This constraint ensures that each row $i \in I$ will be covered with a probability of at least α. Then we define the *generalized* uncertain set covering problem (GUSCP) as

$$\begin{aligned} \min \quad & \sum_{j \in J} c_j y_j \\ \text{s.t.} \quad & P(\sum_{j \in J} a_{ij} y_j \geq 1) \geq \alpha & \forall i \in I \\ & y_j \in \{0,1\} & \forall j \in J \end{aligned} \tag{4}$$

A solution $y^\star \in \{0,1\}^n$ is feasible for GUSCP if and only if $P(\sum_{j \in J} a_{ij} y_j^\star \geq 1) \geq \alpha$ is satisfied for all $i \in I$ which is equivalent to $P(\sum_{j \in J} a_{ij} y_j^\star < 1) = \prod_{j \in \mathcal{C}(y^\star)} p_{ij} \leq 1 - \alpha$ with $\mathcal{C}(y^\star) = \{j \in J \mid y_j^\star = 1\}$ for all $i \in I$.

2 Robust covers and the robust uncertain set covering problem

In most real world applications the actual probabilities p_{ij} are not known precisely. It is more likely to estimate these probabilities on the basis of a given set of data. This leads to natural deviations of the estimated probabilities, in the following called nominal value, from their true but unknown counterparts. In cases with a large data set, these deviations are typically smaller compared to situations with relatively small data sets. These deviations can be quantified by intervals which allow the use of interval uncertainties following the robustness concept of Bertsimas and Sim [2]. Hence, we assume p_{ij} to be uncertain within the interval $[\bar{p}_{ij} - \hat{p}_{ij}, \bar{p}_{ij} + \hat{p}_{ij}] \subset [0,1]$ (note that our results are still valid if p_{ij} has only realizations in $[\bar{p}_{ij}, \bar{p}_{ij} + \hat{p}_{ij}] \subset [0,1]$) where $\bar{p}_{ij} \geq 0$ indicates the nominal value and $\hat{p}_{ij} \geq 0$ denotes the maximum absolute deviation. The goal is to obtain an α-coverage of row $i \in I$ which remains feasible even if up to $\Gamma_i \in \mathbb{N}_0$ values of p_{ij} are realized in their worst case scenario $\bar{p}_{ij} + \hat{p}_{ij}$ and the $n - \Gamma_i$ other realizations of p_{ij} take their nominal values $\bar{p}_{ij}$. A Γ-robust α-cover is defined as follows.

Definition 1 Let $i \in I$ be fixed. Let $\Gamma_i \in \mathbb{N}_0$, $\alpha \in (0,1]$ and let p_{ij} have realizations in $[\bar{p}_{ij} - \hat{p}_{ij}, \bar{p}_{ij} + \hat{p}_{ij}] \subset [0,1]$ for all $j \in J$. A *Γ_i-robust α-cover* of the i-th row is defined as a solution $y^\star \in \{0,1\}^n$ with $P_{\Gamma_i}(\sum_{j \in J} a_{ij} y_j^\star \geq 1) \geq \alpha$ where $P_{\Gamma_i}(\sum_{j \in J} a_{ij} y_j^\star \geq 1)$ denotes the minimum coverage probability on the condition that at most Γ_i realizations of p_{ij} with $j \in \mathcal{C}(y^\star)$ are equal to their worst case scenario $\bar{p}_{ij} + \hat{p}_{ij}$ and the $n - \Gamma_i$ other realizations of p_{ij} with $j \in \mathcal{C}(y^\star)$ are equal to their nominal values $\bar{p}_{ij}$.

If all realizations of p_{ij} only take values within the interval $[\bar{p}_{ij}, \bar{p}_{ij} + \hat{p}_{ij}]$ the definition can be stated in an analogous way. The obtained solutions coincide in both cases. This leads to the following *robust uncertain set covering problem* (RUSCP)

$$\begin{aligned} \min \quad & \sum_{j \in J} c_j y_j \\ \text{s.t.} \quad & P_{\Gamma_i}(\sum_{j \in J} a_{ij} y_j \geq 1) \geq \alpha & \forall i \in I \\ & y_j \in \{0,1\} & \forall j \in J. \end{aligned} \tag{5}$$

In the following, we will show how to derive a mixed-integer linear programming formulation of RUSCP. In the first step, a reformulation of the left-hand side of the chance-constraint (5) is provided in the following proposition.

Proposition 1 *Let $y^\star \in \{0,1\}^n$ and set*

$$w'_{ij} := \begin{cases} \ln(\bar{p}_{ij} + \hat{p}_{ij}) & \text{if } \bar{p}_{ij} + \hat{p}_{ij} > 0 \\ \ln(1-\alpha) & \text{if } \bar{p}_{ij} + \hat{p}_{ij} = 0 \end{cases}$$

and

$$w_{ij} := \begin{cases} \ln(\bar{p}_{ij}) & \text{if } \bar{p}_{ij} > 0 \\ \ln(1-\alpha) & \text{if } \bar{p}_{ij} = 0 \end{cases}$$

for all $i \in I$ and $j \in J$. Let $\mathcal{C}(y^\star) = \{j \in J \mid y^\star_j = 1\}$, then

$$P_{\Gamma_i}(\sum_{j \in J} a_{ij} y^\star_j \geq 1) \geq \alpha \iff \max_{\{U \subset \mathcal{C}(y^\star) \mid |U| \leq \Gamma_i\}} \left\{ \sum_{j \in U} w'_{ij} y^\star_j + \sum_{j \in J \setminus U} w_{ij} y^\star_j \right\} \leq \ln(1-\alpha) \tag{6}$$

holds for all $i \in I$. □

Proof Let $i \in I$ be fixed and let $y^\star \in \{0,1\}^n$. The left-hand side of the i-th Γ_i-robust α-covering constraint can be stated as

$$P_{\Gamma_i}(\sum_{j \in J} a_{ij} y^\star_j \geq 1) = 1 - P_{\Gamma_i}(\sum_{j \in J} a_{ij} y^\star_j < 1) = 1 - \max_{\{U \subset \mathcal{C}(y^\star) \mid |U| \leq \Gamma_i\}} \left\{ \prod_{j \in U} (\bar{p}_{ij} + \hat{p}_{ij}) \prod_{j \in \mathcal{C}(y^\star) \setminus U} \bar{p}_{ij} \right\}.$$

We divide the proof into two cases. At first assume $\bar{p}_{ij} > 0$ for all $j \in J$. Then constraint (5) can be reformulated as follows

$$
\begin{aligned}
& 1-\max_{\{U\subset\mathcal{C}(y^\star)||U|\le\Gamma_i\}}\left\{\prod_{j\in U}(\bar{p}_{ij}+\hat{p}_{ij})\prod_{j\in\mathcal{C}(y^\star)\setminus U}\bar{p}_{ij}\right\}\ge\alpha\\
\iff\quad & \max_{\{U\subset\mathcal{C}(y^\star)||U|\le\Gamma_i\}}\left\{\sum_{j\in U}\ln(\bar{p}_{ij}+\hat{p}_{ij})+\sum_{j\in\{k\in J\setminus U|y^\star_k=1\}}\ln(\bar{p}_{ij})\right\}\le\ln(1-\alpha)\\
\iff\quad & \max_{\{U\subset\mathcal{C}(y^\star)||U|\le\Gamma_i\}}\left\{\sum_{j\in U}\ln(\bar{p}_{ij}+\hat{p}_{ij})y^\star_j+\sum_{j\in J\setminus U}\ln(\bar{p}_{ij})y^\star_j\right\}\le\ln(1-\alpha)\\
\iff\quad & \max_{\{U\subset\mathcal{C}(y^\star)||U|\le\Gamma_i\}}\left\{\sum_{j\in U}w'_{ij}y^\star_j+\sum_{j\in J\setminus U}w_{ij}y^\star_j\right\}\le\ln(1-\alpha).
\end{aligned}
$$

Analogously, one can proof the case of $\bar{p}_{ij}=0$ or $\bar{p}_{ij}+\hat{p}_{ij}=0$, but the index sets have to be split into two sets, where one set contains all indices of positive $\bar{p}_{ij}$ or $\bar{p}_{ij}+\hat{p}_{ij}$ and the other set contains all indices for zero elements. ■

By analogy with the idea of Bertsimas and Sim [2] the maximization subproblem in (6) is defined as

$$\beta_i(y,\Gamma_i):=\max_{\{U\subset\mathcal{C}(y)||U|\le\Gamma_i\}}\left\{\sum_{j\in U}w'_{ij}y_j+\sum_{j\in J\setminus U}w_{ij}y_j\right\}.$$

The next proposition shows how to derive a linear programming formulation of the subproblem of the i-th row $\beta_i(y,\Gamma_i)$ for a given solution $y^\star\in\{0,1\}^n$.

Proposition 2 *Let $i\in I$ be fixed. For a given solution $y^\star\in\{0,1\}^n$ the subproblem*

$$\beta_i(y^\star,\Gamma_i)=\max_{\{U\subset\mathcal{C}(y^\star)||U|\le\Gamma_i\}}\left\{\sum_{j\in U}w'_{ij}y^\star_j+\sum_{j\in J\setminus U}w_{ij}y^\star_j\right\}$$

can be stated as a linear program in the form

$$
\begin{aligned}
\beta_i(y^\star,\Gamma_i)=\sum_{j\in J}w_{ij}y^\star_j+\min\quad & \sum_{j\in J}\zeta_{ij}+\Gamma_i\eta_i\\
s.\,t.\quad & \zeta_{ij}+\eta_i\ge(w'_{ij}-w_{ij})y^\star_j && \forall j\in J\\
& \zeta_{ij}\ge 0 && \forall j\in J\\
& \eta_i\ge 0\,.
\end{aligned}
$$

□

Proof Let $i\in I$ be fixed. It is easy to verify that for a given solution $y^\star\in\{0,1\}^n$ the subproblem

$$\beta_i(y^\star, \Gamma_i) = \max_{\{U \subset \mathcal{C}(y^\star) | |U| \le \Gamma_i\}} \left\{ \sum_{j \in U} w'_{ij} y^\star_j + \sum_{j \in J \setminus U} w_{ij} y^\star_j \right\}$$

can be stated as an integer linear program

$$\begin{aligned} \beta_i(y^\star, \Gamma_i) = \max \quad & \sum_{j \in J} w'_{ij} y^\star_j \xi_j + \sum_{j \in J} w_{ij} y^\star_j (1 - \xi_j) \\ \text{s.t.} \quad & \sum_{j \in J} \xi_j \le \Gamma_i \\ & \xi_j \in \{0, 1\} && \forall j \in J. \end{aligned}$$

The objective function of this problem can be reformulated as

$$\sum_{j \in J} w_{ij} y^\star_j + \max\{\sum_{j \in J} w'_{ij} y^\star_j \xi_j - \sum_{j \in J} w_{ij} y^\star_j \xi_j\} = \sum_{j \in J} w_{ij} y^\star_j + \max\{\sum_{j \in J} (w'_{ij} - w_{ij}) y^\star_j \xi_j\}$$

due to the fact that $\sum_{j \in J} w_{ij} y^\star_j$ is constant. For all $j \in J$ holds $w'_{ij} \ge w_{ij}$ which implies $w'_{ij} - w_{ij} \ge 0$. The polytope $\{\xi \in \mathbb{R}^n \mid \sum_{j \in J} \xi_j \le \Gamma_i, \forall j \in J : \xi_j \le 1, \xi_j \ge 0\}$ is integral which allows us to write the subproblem as:

$$\beta_i(y^\star, \Gamma_i) = \sum_{j \in J} w_{ij} y^\star_j + \max \left\{ \sum_{j \in J} (w'_{ij} - w_{ij}) y^\star_j \xi_j \mid \sum_{j \in J} \xi_j \le \Gamma_i, \forall j \in J : \xi_j \in [0, 1] \right\}. \quad (7)$$

If the primal problem in (7) is feasible and bounded then the dual problem is also feasible and bounded. Applying the strong duality theorem it follows that both objective values coincide. Dualizing the maximization problem in (7) concludes the proof. ■

Replacing the nonlinear constraint (5) we receive a mixed-integer linear formulation for the robust uncertain set covering problem (RUSCP):

$$\begin{aligned} \min \quad & \sum_{j \in J} c_j y_j \\ \text{s.t.} \quad & \sum_{j \in J} w_{ij} y_j + \sum_{j \in J} \zeta_{ij} + \Gamma_i \eta_i \le \ln(1 - \alpha) && \forall i \in I \\ & \zeta_{ij} + \eta_i \ge (w'_{ij} - w_{ij}) y_j && \forall i \in I, \forall j \in J \\ & \zeta_{ij} \ge 0 && \forall i \in I, \forall j \in J \\ & \eta_i \ge 0 && \forall i \in I \\ & y_j \in \{0, 1\} && \forall j \in J. \end{aligned}$$

3 Example

A very important application of the SCP addresses the location of emergency medical service facilities. The goal is to determine locations $j \in J$ of emergency medical service facilities assuring the coverage of all demand nodes $i \in I$ such that induced costs are minimized. A demand node $i \in I$ is covered if it can be reached by an emergency medical vehicle within a pre-specified time limit. In real world applications of this problem, there may exist some demand nodes $i \in I$ which cannot be covered with certainty. Even their covering probabilities cannot be forecasted with certainty. The following figure illustrates the geographical interpretation and the robust covering constraint for a given demand node i:

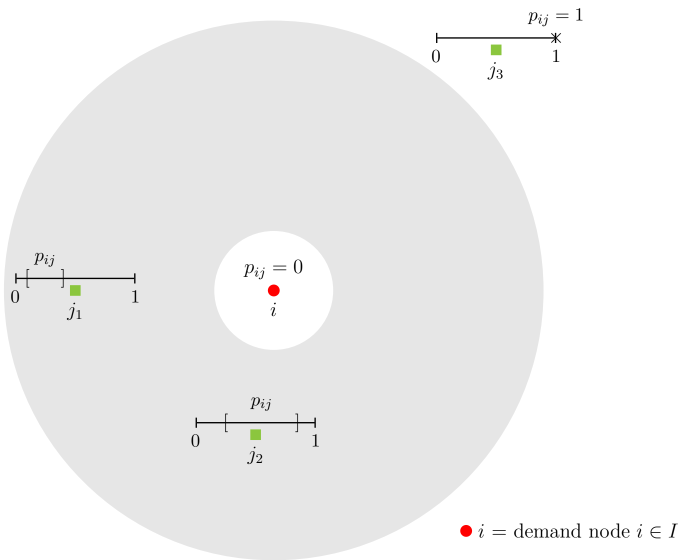

Figure 1: Geographical interpretation of the RUSCP.

The inner radius determines the certain-region. Each vehicle associated to a facility $j \in J$ within this area can reach demand node i with certainty. This corresponds to a disappearing probability of zero. The grey area around i illustrates the uncertain-region. In this illustration facilities j_1 and j_2 belong to the uncertain-region. The probability that a vehicle is not able to reach demand node i from this area cannot be quantified exactly. These probabilities are assumed to have realizations within pre-defined intervals, denoted by the brackets on the $[0, 1]$-line. If facilities, e. g. j_3, are located outside of the mentioned regions, it is not possible to reach demand node i within the time limit.

Assume that we have four demand nodes, $I = \{1, 2, 3, 4\}$, that should be covered with a probability of at least $\alpha = 0.98$ by at most three different facilities, $J = \{1, 2, 3\}$.

Let $y_j = 1$ if facility $j \in J$ is built and $y_j = 0$ otherwise. The nominal value $\bar{p}_{ij}$ of the probability that demand node i cannot be covered by facility j and the maximum deviation $\hat{p}_{ij}$ are given by

$$\bar{p} = \begin{pmatrix} 0.02 & 0.16 & 0.02 \\ 0.16 & 0.02 & 0.02 \\ 0.04 & 0.14 & 0.02 \\ 0.13 & 0.03 & 0.02 \end{pmatrix} \qquad \hat{p} = \begin{pmatrix} 0.05 & 0.1 & 0.12 \\ 0.1 & 0.04 & 0.13 \\ 0.05 & 0.07 & 0.1 \\ 0.05 & 0.04 & 0.1 \end{pmatrix}$$

for all $i \in I, j \in J$. The disappearing probability matrix p has realizations within the interval $[\bar{p}, \bar{p} + \hat{p}]$. The goal is to minimize the number of facilities ($c_j = 1$ for all $j \in J$) such that all nodes are α-covered and protected against $\Gamma := \Gamma_i$, $1 \leq i \leq 4$ worst case realizations of p according definition 1. This leads to Γ-robust 0.98-covering constraints. Table 1 shows the conditional minimum covering probabilities for each demand node depending on the value of Γ. A solution is feasible if the corresponding row only contains entries greater or equal to 0.98. Feasible solutions are highlighted bold while optimal solutions are marked with a star. The first row of the table shows conditional minimal

	demand node	built facility 1	2	3	1 & 2	1 & 3	2 & 3	1 & 2 & 3
$\Gamma = 0$	$i = 1$	0.98	0.84	**0.98***	**0.9968**	**0.9996**	**0.9968**	**0.999936**
	$i = 2$	0.84	0.98	**0.98***	**0.9968**	**0.9968**	**0.9996**	**0.999936**
	$i = 3$	0.96	0.86	**0.98***	**0.9944**	**0.9992**	**0.9972**	**0.999888**
	$i = 4$	0.87	0.97	**0.98***	**0.9961**	**0.9974**	**0.9994**	**0.999922**
$\Gamma = 1$	$i = 1$	0.93	0.74	0.86	**0.9888***	0.9972	0.9776	**0.999552**
	$i = 2$	0.74	0.94	0.85	**0.9904***	0.976	0.997	**0.99952**
	$i = 3$	0.91	0.79	0.88	**0.9874***	0.9952	0.9832	**0.999328**
	$i = 4$	0.82	0.93	0.88	**0.9909***	0.9844	0.9964	**0.999532**
$\Gamma = 2$	$i = 1$	0.93	0.74	0.86	**0.9818***	0.9902	0.9636	**0.998432**
	$i = 2$	0.74	0.94	0.85	**0.9844***	0.961	0.991	**0.99856**
	$i = 3$	0.91	0.79	0.88	**0.9811***	0.9892	0.9748	**0.998488**
	$i = 4$	0.82	0.93	0.88	**0.9874***	0.9784	0.9916	**0.998908**
$\Gamma = 3$	$i = 1$	0.93	0.74	0.86	**0.9818***	0.9902	0.9636	**0.997452**
	$i = 2$	0.74	0.94	0.85	**0.9844***	0.961	0.991	**0.99766**
	$i = 3$	0.91	0.79	0.88	**0.9811***	0.9892	0.9748	**0.997732**
	$i = 4$	0.82	0.93	0.88	**0.9874***	0.9784	0.9916	**0.998488**

Table 1: Γ-robust covering probabilities for all demand nodes and all combinations of facilities ($\Gamma \in \{0, 1, 2, 3\}$)

covering probabilities with no protection against deviations, that means $\Gamma = 0$. Clearly, this solution coincides with the GUSCP with nominal values and the optimal solution is to build only facility 3. The second row shows minimal covering probabilities with $\Gamma = 1$, which means protecting against at most one worst case realization. Building only one facility is no longer feasible. The alternatives to build facilities 1 and 3 as well as building facilities 2 and 3 are not feasible due to demand node 2 respectively demand node 1. The optimal solution is to built facilities 1 and 2. Protecting against at most two or three worst case realizations ($\Gamma = 2$ resp. $\Gamma = 3$) leads to the same optimal solution as in the former case.

4 Conclusion

In this paper a new approach to cope with uncertainties in set covering problems, called RUSCP, was developed. We extended the current state of the art literature regarding the integration of interval uncertainties *in the probability distribution of* $\{0, 1\}$ *parameters*. It was shown that the robust counterpart of the uncertain set covering problem can be stated as a mixed-integer linear programm in a compact formulation.

References

[1] Patrizia Beraldi and Andrzej Ruszczyński. The probabilistic set-covering problem. *Operations Research*, 50(6):956–967, 2002.

[2] Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations research*, 52(1):35–53, 2004.

[3] C. I. Chiang, M. J. Hwang, and Y. H. Liu. An alternative formulation for certain fuzzy set-covering problems. *Mathematical and Computer Modelling: An International Journal*, 42(3–4):363–365, 2005.

[4] Dirk Degel, Sebastian Rachuba, Brigitte Werners, and Lara Wiesche. Multi-criteria covering-based location of volunteer fire departments (workingpaper). 2012.

[5] Reza Zanjirani Farahani, Nasrin Asgari, Nooshin Heidari, Mahtab Hosseininia, and Mark Goh. Covering problems in facility location: A review. *Computers & Industrial Engineering*, 62(1):368–407, 2012.

[6] Matteo Fischetti and Michele Monaci. Cutting plane versus compact formulations for uncertain (integer) linear programs. *Mathematical Programming Computation*, 4(3):239–273, 2012.

[7] M. J. Hwang, C.I. Chiang, and Y. H. Liu. Solving a fuzzy set-covering problem. *Mathematical and computer modelling*, 40(7):861–865, 2004.

[8] Jordi Pereira and Igor Averbakh. The robust set covering problem with interval data. *Annals of Operations Research*, pages 1–19, 2011.

[9] Constantine Toregas, Ralph Swain, Charles ReVelle, and Lawrence Bergman. The location of emergency service facilities. *Operations Research*, 19(6):1363–1373, 1971.