Senior Thesis in Mathematics

The Parks Location Problem

Author:
Alan Antonio Peral Ortiz

Advisor:
Dr. Shahriar Shahriari

Submitted to Pomona College in Partial Fulfillment
of the Degree of Bachelor of Arts

March 24, 2018

Abstract

We take a look at the multiobjective facility location problem of park placement in the city of Bogotá, Colombia. We begin with an introduction to operations research, delving into its development and history, before we move on a to simplified example model. We work up from this example to look at the final model. We then examine more information about multiobjective and facility location problems.

Contents

Chapter 1

Introduction

1.1 What is Operations Research?

We begin with a brief history of Operations Research (OR), which as a modern discipline has its roots in the Soviet political economy as well as the British war effort of World War II.

[SECTION ABOUT SOVIET UNION + OR]

Now, in the year 1934 as Nazi Germany denounced the Treaty of Versailles, Britain sensed the threat and raced to strengthen its defenses. By the following year, radar had been developed and was capable of effectively detecting enemy aircraft, although its utility was stringent on its ability for integration with the existing defense systems: ground observers, interceptor aircraft, and antiaircraft artillery positions. The first task of the newly formed operational research group was to use scientific, rigorous processes to develop a system for the incorporation of radar into existing infrastructure, which was far from the more well-defined mathematical processes that exist today. However, this integrated radar-based air defense system increased the probability of intercepting an enemy aircraft by a factor of ten [CITE HISTORY OF OR IN US MILITARY BOOK].

Soon, the usage of OR spread to investigate other problems, and in 1940 was even called upon to influence high-level strategic policy, when the French requested additional RAF fighter support. Churchill was inclined to acquiesce to the request, but an OR team showed that sending more RAF fighters would weaken Britain beyond recovery in the face of a German attempt to invade Britain [CITE HISTORY OF OR IN US MILITARY]. Churchill was convinced by their presentation, and so he did not send the additional aircraft, preserving the pilots and aircrafts for the Battle of Britain instead. This decision, along with the incorporation of radar, contributed significantly to Britain's victory in the Battle of Britain [CITE HISTORY OF OR IN US MILITARY]. Another significant accomplishment of OR analysts involved the work of deciding upon depth charge settings for bombs. Their work in this field led to an immediate improvement on aerial attacks of German submarines, with estimates of the increased efficiency ranging from 400 to 700 percent [CITE HISTORY OF OR IN US MILITARY]. Other work included reducing the number of artillery rounds required to down one German aircraft from twenty thousand in the summer of 1941 to merely four thousand the following year [CITE HISTORY OF OR IN US MILITARY]. From the British, this discipline spread to the Americans, where it was also used in the war effort. [TALK MORE ABOUT 'MURICA HERE!].

There was a specific need to allocate scarce resources to military operations in an effective manner, and so the British and US militaries had scientists perform *research on* (military) *operations*. In essence, the goal was to make the war machine more efficient, and they succeeded. They developed effective ways to use the new radar technology, as well as came up with better ways to manage convoys and conduct antisubmarine operations [CITE INTRO TO OR BOOK].

The success that OR saw in the war then encouraged interest in nonmilitary applications of the field. A cursory glance at a variety of introductory textbooks will reveal that there is a certain focus on private sector applications of the field [Should probably cite this claim]. These applications tend to be primarily concerned with profit maximization and other aspects of running a business. While these problems provide for some interesting mathematical formulations, the field's ability to

There is a subfield of the discipline called Community-Based Operations Research, which seeks to shift the focus of OR from profit maximization or cost-reduction to improving the quality of life within a community. One of the main advocates of the field and of this lens that focuses on people as opposed to money is Michael P. Johnson, who compiled a textbook containing a variety of case studies that can constitute "Community-Based Operations Research."

One of these case studies presented is the problem of park location in Bogotá, Colombia.

In the introduction I will explain the history of operations research and how it tends to be most utilized in the private sector. I will detail the development of (and necessity for) community-based operations research, and then explain that I will be examining a particular case study: the problem of developing new public parks in Bogotá, Colombia. I will perhaps provide a literature review here, a roadmap of what I will cover in the rest of the thesis, and anything else that may fit and come to mind later.

1.2 An Introduction to the Parks Location Problem

In rapidly developed urban areas and large cities, the presence of public parks, green spaces, and other recreational facilities has been associated with a marked improvement in quality of life, mental health, and general wellbeing [CITE SOURCE]. Knowing this, the city of Bogotá (Colombia), one of the largest cities in Latin America with a population of about eight million that is expected to reach ten million by 2025, has implemented a number of changes that include the recovery of public spaces and the improvement of public parks [CITE THE SOURCE].

In 2006, the mayor and city council of Bogotá threw their support behind a sports and recreation master plan for the city. This plan indicated that by 2019 the city must reach a minimum level of 2.71 m^2 of neighborhood park area per resident. It then became the *Instituto Distrital de Recreación y Deporte* (IDRD), or Recreation and Sports Institute of Bogotá's job to implement the master plan. As such, the IDRD was faced with a monumental challenge: they had to execute the construction of numerous new parks and revitalize dilapidated public spaces in a manner that balanced the differing geographic, social, and economic needs of the city.

Because of the many needs they had to consider and due to the nature of the problem, this problem was then modeled in such a way that it became a multiobjective facility location problem, which we will examine in further detail at a later point. For now we begin to consider the model.

Chapter 2

A Look at the Model

2.1 A Simplified Model

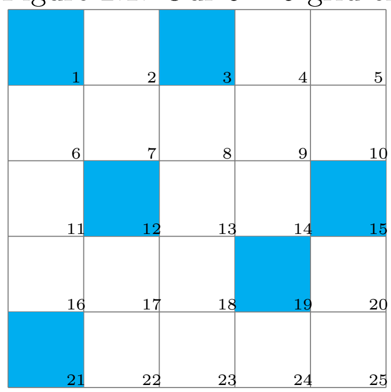
2.1.1 Setting up the Problem

In this chapter I present a simplified version of the model. We will take a look at some simpler examples and gradually increase the complexity of our example until we arrive at the one presented in the paper. We begin with a 5×5 grid, which will represent our simplified fictional city. Each block represents a plot of land, which might either be empty or occupied. If it is occupied, this means that there are people living on the block. Otherwise, the lot is empty and it is a candidate parcel, to potentially be turned into a park. We number each block from one to 25, and we have colored every candidate parcel cyan. We want to build one park, and so we can choose from amongst all the cyan parcels. This is shown in Figure 2.1. We note that every lot is the same size. We also define the service area of the candidate parcels to consist of all the populated blocks adjacent and diagonal to the candidate parcel. For example, candidate parcel one's service area will consist of the set $\{2, 6, 7\}$. In other words, if we built a park on candidate parcel one, then the visitors would be from **only** those three surrounding lots. The first question we are interested in asking then becomes:

Which candidate parcel will maximize the geographical coverage, as measured by the service area of the parcel?

In this example, it is possible to answer our question after merely observing the grid. Candidate parcel 12 would result in the highest number of lots

Figure 2.1: Our 5×5 grid city



served with 8. We now move on to defining a certain number of sets and variables so that we can present a mathematical formulation of the problem.

We first let $\mathcal{J}$ be the set of all blocks that would benefit from the construction of a new park. Looking back to figure 2.1, we see that $\mathcal{J} = \{2, 4, 6, 7, \ldots, 22, 23, 24, 25\}$. The candidate parcels have not been included in this set, and in this case we note that lot five is not included in $\mathcal{J}$ either because it is outside the service area of every candidate parcel. We also then define the set $\mathcal{I}$ to be the set of candidate parcels. In our case, $\mathcal{I} = \{1, 3, 12, 15, 19, 21\}$. We finally define the set $\mathcal{W}_j$ for $j \in \mathcal{J}$, which consists of all the candidate parcels that service block j. In our example, $\mathcal{W}_7 = \{1, 3, 12\}$.

We now define z_j as a binary decision variable, taking on the value 1 if block $j \in \mathcal{J}$ is covered by at least one park, and 0 otherwise. So if we decided to build a park on candidate parcel 21, then $z_{16} = z_{17} = z_{22} = 1$, and for every other possible value of j we would have $z_j = 0$. We also define the binary decision variable y_i, which will take on value 1 if candidate parcel i is selected to become a park, and 0 otherwise. If we turned candidate parcel

21 into a park, then $y_{21} = 1$. We can now formulate our initial model:

$$\begin{aligned}
\max\ f_1 &= \sum_{j \in \mathcal{J}} z_j \\
\text{subject to } z_j &\leq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} \\
|\mathcal{W}_j|\, z_j &\geq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} \\
p_{max} &\geq \sum_{i \in \mathcal{I}} y_i \\
z_j &\in \{0, 1\}, j \in \mathcal{J} \\
y_i &\in \{0, 1\}, i \in \mathcal{I}
\end{aligned}$$

Our objective function seeks to maximize the geographical coverage of the potential parks to be built. The first constraint guarantees that if block j is covered, then at least one parcel servicing it has been selected as a park. So if block 2 is covered in figure 2.1, then either candidate parcel 1 or candidate parcel 3 should have been selected to become a park. Conversely, the second constraint guarantees that if block j is not covered, then none of the candidate parcels servicing it should be selected as parks. While I have referred to these first two constraints as being one individual constraint each, the observation can be made that they must be satisfied for all values of $j \in \mathcal{J}$, meaning that each constraint must be repeated $|\mathcal{J}|$ times.

The next constraints indicates how many parks we would like to build. The variable p_{max}, or the maximum allowable amount of parks we want built, is a constant that is decided upon beforehand. In our case, we decided that we wanted to build one park, so $p_{max} = 1$. The last two constraints define our z_j and y_i as binary decision variables.

We can see an expansion of the model above to include all true constraints in Figure 2.2.

As we realized earlier, wanting to select only one candidate parcel will result in the selection of candidate parcel 12. What if we wanted to build two parks? Looking at Figure 2.3 we see that picking the two green parcels would result in a total coverage area of 13 blocks, highlighted in orange, which is

Figure 2.2: Fully expanded model

$$\max\ f_1 = z_2 + z_4 + z_6 + z_7 + z_8 + z_9 + z_{10} + z_{11} + z_{13} + z_{14} + z_{16} + z_{17} + z_{18} + z_{20} + z_{22} + z_{23} + z_{24} + z_{25}$$

$$\begin{aligned}
\text{subject to } z_2 &\le y_1 + y_3 & |W_2|z_2 &\ge y_1 + y_3 \\
z_4 &\le y_3 & |W_4|z_4 &\ge y_3 \\
z_6 &\le y_1 + y_{12} & |W_6|z_6 &\ge y_1 + y_{12} \\
z_7 &\le y_1 + y_3 + y_{12} & |W_7|z_7 &\ge y_1 + y_3 + y_{12} \\
z_8 &\le y_3 + y_{12} & |W_8|z_8 &\ge y_3 + y_{12} \\
z_9 &\le y_3 + y_{15} & |W_9|z_9 &\ge y_3 + y_{15} \\
z_{10} &\le y_{15} & |W_{10}|z_{10} &\ge y_{15} \\
z_{11} &\le y_{12} & |W_{11}|z_{11} &\ge y_{12} \\
z_{13} &\le y_{12} + y_{19} & |W_{13}|z_{13} &\ge y_{12} + y_{19} \\
z_{14} &\le y_{15} + y_{19} & |W_{14}|z_{14} &\ge y_{15} + y_{19} \\
z_{16} &\le y_{12} + y_{21} & |W_{16}|z_{16} &\ge y_{12} + y_{21} \\
z_{17} &\le y_{12} + y_{21} & |W_{17}|z_{17} &\ge y_{12} + y_{21} \\
z_{18} &\le y_{12} + y_{19} & |W_{18}|z_{18} &\ge y_{12} + y_{19} \\
z_{20} &\le y_{15} + y_{19} & |W_{20}|z_{20} &\ge y_{15} + y_{19} \\
z_{22} &\le y_{21} & |W_{22}|z_{22} &\ge y_{21} \\
z_{23} &\le y_{19} & |W_{23}|z_{23} &\ge y_{19} \\
z_{24} &\le y_{19} & |W_{24}|z_{24} &\ge y_{19} \\
z_{25} &\le y_{19} & |W_{25}|z_{25} &\ge y_{19}
\end{aligned}$$

$$\begin{aligned}
p_{max} &\ge y_1 + y_3 + y_{12} + y_{15} + y_{19} + y_{21} \\
z_2, z_4, z_6, z_7, z_8, z_9, z_{10}, z_{11}, z_{12}, z_{13}, z_{14}&, \\
z_{16}, z_{17}, z_{18}, z_{20}, z_{22}, z_{23}, z_{24}, z_{25} &\in \{0, 1\} \\
y_1, y_3, y_{12}, y_{15}, y_{19}, y_{21} &\in \{0, 1\}
\end{aligned}$$

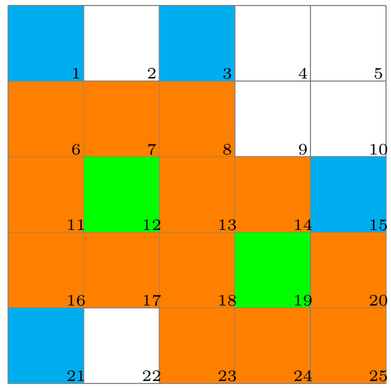


Figure 2.3: Picking two candidate parcels to maximize geographical coverage.

the maximum in this case with two candidate parcels.

We were able to see this result without having to do any calculations because of the simplicity of our model. But this is not always so apparent. We now begin to alter our model and consider other objectives.

2.1.2 Changing the Objective

We were previously only concerned with maximizing the geographical coverage of our model. In real life, prioritizing the geographical coverage of the potential parks serves to spread out the location of the parks and avoids building excessive parks in densely populated areas. It also proactively locates parks in places that may have yet to be developed or may be susceptible to rapid population growth in the future. However, we do not want to fully neglect the population density of the various city blocks. How does our model change when we add the number of inhabitants in each block? The question we are interested in now becomes:

How do we maximize the geographical coverage of the parks as well as the number of beneficiaries that would result from the construction of the parks?

We define the number of beneficiaries from a park construction as the sum of the population of all the blocks in the new park's service area.

We now turn to Figure 2.4 to see an updated version of our model city. The residential blocks have been updated with a number that indicates the

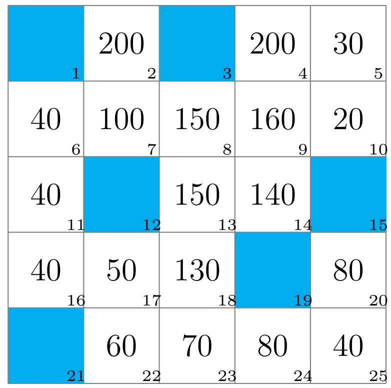


Figure 2.4: City blocks with population numbers.

number of people living in that city block. This allows us to calculate the number of beneficiaries that would result from the construction of a park on any candidate parcel. Building a park on candidate parcel 21, for example, would result in 150 beneficiaries. We define one additional variable.

Let p_i be the number of beneficiaries resulting from building a park on candidate parcel i. In the example of parcel 21, we would have $p_{21} = 150$. We can now update our model:

$$
\begin{aligned}
\max\ f_1 &= \sum_{j \in \mathcal{J}} z_j \\
\max\ f_3 &= \sum_{i \in \mathcal{I}} p_i y_i \\
\text{subject to } z_j &\leq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} \\
|\mathcal{W}_j|\, z_j &\geq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} \\
p_{max} &\geq \sum_{i \in \mathcal{I}} y_i \\
z_j &\in \{0, 1\}, j \in \mathcal{J} \\
y_i &\in \{0, 1\}, i \in \mathcal{I}
\end{aligned}
$$

Our model now has two objective functions. The first objective f_1 still seeks to maximize the geographical coverage of the parks. The second objective f_3 seeks to maximize the number of beneficiaries. But how do we maximize two things at the same time? We cannot guarantee that there will be a solution that maximizes the number of beneficiaries and the geographical coverage simultaneously. If we consider individual parcels, we notice that candidate parcel 3 serves the most amount of people, with $f_3 = 710$. However, candidate parcel 12 still has the most expansive geographical coverage, with $f_1 = 8$. How do we reconcile these two solutions?

2.2 A Solution Strategy

Because we are now trying to optimize multiple objectives, our question turns into a multiobjective optimization problem. We will present one possible way to solve these types of problems, borrowing the method used by the researchers from Bogotá. They implemented a lexicographic ordering of the objectives, which in practice meant that they ranked their objectives in accordance with their priorities. These priorities are reflected by the subscripts of the objective functions. Geographical coverage was deemed the most important criterion, which is why it is labeled f_1. Maximizing the number of beneficiaries was likewise deemed the third-most important criterion, so it is accordingly labeled f_3. For our model these are the only ones that we are concerned with, but it could work with any number of objective functions (for example, six).

The solution strategy is composed of two main parts. The first part is concerned with setting a benchmark for each objective function. The second part is about compromising to find a best possible solution. We will go over both parts in great detail.

2.2.1 Solving in Isolation

For the first part of our solution strategy, we will try to set a benchmark for each objective function by optimizing each of them in isolation. What does this mean? We return to our multiobjective optimization problem to explain the process.

$$
\begin{aligned}
\max\ f_1 &= \sum_{j\in\mathcal{J}} z_j \\
\max\ f_3 &= \sum_{i\in\mathcal{I}} p_i y_i \\
\text{subject to } z_j &\leq \sum_{i\in\mathcal{W}_j} y_i, j \in \mathcal{J} \\
|\mathcal{W}_j|\, z_j &\geq \sum_{i\in\mathcal{W}_j} y_i, j \in \mathcal{J} \\
p_{max} &\geq \sum_{i\in\mathcal{I}} y_i \\
z_j &\in \{0,1\}, j \in \mathcal{J} \\
y_i &\in \{0,1\}, i \in \mathcal{I}
\end{aligned}
$$

We see that our model has two objective functions, and we wish to solve for them in isolation. Our aim is to be able to treat each objective function as if it's the only objective function in our linear program. Since we have only two objective functions in this model, then one way to do this is by turning this model into two separate linear programs, as such:

$$
\begin{aligned}
\max\ f_1 &= \sum_{j\in\mathcal{J}} z_j \\
\text{subject to } z_j &\leq \sum_{i\in\mathcal{W}_j} y_i, j \in \mathcal{J} \\
|\mathcal{W}_j|\, z_j &\geq \sum_{i\in\mathcal{W}_j} y_i, j \in \mathcal{J} \\
p_{max} &\geq \sum_{i\in\mathcal{I}} y_i \\
z_j &\in \{0,1\}, j \in \mathcal{J} \\
y_i &\in \{0,1\}, i \in \mathcal{I}
\end{aligned}
\qquad
\begin{aligned}
\max\ f_3 &= \sum_{i\in\mathcal{I}} p_i y_i \\
\text{subject to } z_j &\leq \sum_{i\in\mathcal{W}_j} y_i, j \in \mathcal{J} \\
|\mathcal{W}_j|\, z_j &\geq \sum_{i\in\mathcal{W}_j} y_i, j \in \mathcal{J} \\
p_{max} &\geq \sum_{i\in\mathcal{I}} y_i \\
z_j &\in \{0,1\}, j \in \mathcal{J} \\
y_i &\in \{0,1\}, i \in \mathcal{I}
\end{aligned}
$$

Splitting our original multiobjective problem into multiple linear programs in this manner allows us to reduce our initial complicated problem into multiple, familiar, subproblems. We then solve each individual linear program, and we get the best possible result for each objective. Referring back to

Figure 2.4, we again see that the first linear program will give us the result $f_1 = 8$, because that is the maximum geographical coverage possible. The second linear program will give us the result $f_3 = 710$, because as we established earlier, this is the maximum amount of beneficiaries served by a candidate parcel.

Performing this process, however, brings us no closer to finding a suitable solution for both objective functions simultaneously, because as we saw, each optimal solution for the linear programs results in different candidate parcels being picked (parcels twelve and three, respectively). What we have now, though, are figures that serve as an upper bound and a benchmark for our multiobjective problem. We know that, given no other considerations, when we want to build one park we can serve at most eight lots. Similarly, we know that when only considering the amount of people living in surrounding lots, we can serve at most 710 of them by building one park. In a problem that seeks to maximize both of these figures, we can do no better than the individual optimal solutions we found.

In our final solution to the multiobjective problem, since we are trying to maximize both the geographical coverage and the number of beneficiaries served, then we want our solution to come as close to both optimal values as possible.

There are two stages to solving this model, as developed by the mathematicians working on this project. The first stage involves solving for each objective function independently, with the given constraints.

We will discuss the second stage at a later date.

Chapter 3

Multiobjective Facility Location Problems

3.1 What?

Here we will look at the general class of Multiobjective facility location problems, which is the class of problems that contains the main model/problem I will be looking at for my thesis.

Chapter 4

Discrete Facility Location Problems

4.1 What? pt. II

The problem I will be examining also falls under this category. I am not sure how much this would vary from the previous chapter. But if it turns out there is a significant difference between both types of problems, it might be helpful to provide a dedicated chapter for both. Maybe the only difference is that the other class of problems has many objective functions.

Chapter 5

Applications

If I have time to apply what I have learned to another problem, maybe this will go here. But this may be a bit ambitious. Stay tuned.