A New Mathematical Approach to Geographic Profiling

Mike O'Leary
Department of Mathematics
Towson University

December 14, 2009

Abstract

The primary question in geographic profiling is, given the locations of a series of crimes committed by the same serial offender, to estimate the location of that offender's anchor point. Currently, there are three main approaches to the problem, exemplified by the three software systems- CrimeStat, Dragnet, and Rigel. Though the details of the approaches taken by these software packages differ, they share a common mathematical heritage.

In our work, we developed an fundamentally new mathematical framework for the geographic profiling problem. Our framework meets our five essential goals:

1. The framework is mathematically rigorous;
2. There are explicit connections between assumptions on offender behavior and the components of the mathematical model;
3. The framework takes into account local geographic features; in particular it accounts for
   (a) Geographic features that influence the selection of a crime site, and
   (b) Geographic features that influence the potential anchor points of offenders;
4. The framework is based exclusively on data that is local to the jurisdiction(s) where the offenses occur; and
5. The framework returns a prioritized search area for law enforcement officers.

Our mathematical approach to this problem is based on Bayesian inference, and begins with the explicit ansatz that the offender's choice of targets depends only on (1) the distance between the target and the offender's anchor point, and (2) local geographic features of the target location. The algorithm requires a representative list of historical crimes of the same type as the series; this is used to estimate the local target attractiveness. The algorithm uses an estimate of the prior distribution of offender anchor points taken from local population density. The algorithm also estimates the structure of the offender's distance decay behavior; as a prior it takes a list of solved crimes together with the corresponding anchor point. With the model and appropriate priors selected, the mathematical framework then used Bayesian methods to develop an estimate for the probability distribution of the offender's anchor point based on the locations of the crime series.

This mathematical model has been implemented in software. Our tool is broken down into two packages. The first is a graphical user interface that lets the analyst or officer enter the required data, and comes complete with help features, and a set of instructions. The second tool performs the actual mathematical analysis; because it is separate from the user interface, it could potentially be used as a component in other tools.

The software tool is freely available for download, and is now currently being tested by both the Baltimore County Police Department as well as the Los Angeles Police Department. The source code for both tools are available, together with extensive documentation.

Contents

Chapter 1

Executive Summary

The primary question in geographic profiling is, given the locations of a series of crimes committed by the same serial offender, to estimate the location of that offender's anchor point. Currently, there are three main approaches to the problem, exemplified by the three software systems- CrimeStat, Dragnet, and Rigel. Though the details of the approaches taken by these software packages differ, they share a common mathematical heritage.

In our work, we developed an fundamentally new mathematical framework for the geographic profiling problem and implemented in prototype software that is now freely available for download[1]. Both our mathematical approach and our software tool are able to account for geographic features that affect the selection of the crime sites, geographic features that affect the distribution of potential anchor points, differences in the travel distances of different offenders, and certain demographic characteristics (race/ethnic group, age, and sex) of the offender.

To begin our discussion of our mathematical model, let us agree to adopt some common notation. A point $\mathbf{x}$ will have two components $\mathbf{x} = (x^{(1)}, x^{(2)})$. These can be latitude and longitude, or distances from a fixed pair of perpendicular reference axes. We presume that we are working with a series of n linked crimes, and the crime sites under consideration are labeled $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. We use the symbol $\mathbf{z}$ to denote the offender's anchor point. The anchor point can be the offender's home, place of work, or some other location of importance to the offender.

We begin by assuming that our offender chooses potential locations to offend randomly according to some unknown probability density function $P(\mathbf{x})$. The use of a probability distribution here is not meant to indicate that the offender is choosing targets randomly, though that may be the case. Instead it represents our lack of knowledge of the decision making process of the offender.

We assume that P depends on just two factors; first is the anchor point $\mathbf{z}$ of the offender. Second is the average distance that our offender is willing to travel to offend. We know that different offenders have a different willingness to travel and that the travel patterns of a fifteen year old will be different that those of a forty year old. For that reason we explicitly allow for the possibility that the offender's average travel distance may affect the choice of targets by the offender.

Thus, our initial mathematical model is that the offender with anchor point $\mathbf{z}$ and average offense distance α chooses to offend at the location $\mathbf{x}$ according to an unknown probability distribution $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$. The elements of the crime series $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ then represent a sample from this unknown distribution.

[1] `http://pages.towson.edu/moleary/Profiler.html`

Now we suppose that we specify the form of $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$; then the geographic profiling problem becomes a parameter identification problem for the unknown parameters $\mathbf{z}$ and α. We approach this problem using Bayesian methods to try to find a probability distribution for the unknown parameters. Standard Bayesian theory tells us that the posterior distribution of $\mathbf{z}$ and α given a series of crimes $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ is

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \frac{P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \,|\, \mathbf{z}, \alpha)\pi(\mathbf{z}, \alpha)}{P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)}.$$

Here $P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \,|\, \mathbf{z}, \alpha)$ is our model of offender behavior; it specifies the probability density that the offender will offend at all of the locations $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ given that they have anchor point $\mathbf{z}$ and average offense distance α. The factor $\pi(\mathbf{z}, \alpha)$ is the prior distribution of anchor point and offense distance; it represents what we know about the offender before we take into account the information from the crime series itself. The factor $P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ is the marginal distribution; since it does not depend on either $\mathbf{z}$ or α, we can remove it provided we replace the equality by a proportionality, so

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \propto P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \,|\, \mathbf{z}, \alpha)\pi(\mathbf{z}, \alpha).$$

We next assume that the different crime sites were chosen independently of one another. This choice is made because it is the simplest way we can relate the distribution $P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \,|\, \mathbf{z}, \alpha)$ to our model of offender behavior $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$. However, there is some evidence that criminal offense sites are not independent, which would require a more sophisticated model for offender behavior. Proceeding with the simplest model of independence though, we can then say that

$$P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \,|\, \mathbf{z}, \alpha) = P(\mathbf{x}_1 \,|\, \mathbf{z}, \alpha)P(\mathbf{x}_2 \,|\, \mathbf{z}, \alpha) \dots P(\mathbf{x}_n \,|\, \mathbf{z}, \alpha).$$

The simplest approach to the prior $\pi(\mathbf{z}, \alpha)$ is to assume that anchor points and offense distances are independent; then we can write

$$\pi(\mathbf{z}, \alpha) = H(\mathbf{z})\pi(\alpha).$$

Here $H(\mathbf{z})$ represents our knowledge of the distribution of offender anchor points before we incorporate information about the crime series. We take the approach that the distribution of anchor points can be modeled by population density; areas with high population density will correspond to areas with high anchor point density. We can then calculate $H(\mathbf{z})$ by simply using U.S. Census data; moreover because U.S. Census data is sorted by sex, age, and race / ethnic group all the way to the block level, we can incorporate these demographic factors when $H(\mathbf{z})$ is calculated. In practice, we calculate $H(\mathbf{z})$ from the Census data by using a kernel density parameter estimation technique

$$H(\mathbf{z}) = \sum_{i=1}^{N_{\text{blocks}}} = p_i K(\mathbf{z} - \mathbf{q}_i \,|\, \sqrt{A_i})$$

where each block has population p_i, center $\mathbf{q}_i$ and for each block we have chosen a different bandwidth equal to the side length of a square with the same area A_i as the block. Here $K(\mathbf{x} \,|\, \lambda)$ is a truncated quartic kernel with bandwidth λ, so that

$$K(\mathbf{x}|\lambda) = \begin{cases} \dfrac{3}{\pi\lambda^6}(|\mathbf{x}|^2 - \lambda^2)^2 & \text{if } |\mathbf{x}| \leq \lambda, \\ 0 & \text{if } |\mathbf{x}| \geq \lambda. \end{cases}$$

The prior distribution of offender average offense distance $\pi(\alpha)$ is more difficult to estimate. Indeed, though distribution of distances are available across offenses from aggregated historical data, we need to obtain the distribution of average offense distances across offenders. However, we are able to link the historical aggregate data with the behavior of individual offenders through a Fredholm integral equation, whose solution can then be estimated. This process is entirely automated within our software tool, and the user simply needs to obtain a list of historical solved crimes of the same general type as the crime series.

Combining these, we then see that we have the relationship

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \propto P(\mathbf{x}_1 \,|\, \mathbf{z}, \alpha) P(\mathbf{x}_2 \,|\, \mathbf{z}, \alpha) \ldots P(\mathbf{x}_n \,|\, \mathbf{z}, \alpha) H(\mathbf{z}) \pi(\alpha)$$

and since we are only interested in the distribution of the anchor points, we take the conditional distribution to obtain

$$P(\mathbf{z} \,|\, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \propto \int_0^\infty P(\mathbf{x}_1 \,|\, \mathbf{z}, \alpha) P(\mathbf{x}_2 \,|\, \mathbf{z}, \alpha) \ldots P(\mathbf{x}_n \,|\, \mathbf{z}, \alpha) H(\mathbf{z}) \pi(\alpha) \, d\alpha$$

which gives the probability distribution that the offender has anchor point $\mathbf{z}$, given that they have committed crimes at $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.

For this approach to be useful though, we still need to specify a model for offender behavior $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$. We take the approach that this has the form

$$P(\mathbf{x} \,|\, \mathbf{z}, \alpha) = D(d(\mathbf{x}, \mathbf{z}), \alpha) G(\mathbf{x}) N(\mathbf{z}, \alpha).$$

Here the factor $D(d(\mathbf{x}, \mathbf{z}), \alpha)$ accounts for the distance decay behavior of the offender. The term $d(\mathbf{x}, \mathbf{z})$ is the distance between the points $\mathbf{x}$ and $\mathbf{z}$. The mathematical framework does not force any particular choice of the distance metric on us, however since our software uses latitude and longitude internally to represent points, we use a spherical distance metric. Similarly, the mathematical framework does not force any particular form for the distance decay term. Our software uses a Rayleigh distribution; this was chosen because this is the distribution of distances that results from a bivariate normal distribution, and bivariate normal distributions appear naturally as the limit of random walks. Thus, we use the factor

$$D(d, \alpha) = \frac{\pi d}{2\alpha^2} \exp\left(-\frac{\pi d^2}{4\alpha^2} \right).$$

The factor $G(\mathbf{x})$ is present because some crimes can only occur at certain pre-defined locations, such as liquor store robberies or bank robberies. Even if the crime type is not limited to certain locations, it is still more likely to occur in some regions than others, like street robberies. This factor lets us account for these variations. It represents the local target attractiveness of a location, so that areas where $G(\mathbf{x})$ are considered to be more likely locations of an offense that regions where $G(\mathbf{x})$ is small. Rather than try to determine some *a priori* form for $G(\mathbf{x})$ based on criminological models, we instead simply assume that past behavior is a reasonable predictor of future behavior. In particular, to calculate $G(\mathbf{x})$, we start with a list of historical crimes $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N$. We then construct the local target attractiveness function using a kernel density parameter estimation technique by calculating

$$G(\mathbf{x}) = \sum_{i=1}^{N} K(\mathbf{x} - \mathbf{c}_i \,|\, \lambda)$$

where the bandwidth λ is twice the mean nearest neighbor distance between historical crime sites. This is essentially the same as one of the methods used to generate crime hot spots.

The last factor, $N(\mathbf{z}, \alpha)$ is a normalization factor, which ensures that P actually represents a probability distribution. It is determined by our choice of D and G, and has the form

$$N(\mathbf{z}, \alpha) = \frac{1}{\iint D(d(\mathbf{y}, \mathbf{z}), \alpha) G(\mathbf{y}) dy^{(1)} dy^{(2)}}.$$

This completes the specification of our mathematical framework, and shows us how we can estimate the location of the anchor point of an offender using as data:

- The locations of the crime series,
- The locations of historically similar crimes, to generate our estimate for $G(\mathbf{x})$,
- A list of solved historical crimes, with both the locations of the offense site and the corresponding anchor point, to generate our estimate of $\pi(\alpha)$, and
- Census data, along with any available demographic information about the offender, this lets us estimate $H(\mathbf{z})$.

This mathematical approach meets our five essential goals:

1. The framework is mathematically rigorous;
2. There are explicit connections between assumptions on offender behavior and the components of the mathematical model;
3. The framework takes into account local geographic features; in particular it accounts for
   (a) Geographic features that influence the selection of a crime site, and
   (b) Geographic features that influence the potential anchor points of offenders;
4. The framework is based exclusively on data that is local to the jurisdiction(s) where the offenses occur; and
5. The framework returns a prioritized search area for law enforcement officers.

Once the mathematical development was complete, we needed to turn these ideas into a practical tool that could be useful to law enforcement. To that end, we have developed software that implements these mathematical algorithms. The software was built in two parts:

- A program called Profiler that performs all of the mathematical analysis, and
- A program called ProfilerGUI with which the user interacts.

An analyst who wishes to use our algorithm need only download the software package and run ProfilerGUI. That program provides a convenient user interface that allows the user to enter all of the data necessary for the analysis in a simple form; it then calls the Profiler program to actually perform the analysis. When finished, it provides the user with a map of the proposed search area in .kml format; is also provides the analyst with a map of the target attractiveness function $G(\mathbf{x})$, as well as a map of the local population density $H(\mathbf{z})$.

Because the software was built in two components, others who wish to incorporate the mathematical analysis tools of Profiler are able to do so.

This project has resulted in two main implications for policy and practice. First by providing a new tool for the geographic profiling problem to police agencies, we hope to directly improve the clearance rate for series crimes.We presented our first functional prototype software package at the NIJ Conference in June 2009. The program is now being used by both the Los Angeles Police Department and the Baltimore County Police Department, both of whom are examining the effectiveness and usefulness of the tool.

It should be noted however, that we have not yet made a study of the effectiveness of the tool or of the mathematical algorithms that it contains.

On the other hand, by making our mathematics, algorithms and code widely and publicly available, we also hope that we can provide valuable insights to other researchers.

Chapter 2

Introduction

2.1 Statement of the problem

The purpose of our project was to develop and implement improved mathematical methods for the geographic profiling problem. Geographic profiling is a technique to used to make predictions about the location of a serial offender's anchor point or home base, based on the known locations of that offender's offenses. Prior to this work, three main software suites existed for geographic profiling. They are CrimeStat- developed by Ned Levine, Dragnet- developed by David Canter, and Rigel- developed by Kim Rossmo. These three software suites share a common mathematical heritage, and use mathematically related methods to determine their estimates of the offender's anchor point.

In our work, we developed an fundamentally new mathematical framework for the geographic profiling problem. Our framework meets our five essential goals:

1. The framework is mathematically rigorous;
2. There are explicit connections between assumptions on offender behavior and the components of the mathematical model;
3. The framework takes into account local geographic features; in particular it accounts for
    (a) Geographic features that influence the selection of a crime site, and
    (b) Geographic features that influence the potential anchor points of offenders;
4. The framework is based exclusively on data that is local to the jurisdiction(s) where the offenses occur; and
5. The framework returns a prioritized search area for law enforcement officers.

Ensuring that the mathematical algorithms are rigorous and making explicit the connections between the assumptions on offender behavior and components of the model helps in the analysis of the model. In particular, it gives researchers another tool to evaluate a model; one issue with the current generation of software suites for geographic profiling is that there is little agreement between the principals on how to compare the effectiveness of these differing approaches.

It is important that the mathematics explicitly allows for the influence of the local geography and demography. It is well known that there are relationships between the physical environment and crime rates; see for example Brantingham and Brantingham [8]. It is essential that a good mathematical framework possess the ability to incorporate this information into the model. That said, it is equally important that the model uses only data that is available to the appropriate law enforcement agency. Even if we were somehow able to create the perfect algorithm that always correctly predicts an offender's anchor point, that algorithm would still be worthless if it required data unavailable to law enforcement.

Finally, we recognize that simple point estimates of offender anchor points are not very valuable to practicing law enforcement officers. Rather, to be useful to practitioners, the algorithm must produce prioritized search areas.

There is clearly a need for improvement in geographic profiling. Indeed, the National Institute of Justice's Mapping Analysis and Public Safety web site [1] says

> "Though there have been anecdotal successes with geographic profiling, there have also been several instances where geographic profiling has either been wrong on predicting where the offender lives/works or has been inappropriate as a model. Thus far, none of the geographic profiling software packages have been subject to rigorous, independent or comparative tests to evaluate their accuracy, reliability, validity, utility, or appropriateness for various situations."

Some of the issues that face existing algorithms may be due to their common mathematical heritage. Indeed, it may be the case that the limiting feature of these existing algorithms actually lies in the mathematics upon which they are based. Clearly, no approach to geographic profiling can be more accurate than their underlying algorithm will let them be.

In our work we have developed a fundamentally new and rigorous mathematical technique that estimates the location of an offender's anchor point.

Our approach is to start with the assumption that an offender has a single stable anchor point. We then assume that the offender chooses crime locations according to a probability density function with two factors: first is a distance decay from the offender's anchor point, while the second is a local target attractiveness.

We make no *a priori* mathematical restrictions on the choice and form of the distance decay component. The mathematics allows for the use of the Euclidean, Manhattan, or any other distance metric; the form of the distance decay function can have any of the common mathematical forms-normal, negative exponential, or experimentally determined.

The local target density is determined by direct calculation from historical data. For example, if an analyst is examining a series of late night street robberies, then the algorithm requires a representative sample of historical late night street robberies. From this information, we estimate the likelihood that this particular crime occurs at any given location. This mathematical analysis proceeds under the assumption that the historical information is a reasonable predictor of current crime rates.

Given both the distance decay functional form, and the local target density, we obtain a family of probability density functions, one for each potential anchor point for the offender. We could then use maximum likelihood methods to estimate the anchor point of the offender [14]. However, this

[1] `http://www.ojp.usdoj.gov/nij/maps/gp.htm`, accessed August 2009

approach only provides a point estimate for the location of the offender's anchor point. Instead, we apply Bayesian techniques to estimate the probability distribution for the offender's anchor point, and so provide a map of the regions that are more or less likely to contain the offender's anchor point.

We have created and developed the mathematics necessary for this analysis, and implemented the resulting algorithms in software. The tool we have developed has now been released, and is being tested by both the Baltimore County Police Department and the Los Angeles Police Department.

2.2 Literature

To begin our review of the current state of geographic profiling, let us agree to adopt some common notation. A point $\mathbf{x}$ will have two components $\mathbf{x} = (x^{(1)}, x^{(2)})$. These can be latitude and longitude, or distances from a fixed pair of perpendicular reference axes. We presume that we are working with a series of n linked crimes, and the crime sites under consideration are labeled $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. We use the symbol $\mathbf{z}$ to denote the offender's anchor point. The anchor point can be the offender's home, place of work, or some other location of importance to the offender.

An important first question in geographic profiling is how to measure the distance between points. One common method is to use the usual notion of distance, called the Euclidean distance. In this case, the distance $d(\mathbf{x}, \mathbf{y})$ between two points $\mathbf{x}$ and $\mathbf{y}$ is given by $d_2(\mathbf{x}, \mathbf{y})$ where

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{[x^{(1)} - y^{(1)}]^2 + [x^{(2)} - y^{(2)}]^2}$$

Another approach is to use the Manhattan distance; in this case the distance between $\mathbf{x}$ and $\mathbf{y}$ is given by $d_1(\mathbf{x}, \mathbf{y})$ where

$$d_1(\mathbf{x}, \mathbf{y}) = |x^{(1)} - y^{(1)}| + |x^{(2)} - y^{(2)}|$$

There are other choices that can be used as well; for example the total street distance following the local road network, or the total time to make the trip while following the local road network.

One important difference to note between these distances is that the Euclidean distance gives the same results regardless of the choice of the coordinate axes; in particular, rotating a pair of points around a third does not change the distance between the pair. This result does not hold for the Manhattan distance, and so the coordinate axes need to be chosen with care when using the Manhattan distance.

Following Snook, Zito, Bennell and Taylor [52], we classify algorithms for geographic profiling into two general categories- spatial distribution strategies and probability distance strategies.

Spatial distribution strategies. The simplest of the spatial distribution strategies is to estimate the offenders anchor point $\mathbf{z}$ by the centroid ζ_{centroid} of the crime series [27]. The centroid, also known as the center of mass is defined to be

$$\zeta_{\text{centroid}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i .$$

Another approach is to estimate the anchor point $\mathbf{z}$ by the center of minimum distance ζ_{cmd}. This is chosen to be the value of $\mathbf{y}$ for which the sum of the distances from the point to the crime sites

$$D(\mathbf{y}) = \sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{y})$$

is a minimum. Here different choices of the distance function d lead to different choices for the center of minimum distance. Unlike the centroid, there is no simple formula that gives the value of ζ_{cmd}; however, there are a number of efficient algorithms that can approximate it to any desired accuracy.

Another method used to estimate the anchor point is the circle method of Canter and Larkin [12]. They make a distinction between two types of offenders; marauders and commuters. A marauder is assumed to move from a home base, commit a crime, and then return, where the base acts as a focal point for the crime series. In this case, the offender's home range and criminal range overlap. A commuter, on the other hand, first moves from a home base to another area, then commits the crime. As a consequence, the focal point for the crime series is different from the home base, and the offender's home range and criminal range do not overlap

Their circle hypothesis is the following. Given a series of linked crimes committed by a marauder, draw a circle whose diameter are the two crimes locations that are the farthest apart. Then all of the crimes in the series would be in the resulting circle, and the offender's home base would also lie in the circle so drawn.

There is evidence for the validity of the circle hypothesis. In their original paper, Canter and Larkin [12] examined a collection of 45 male sexual assaults in Britain. In 41 of the 45 cases the circle correctly encompassed all of the crime sites and in 39 of the 45 cases the circle correctly contained a base for the offender. Kocsis and Irwin [22] examined 24 rape series, 22 arson series, and 27 burglary series in Australia. The circle contained all of the crimes for 79% of the rape series, 82% of the arson series, and 70% of the burglary series, while the circle correctly contained the home base of the offender for 71% of the rape cases, 82% of the arson cases, and 48% of the burglary cases. This last result suggests that the marauder hypotheses may not necessarily be appropriate for burglary.

These results were amplified by Meany [35], who showed that burglars were more likely to act as commuters than non-burglars, while arsonists and sex offenders were more likely to act as marauders than non-arsonists / non-sex offenders. Similarly, Kocsis, Cooksey, Irwin and Allen [21] found in their study of 58 burglaries that occurred in rural Australian towns, that the circle theory was less effective. Laukkanen and Santtila [25] examined 76 commercial robbery series, and found only 30 (=39%) that satisfied the circle hypothesis. Note however, that many of these series were very short; 62 of the 76 series analyzed contained either two or three crimes.

Probability distribution strategies In contrast to the spatial distribution strategies are the probability distance strategies. These methods are currently employed in the major computer programs for geographic profiling (CrimeStat, Dragnet, and Rigel). All have the common idea of constructing a hit score by summing the values of some decay function of the distances between a general point and the elements of the crime series. Mathematically, they all begin by first making a choice of distance metric d; they then select a decay function f and construct a hit score function $S(\mathbf{y})$ by computing

$$S(\mathbf{y}) = \sum_{i=1}^{n} f(d(\mathbf{x}_i, \mathbf{y})) = f(d(\mathbf{x}_1, \mathbf{y})) + \cdots + f(d(\mathbf{x}_n, \mathbf{y})). \tag{2.1}$$

Regions with a high hit score are considered to be more likely to contain the offender's anchor point than regions with a low hit score. In practice, the hit score $S(\mathbf{y})$ is not evaluated everywhere, but simply on some rectangular array of points $\mathbf{y}_{jk} = (y_j^{(1)}, y_k^{(2)})$ for $j \in \{1, 2, \ldots J\}$ and $k \in \{1, 2, \ldots, K\}$, giving us the array of values $S_{jk} = S(\mathbf{y}_{jk})$.

In effect, these approaches center a tent function on each of the crime sites; the hit score is then found by summing the results. The differences between the different methods are solely in the choice of the tent function.

Rossmo's method, as described in [41, Chapter 10] chooses the Manhattan distance function for d and the decay function

$$f(d) = \begin{cases} \dfrac{k}{d^h} & \text{if } d < B, \\ \dfrac{kB^{g-h}}{(2B-d)^g} & \text{if } d \geq B. \end{cases}$$

We remark that Rossmo also considers the possibility of forming hit scores by multiplication; see [41, p. 200]

The method described by Canter, Coffey, Huntley and Missen in [11] is to use a Euclidean distance, and to choose either a decay function in the form

$$f(d) = e^{-\beta d}$$

or functions with a buffer and plateau, with the form

$$f(d) = \begin{cases} 0 & \text{if } d < A, \\ B & \text{if } A \leq d < B, \\ Ce^{-\beta d} & \text{if } d \geq B. \end{cases}$$

The CrimeStat program described in [29] uses Euclidean or spherical distance and gives the user a number of choices for the decay function, including

- Linear: $f(d) = A + Bd$,
- Negative exponential: $f(d) = Ae^{-\beta d}$,
- Normal: $f(d) = A(2\pi S^2)^{-1/2} \exp\left[-(d - \bar{d})^2/2S^2\right]$,
- Lognormal: $f(d) = A(2\pi d^2 S^2)^{-1/2} \exp\left[-(\ln d - \bar{d})^2/2S^2\right]$, and
- Truncated negative exponential: $f(d) = Bd$ if $d < C$ and $f(d) = Ae^{-\beta d}$ if $d \geq C$.

CrimeStat also allows the user to use empirical data to create a different decay function matching a set of provided data as well as the use of indirect distances.

Though each of these approaches are distinct, they share the same underlying mathematical structure; they vary only in the choice of decay function and the choice of distance metric.

Bayesian methods We remark that the latest version (3.2) of CrimeStat contains a new Bayesian Journey to Crime Module that integrates information on the origin location of other offenders who committed crimes in the same location with the distance decay estimates [30]. In this approach, the geographic region under consideration is subdivided into subregions R_i; then the matrix C_{ij}

is formed, which counts the number of crimes that occur in region R_i committed by an offender with anchor point in R_j. This is then used as the prior distribution for the Bayesian analysis. In particular if a series of crimes is concentrated in region R_i, then by examining the matrix C_{ij}, regions R_j for which C_{ij} is large are considered to be more likely to contain the offender's anchor point than regions where C_{ij} is small. Levine and Block [31]have tested this method with data from Baltimore County and from Chicago.

Controversies. There is some question as to whether or not computer GIS systems implementing the existing algorithms are as effective as simply providing humans with some simple heuristics; *c.f.* the discussion in Snook, Canter, and Bennell [48], Snook, Taylor and Bennell [49], Rossmo [43] and Snook, Taylor and Bennell [51]; see also the discussion in Snook, Taylor and Bennell [50], Rossmo and Filer [44], Bennell, Snook and Taylor [4] and Rossmo, Filey and Sesley [45]; see also the paper of Bennell, Snook, Taylor, Corey, and Keyton [5].

There have also been significant disagreements in the literature as to what is the best methodology to evaluate the currently existing geographic profiling software. See the original report prepared for NIJ by Rich and Shively [40], the critique of Rossmo [42], and the response of Levine [28].

Distance Decay. An important first question for all geographic profiling methods is to determine the appropriate metric for measuring distance; we have already seen that different existing methods choose different metrics for measuring distance. Wiles and Costello [58] point out that British cities lack the typical rectangular gird pattern of streets common to American cities, and that it was generally difficult to identify the most likely route between two locations. As a consequence, they felt that in their study it was better to use Euclidean distance as opposed to the Manhattan distance or actual travel times or distances.

It is generally accepted that there is a real distance decay function, and that offenders are less likely to commit crimes as the distance from the offense site to the anchor point increases. Snook [47] studied serial burglars in Newfoundland. He found targets close to offender's homes were far more likely to be burgled than targets farther away. Snook also found that the travel distance was found to have significant relationships with the offender's age, method of transportation and the value of the stolen property.

Warren, Reboussin, Hazelwood, Cummings, Gibbs and Trumbetta [56] analyzed a series of 565 rapes. They also found distance decay, and analyzed the relationships between crime scene behavior and the distances traveled by the offender. Van Koppen and Jansen [24] studied 434 commercial robberies committed by 585 robbers in the Netherlands, and examined the relationship between the distance traveled to the crime scene with characteristics of the offenders and offenses. Fritzon [17] studied the relationships between the characteristics and motivations of arsonists with the distance the arsonist traveled to set the fire. Finally, Lu [33] looked at distances traveled by offenders after a crime. In particular, she studied auto thefts in Buffalo NY, and examined the distance from the location of the theft to the location the vehicle was recovered.

Van Koppen, and De Keijser [23] took issue with the use of aggregate data to determine individual distance decay functions; however Rengert, Piquero, and Jones [39] disagree with many of their conclusions.

Finally, we mention López [32] who studied 20 residential burglars in the Kennemerland region. In this study, they normalized each of the distances that the burglars traveled to their crime sites, and concluded that their data provided no evidence for the existence of a distance decay function.

Influence of geographic and demographic factors. Relationships between environmental factors and crimes have been studied extensively. White [57] examined relationships between felony rates in Indianapolis and social and geographic factors sorted by census tracts. Brantingham and Brantingham [7] examined burglaries in Tallahassee to study the spatial patterning of crime; they found that burglary rates were higher in blocks that formed the boundary of a neighborhood than blocks that were in the interior of a neighborhood. Brown [9] studied the relationship between crime rates and various environmental factors in Chicago, while Wang and Minor [55] found relationships between geographic availability of employment and crime rates in Cleveland.

Bernasco and Nieuwbeerta [6] studied residential burglaries in The Hague, Netherlands. They used a discrete spatial choice approach to ascertain the existence of a relationship between neighborhood burglary rates and neighborhood characteristics, including ethnic heterogeneity and the number of residential units in the neighborhood. Osborn and Tseloni [37] analyzed social and demographic data from the British Crime survey, and found that some household characteristics affect the likelihood of a number of property crimes. Similar results were found by Malczewski, Poetz, and Iannuzzi [34] in their study of residential burglaries in London Ontario. Buettner and Spengler [10] also showed that some local socio-economic factors were significant for both property and violent crime. Groff and La Vigne [18] created a model that looked at local variables like land use to predict areas with potentially desirable targets for burglary; they then empirically tested their model.

Interestingly, Tseloni, Wittebrood, Farrell and Pease [53] compared geographic features that influenced burglary rates across three different countries (Britain, the U.S., and the Netherlands). Though the effect of some variables on crime rates appeared consistent across the different nations, there were some variables that were significant in different nations, but in opposite directions. For example, increased household affluence indicated higher burglary rates in Britain, while it indicated lower burglary rates in the U.S.

2.3 Statement of hypothesis

The goal of our research was to develop a new mathematical approach to the geographic profiling problem and to produce a tool that could be used as a practical aid to investigations.

As valuable as existing tools have been, they are based on the mathematical notion of hit scores, and it is unclear how this technique might be extended to incorporate available geographic information that may affect the selection of a crime site or the location of an offender's anchor point. Another issue is the fact that each of the existing approaches to the geographic profiling problem is, at its core, a model of how offenders behave. However, it is unclear what each of these approaches actually says about how offenders behave.

Our first goal then, was to develop a more robust mathematical framework for the geographic profiling problem. We wanted to create a mathematical approach that allowed us to incorporate available sources of geographic information affecting both crime site selection and anchor point selection. We also wanted to make sure that our approach made explicit any assumptions about offender behavior so that they could be examined, studied, critiqued, and improved.

The second goal of our project was to turn these new mathematical ideas into a practical tool. Indeed, we wanted to be sure that the mathematical techniques that we developed were more than just theoretical constructs, but something that will have a practical impact on law enforcement.

In particular, we wanted to create a piece of software that can be provided to police departments for their use and evaluation. The source code for this software will be made available under an appropriate open source license so that other researchers and tool developers can incorporate not only the model but the code as well.

That said, we are not claiming that the approach we have created is necessarily better than existing methods. So far, we have been able to develop the mathematical approach and implement it in software; currently the technique is now being tested for effectiveness.

It is also important to note that the idea behind our research is to find a way to extract all of the information about the geographic location of the offender's anchor point from the information in the crime series. However, this does not necessarily mean that the tool will be successful. As one example, the crime series may not contain much geographic information about the offender, as might be the case for a series of bank robberies that are widely dispersed and occur near ramps to major interstates. It also may be the case that some of the underlying assumptions about the offender may be incorrect- perhaps the offender does not have a single stable anchor point during the series.

Chapter 3

Methods

3.1 The mathematical model

We begin by looking for an appropriate model for offender behavior and start with the simplest possible situation– where we know nothing about the offender. Thus, we assume that our offender chooses potential locations to offend randomly according to some unknown probability density function $P(\mathbf{x})$. For any geographic region R, the probability that our offender will choose a crime site in R can be found by adding up the values of P in R, giving us the probability $\iint_R P(\mathbf{x})\, dx^{(1)}\, dx^{(2)}$.

At first glance, it may seem odd to use a probabilistic model to describe human behavior. In fact, probabilistic models are commonly used to describe many kinds of apparently deterministic phenomena. For example, classical models of the diffusion of heat or chemical concentration can be derived probabilistically; they also see application in models of the stock market [1], [59], [60], in models of population genetics [16], and in many other models [3].

More precisely, the probability density function P represents our knowledge of the behavior of the offender. We use a probability distribution, not because the offender's decision has a random component, although it may. Rather, we use a probability density because we lack complete information about the offender. Indeed, consider the following thought experiment. If we want to model the flip of a coin, we use probability and assume that each side of the coin is apt to occur half the time. Now instead of flipping the coin, let us take the coin to a colleague and ask them to choose a side. In this case the outcome is the deliberate result of a decision by an individual. However without knowing more information about our colleague's preferences, the best choice to model the outcome of that experiment is still the use of a probability distribution.

Returning to our model of offender behavior, we begin with a question: upon what sorts of variables should our probability density function P depend? One of the fundamental assumptions of geographic profiling is that the choice of an offender's target locations is influenced by the location of the offender's anchor point $\mathbf{z}$. Therefore, we assume that P depends upon $\mathbf{z}$. Underlying this approach are the requirements that the offender has a single anchor point and that it is stable during the crime series.

A second important factor is the distance our offender is willing to travel to commit a crime. Let α denote the average distance that our offender is willing to travel to offend. We allow for the possibility that this value varies between offenders. Combining these, we assume that there is a probability density function $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$ for the probability that an offender with a single stable anchor point $\mathbf{z}$ and average offense distance α commits a crime at the location $\mathbf{x}$.

We assume that this model is local to the jurisdiction under consideration. In particular, we explicitly allow for the possibility that different models $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$ may need to be chosen for different jurisdictions.

The key mathematical point is that the unknown is now the entire distribution $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$, rather than just the anchor point $\mathbf{z}$. On its face, it seems a step backwards, but in fact, it is not. Indeed, let us suppose that the *form* of the distribution P is known, but that the values of the anchor point $\mathbf{z}$ and average offense distance α are unknown. Then the problem can be stated mathematically as, given a sample $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ (the crime site locations) from the distribution $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$ with parameters $\mathbf{z}$ and α to determine the best way to estimate the parameter $\mathbf{z}$ (the anchor point).

For the moment, let us set aside the question of what reasonable choices can be made for the form of the distribution $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$, and focus on how we can estimate the anchor point $\mathbf{z}$ from our knowledge of the crime locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

It turns out that this is a well studied mathematical problem. One approach is to use the maximum likelihood estimator. To do so, one first forms the likelihood function:

$$L(\mathbf{y}, a) = \prod_{i=1}^{n} P(\mathbf{x}_i \,|\, \mathbf{y}, a) = P(\mathbf{x}_1 \,|\, \mathbf{y}, a) \cdots P(\mathbf{x}_n \,|\, \mathbf{y}, a).$$

Then the maximum likelihood estimates $\hat{\mathbf{z}}_{\text{mle}}$ and $\hat{\alpha}_{\text{mle}}$ are the values of $\mathbf{y}$ and a that make L as large as possible. Equivalently, one can maximize the log-likelihood function

$$\lambda(\mathbf{y}, a) = \sum_{i=1}^{n} \ln P(\mathbf{x}_i \,|\, \mathbf{y}, a) = \ln P(\mathbf{x}_1 \,|\, \mathbf{y}, a) + \cdots + \ln P(\mathbf{x}_n \,|\, \mathbf{y}, a).$$

Though rigorous, this approach is unsuitable as simple point estimates for the offender's anchor point are not operationally useful. Instead, we continue our analysis by using Bayes' Theorem.

3.1.1 Bayesian Analysis

To see how Bayesian methods can be applied to geographic profiling, we begin with the simplest case where the offender has only committed one crime at the location $\mathbf{x}$. We would like to use the information from this crime location to form an estimate for the probability distribution for the anchor point $\mathbf{z}$. Bayes' Theorem gives us the estimate

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}) = \frac{P(\mathbf{x} \,|\, \mathbf{z}, \alpha)\pi(\mathbf{z}, \alpha)}{P(\mathbf{x})} \tag{3.1}$$

[13, 14]. Here $P(\mathbf{z}, \alpha \,|\, \mathbf{x})$ is the posterior distribution, which gives the probability density that the offender has anchor point $\mathbf{z}$ and average offense distance α, given that the offender has committed a crime at the location $\mathbf{x}$.

The term $P(\mathbf{x})$ is the marginal distribution. The important thing to note is that it is independent of $\mathbf{z}$ and α, therefore it can be ignored provided we replace the equality in (3.1) with proportionality.

The term $\pi(\mathbf{z}, \alpha)$ is the prior distribution. It represents our knowledge of the probability density that the offender has anchor point $\mathbf{z}$ and average offense distance α before we incorporate any information about the crime series. One approach to the prior is to assume that the anchor point $\mathbf{z}$ is mathematically independent of the average offense distance α. In this case, we can factor to obtain

$$\pi(\mathbf{z}, \alpha) = H(\mathbf{z})\pi(\alpha) \tag{3.2}$$

where $H(\mathbf{z})$ is the prior probability density function for the distribution of anchor points before any information from the crime series is included and $\pi(\alpha)$ is the probability density function for the prior distribution of the offender's average offense distance, again before any information from the crime series is included.

Combining these, we then obtain the expression

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}) \propto P(\mathbf{x} \,|\, \mathbf{z}, \alpha)H(\mathbf{z})\pi(\alpha).$$

Of course, we are interested in crime series, and we would like to estimate the probability density for the anchor point $\mathbf{z}$ given our knowledge of all of the crime locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$. To do so, we proceed in a similar fashion; now Bayes' Theorem implies

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{P(\mathbf{x}_1, \ldots, \mathbf{x}_n \,|\, \mathbf{z}, \alpha)\pi(\mathbf{z}, \alpha)}{P(\mathbf{x}_1, \ldots, \mathbf{x}_n)}.$$

Here $P(\mathbf{z}, \alpha \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n)$ is again the posterior distribution, which gives the probability density that the offender has anchor point $\mathbf{z}$ and average offense distance α, given that the offender has committed a crime at each of the locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$. The marginal $P(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ remains independent of $\mathbf{z}$ and α, and can be ignored; the prior π can be handled by (3.2). Then

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n) \propto P(\mathbf{x}_1, \ldots, \mathbf{x}_n \,|\, \mathbf{z}, \alpha)H(\mathbf{z})\pi(\alpha). \tag{3.3}$$

The factor $P(\mathbf{x}_1, \ldots, \mathbf{x}_n \,|\, \mathbf{z}, \alpha)$ on the right side is the joint probability that the offender committed crimes at all of the locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ given that they had anchor point $\mathbf{z}$ and average offense distance α. The simplest assumption we can make is that all of the offense sites are mathematically independent; then we have the reduction

$$P(\mathbf{x}_1, \ldots, \mathbf{x}_n \,|\, \mathbf{z}, \alpha) = P(\mathbf{x}_1 \,|\, \mathbf{z}, \alpha) \cdots P(\mathbf{x}_n \,|\, \mathbf{z}, \alpha). \tag{3.4}$$

Substituting this into (3.3) gives

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n) \propto P(\mathbf{x}_1 \,|\, \mathbf{z}, \alpha) \cdots P(\mathbf{x}_n \,|\, \mathbf{z}, \alpha)H(\mathbf{z})\pi(\alpha).$$

Finally, since we are only interested in the location of the anchor point $\mathbf{z}$, we take the conditional distribution to obtain our fundamental mathematical result:

$$P(\mathbf{z} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n) \propto \int_0^\infty P(\mathbf{x}_1 \,|\, \mathbf{z}, \alpha) \cdots P(\mathbf{x}_n \,|\, \mathbf{z}, \alpha)H(\mathbf{z})\pi(\alpha)\, d\alpha. \tag{3.5}$$

The expression $P(\mathbf{z} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n)$ gives us the probability density that the offender has anchor point $\mathbf{z}$ given that they have committed crimes at the location $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Because we are calculating probabilities, this immediately provides us a rigorous search area for the offender. Indeed regions with larger values of $P(\mathbf{z} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n)$ by definition are more likely to contain the offender's anchor point than regions where $P(\mathbf{z} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n)$ is lower.

This is a very general framework for the geographic profiling problem. There are many choices for the model of offender behavior $P(\mathbf{x} \mid \mathbf{z}, \alpha)$, and we will later examine a number of reasonable choices. Though the preceding used a model for offender behavior with one parameter α other than the anchor point, the mathematics continues to hold with elementary modifications if we either add additional parameters or remove the parameter α.

In addition to an assumption as to the form of $P(\mathbf{x} \mid \mathbf{z}, \alpha)$, we have made two other fundamental assumptions. One is that the prior $H(\mathbf{z})$ for $\mathbf{z}$ is independent of the prior $\pi(\alpha)$ for α. This is a reasonable first assumption, and it is what allows us the factorization in (3.2). Its significance is that we are assuming that the average distance that the offender is willing to travel is independent of the offender's anchor point. This is probably most appropriate in urban areas and for regions where offenders travel short distances to offend. On the other hand, the assumption may be less valid for example, in a town surrounded by a less populated rural area. If potential offense locations are concentrated in the town, then offenders with anchor points far from the town will likely have a higher average offense distance than offenders with anchor points in town.

The remaining fundamental assumption is that the offender's choice of crime sites are independent; this is necessary for the factorization in (3.4). It can be replaced by other assumptions, but would require a different model for the joint distribution than the simple expression in (3.4). Though reasonable as a first assumption, there is evidence of deviation from independence in the literature. For example Kocsis, Cooksey, Irwin and Allen [21] found in their analysis of 58 multiple burglary cases in rural Australia that the crime sites tended to lie in narrow corridors emanating from the offenders anchor point. Meaney [35] examined 83 burglary series, 23 sexual offense series, and 21 arson series; she found that the first offense occurred closer to the offender's home than the last offense, suggesting that there is a temporal component to offender's site selection. On the other hand, Laukkanen and Santtila [25] concluded that the distance a robber travels to offend did not increase as the crime series progressed. We also mention Ratcliffe [38] who examined some of the interrelation between temporal data and routine activity theory.

3.1.2 Simple Models for Offender Behavior

If our fundamental mathematical result is to have any practical or investigative value, we need to be able to construct reasonable choices for our model of offender behavior. One simple model is to assume that the offender chooses a target location based only on the Euclidean distance from the offense location to the offender's anchor point and that this distribution is (bivariate) normal. In this case we obtain

$$P(\mathbf{x} \mid \mathbf{z}, \alpha) = \frac{1}{4\alpha^2} \exp\left(-\frac{\pi}{4\alpha^2}|\mathbf{x} - \mathbf{z}|^2\right). \tag{3.6}$$

If we make the prior assumptions that all offenders have the same average offense distance α and that all anchor points are equally likely, then

$$P(\mathbf{z} \mid \mathbf{x}_1, \dots, \mathbf{x}_n) = \left(\frac{1}{4\alpha^2}\right)^n \exp\left(-\frac{\pi}{4\alpha^2}\sum_{i=1}^{n}|\mathbf{x}_i - \mathbf{z}|^2\right).$$

We see that the posterior anchor point probability distribution is just a product of normal distributions, one centered at each crime site; compare this to sums used in the calculation of hit scores (2.1). We also mention that in this model of offender behavior, the maximum likelihood estimate for the anchor point is simply the mean center of the crime site locations; this is also the mode of the posterior anchor point probability distribution $P(\mathbf{z} \mid \mathbf{x}_1, \ldots, \mathbf{x}_n)$.

Another reasonable choice of a model for offender behavior is to assume that the offender chooses a target location based only on the Euclidean distance from the offense location to the offender's anchor point, but that now the distribution is a (bivariate) negative exponential so that

$$P(\mathbf{x} \mid \mathbf{z}, \alpha) = \frac{2}{\pi\alpha^2} \exp\left(-\frac{2}{\alpha}|\mathbf{x} - \mathbf{z}|\right). \tag{3.7}$$

Once again, if our prior assumptions are that all offenders have the same average offense distance and that all anchor points are equally likely, then

$$P(\mathbf{z} \mid \mathbf{x}_1, \ldots, \mathbf{x}_n) = \left(\frac{2}{\pi\alpha^2}\right)^n \exp\left(-\frac{2}{\alpha}\sum_{i=1}^{n}|\mathbf{x}_i - \mathbf{z}|\right).$$

We see that this is just a product of negative exponentials centered at each crime site. Further, the corresponding maximum likelihood estimate for the offender's anchor point is simply the center of minimum distance for the crime series locations. Finally, if we construct the function $\tilde{S}(\mathbf{z}) = \ln P(\mathbf{z} \mid \mathbf{x}_1, \ldots, \mathbf{x}_n)$, then $\tilde{S}$ is a hit score in the same form as (2.1) with a linear decay function f and Euclidean distance d.

This preceding analysis was predicated on the assumption that all offenders have the same average offense distance α and that this was known in advance. Similarly, the existing hit score methods all rely on decay functions f with one or more parameters that also need to be determined in advance. Unlike the hit score techniques however, our method does not require that we make a choice for the parameter α in advance. For example, if we assume only that the offender has a distance decay in the form (3.6) (or in the form (3.7)), with α unknown, then the maximum likelihood technique will estimate both the anchor point $\mathbf{z}$ and the average offense distance α. Our fundamental mathematical result (3.5) also does not require that the parameter α be determined in advance, though it does require a prior estimate $\pi(\alpha)$ for the distribution of average offense distances.

3.1.3 More Realistic Models for Offender Behavior

These simple models for offender behavior show that our framework recaptures many existing geographic profiling techniques; however, this new method is more general and allows us a simple way to incorporate geographic features into the model. Indeed, let us suppose that offender target selection depends on more than just the distance from the anchor point to the crime site locations, but that it depends on some features in the local geography. One way to account for this is to suppose that the offense probability density is proportional to both a distance decay term and to a function that measures the attractiveness of a particular target location. Doing so, we obtain the following expression

$$P(\mathbf{x} \mid \mathbf{z}, \alpha) = D(d(\mathbf{x}, \mathbf{z}), \alpha)G(\mathbf{x})N(\mathbf{z}, \alpha). \tag{3.8}$$

Here the factor D models the effect of distance decay using the distance metric $d(\mathbf{x}, \mathbf{z})$. For example, we can specify a normal decay, so that

$$D(d, \alpha) = \frac{1}{4\alpha^2} \exp\left(-\frac{\pi}{4\alpha^2}d^2\right).$$

We could also specify a negative exponential decay, so

$$D(d, \alpha) = \frac{2}{\pi \alpha^2} \exp\left(-\frac{2}{\alpha} d\right),$$

but of course there are many other reasonable possibilities.

One of the consequences of this approach to distance decay is that it assumes uniformity of travel direction with respect to the given distance metric; in this way it simplifies actual travel behavior. It may be the case that certain directions are preferred by the offender; for example when searching for potential targets, the offender may prefer to move closer to an urban area than farther away. A new approach to the modeling this distance decay effect is the kinetic random walk model of [36].

The factor $G(\mathbf{x})$ is used to account for the local geographic features that influence the selection of a crime site. High values for $G(\mathbf{x})$ indicate that $\mathbf{x}$ is a likely target for typical offenders; low values indicate $\mathbf{x}$ is a less likely target.

The remaining factor N is a normalization required to ensure that P is a probability distribution. Its value is completely determined by the choices of D and G and has the form

$$N(\mathbf{z}, \alpha) = \frac{1}{\iint D(d(\mathbf{y}, \mathbf{z}), \alpha) G(\mathbf{y}) dy^{(1)} dy^{(2)}}. \tag{3.9}$$

Returning to the influence of geography on target selection, one simple example of $G(\mathbf{x})$ is to account for jurisdictional boundaries. Suppose that all known crimes in the series must occur in a region J. The offender can commit crimes outside J; but these are presumed unknown to the analyst; the offender's anchor point may also reside outside the region J. We can account for this with the simple model

$$G(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in J, \\ 0 & \text{if } \mathbf{x} \notin J. \end{cases}$$

In practice, the region J corresponds to one or more jurisdictions sharing information about the offender's crime series.

The incorporation of this very simple geographic information has some surprising consequences. In particular, the algorithm is able to distinguish between areas where no crimes in the series have occurred (inside J) from areas where there is no information as to whether or not a crime in the series may have occurred (outside J). For example, suppose that the elements of a hypothetical crime series are all near the southern boundary of a jurisdiction J. Then the algorithm will return a search area skewed to the south of the crime series because the algorithm "knows" that no known crimes take place north of the series, but that there may be crimes to the south of the series that are unknown to the analyst; thus the offender is more likely to live south of the series than to the north. As a consequence, this model does not suffer from the convex hull effect described by Levine [28].

This simple approach to geographic information affecting the selection of the target is primarily illustrative; clearly a better model can be chosen. To do so, one approach would be to use available geographic and demographic data and the correlations between crime rates and these variables that have already been published to construct an appropriate choice for $G(\mathbf{x})$. However, this approach has a number of issues. First, is the fact that different crime types have different etiologies; in particular their relationship to the local geographic and demographic backcloth depends strongly on the particular type of crime. This would limit the method to only those crimes where this relationship has been well studied. Moreover, even for well studied crimes, there are regional differences. Indeed, [53] noted that increased household affluence indicated higher burglary rates in Britain, and indicated lower burglary rates in the U.S.

The primary issue here is that this approach posits a method to explain crime rates by looking for explanatory variables. However, from the perspective of geographic profiling, it is unnecessary to explain; instead we can simply acknowledge these differences, and work on measuring the resulting differences. Rather than look at the local geographic variables, we can use historical data to model the geographic target attractiveness.

In particular, let us assume that historical crime rates are reasonable predictors of the likelihood that a particular region will be the site of an offense. Then, given a crime series that we wish to analyze, we require a representative list of historically committed crimes of the same type. Clearly, this process requires the presence of a skilled analyst to determine which historical crimes are of the same type as the series under consideration. As an example, when looking at a series of street robberies, it is likely that the geographic distribution of street robbery rates are different for daylight robberies as opposed to late night robberies. This approach then inserts the crime analyst and their relevant real world experience directly into the modeling process and the algorithm. This approach also lets us handle different crime types within the same mathematical framework, as different crime types will have different historical patterns. Of course, the local analyst will need to have access to the necessary data.

Once we have the historical data, we need to estimate the target density function $G(\mathbf{x})$. Perhaps the simplest method is kernel density parameter estimation. To use this method, let us suppose that we have a representative list of the crimes of a given type and that they have occurred at the points $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N$. Choose a kernel density function $K(\mathbf{y} \mid \lambda)$ with bandwidth λ. There are a number of reasonable choices for the kernel density function K, including normal or truncated quartic. It turns out that the mathematical properties of this method do not depend strongly on the mathematical form of the kernel, but that they do depend on the bandwidth of the kernel [46]. The bandwidth λ of a given kernel is related to the width of the function; as an example the bandwidth of a normal curve is the variance of that normal; when using a truncated quartic, the bandwidth is the size of the interval for which the quartic is nonzero.

We then construct the local target attractiveness function by calculating

$$G(\mathbf{x}) = \sum_{i=1}^{N} K(\mathbf{x} - \mathbf{c}_i \mid \lambda) \tag{3.10}$$

for a reasonable choice of bandwidth λ, say the mean nearest neighbor distance between historical crime sites. This is essentially the same as one of the methods used to generate crime hot spots described in [15]. Similar techniques are used in mathematical biology to estimate the home range of an animal species based on observations on individuals in the environment [61].

Once we have selected a model for offender behavior $P(\mathbf{x} \mid \mathbf{z})$, we also need to make a choice for the prior probability density for offender anchor points $H(\mathbf{z})$ and the prior distribution of the offender's average offense distance $\pi(\alpha)$ before we can use our fundamental result (3.5).

The prior probability density for offender anchor points $H(\mathbf{z})$ represents our knowledge of the offender's anchor point before we use any of the information from the crime series itself. There are a number of mathematically and criminologically reasonable choices for this prior distribution. The simplest choice would be to assume all potential anchor points are equally likely; we can do this by simply choosing $H(\mathbf{z}) = 1$.

Before we examine more sophisticated priors, we return to the question of what is an anchor point. If we assume that the anchor point is the offender's home, or more generally that the distribution of anchor points follows local population density, then we can use demographic data to generate an estimate for the prior. In this case, we we can choose $H(\mathbf{z})$ so that it is proportional to local population density. U.S. Census data gives population counts at the block level together with the land area of the block. We can use this data and kernel density parameter estimation technique to generate $H(\mathbf{z})$ by calculating

$$H(\mathbf{z}) = \sum_{i=1}^{N_{\text{blocks}}} = p_i K(\mathbf{z} - \mathbf{q}_i \mid \sqrt{A_i})$$

where each block has population p_i, center $\mathbf{q}_i$ and for each block we have chosen a different bandwidth equal to the side length of a square with the same area A_i as the block. We mention that U.S. Census population data at the block level is also available sorted by age, sex, and race/ethnic group. Thus, if demographic information is available about the offender, then this information can be incorporated when the prior distribution of anchor points $H(\mathbf{z})$ is calculated.

Our framework does not require that the anchor point be the offender's home or that the distribution of anchor points follows local population density. Another reasonable approach to calculating $H(\mathbf{z})$ would be to begin with the anchor points of previous offenders who have committed similar crimes. Then the same kernel density process used to generate $G(\mathbf{x})$ in (3.10) can be used to generate $H(\mathbf{z})$. These historical anchor points can be determined on an offender-by-offender basis; they can be homes, places of work or even the offender's favorite bar. Recall however that one of our assumptions is that each offender has a unique stable anchor point.

The last element needed to implement our fundamental mathematical result is some estimate of the prior distribution $\pi(\alpha)$ of the average distance to crime. Estimates of these types of distance to crime distributions are commonly performed by choosing a common statistical function and using best fit estimates; see [29, Chapter 10] for an example of the process. However, our framework does not require a particular parametrized form for the prior distribution $\pi(\alpha)$; we can instead directly use appropriate empirical data in the construction. Further, there is no requirement that the same choice of $\pi(\alpha)$ needs to be made for different crime types. Again, an analyst can choose which historical data to use when generating $\pi(\alpha)$.

Prototype software that implements this framework has been developed and released to for public use and evaluation. Empirical tests are being arranged to evaluate the accuracy and precision of this approach.

3.1.4 Future Offense Prediction

The focus of our attention so far has been on the traditional geographic profiling problem of estimating the location of the offender's anchor point by using the geographic information contained in the crime series. However, this is not the only question of interest to law enforcement. Indeed, another question of nearly equal importance is to estimate the location of the serial offender's next target.

This question can be posed in the following mathematical form. Given a series of crimes at the locations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ committed by a single serial offender, estimate the probability density $P(\mathbf{x}_{\text{next}} \,|\, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, that $\mathbf{x}_{\text{next}}$ will be the location of the next offense. The Bayesian approach to this problem is to calculate the posterior predictive distribution

$$P(\mathbf{x}_{\text{next}} \,|\, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = \iiint P(\mathbf{x}_{\text{next}} \,|\, \mathbf{z}, \alpha) P(\mathbf{z}, \alpha \,|\, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)\, dz^{(1)}\, dz^{(2)}\, d\alpha.$$

Once again, we can use (3.3) and (3.4) to simplify, and so obtain the expression

$$P(\mathbf{x}_{\text{next}} \,|\, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \propto \iiint P(\mathbf{x}_{\text{next}} \,|\, \mathbf{z}, \alpha) P(\mathbf{x}_1 \,|\, \mathbf{z}, \alpha) P(\mathbf{x}_2 \,|\, \mathbf{z}, \alpha) \ldots \\ P(\mathbf{x}_n \,|\, \mathbf{z}, \alpha) H(\mathbf{z}) \pi(\alpha)\, dz^{(1)}\, dz^{(2)}\, d\alpha$$

This approach makes the same independence assumptions about offender behavior as our fundamental result (3.5).

3.2 Implementing the mathematical model

Our fundamental result was (3.5) which said that if we assume that the an offender has anchor point $\mathbf{z}$ and average offense distance α commits a crime at the location $\mathbf{x}$ according to the probability density $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$, then the probability density that the offender has anchor point $\mathbf{z}$ given that they have committed crimes at $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ satisfies

$$P(\mathbf{z} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n) \propto \int_0^\infty P(\mathbf{x}_1 \,|\, \mathbf{z}, \alpha) \cdots P(\mathbf{x}_n \,|\, \mathbf{z}, \alpha) H(\mathbf{z}) \pi(\alpha)\, d\alpha.$$

Here $H(\mathbf{z})$ is a prior estimate for the distribution of offender anchor points, and $\pi(\alpha)$ is a prior estimate for the distribution of average offense distances.

We then considered how the function $P(\mathbf{x} \,|\, \mathbf{z})$ might be constructed and presented the option (3.8)

$$P(\mathbf{x} \,|\, \mathbf{z}, \alpha) = D(d(\mathbf{x}, \mathbf{z}), \alpha) G(\mathbf{x}) N(\mathbf{z}, \alpha).$$

Here D is the distance decay function, G accounts for local geographic features that affect target selection, and N is a required normalization term.

Before we could turn the mathematical framework described in the previous section into a functioning tool, we need to

- Make choices for the distance decay function
- Estimate the prior distribution of average offense distance
- Determine how to represent the geographic data in a form amenable to computation, and
- Determine how to evaluate the normalization function $N(\mathbf{z}, \alpha)$.

In this section we discuss the mathematical details involved in these issues.

3.2.1 Measuring distance

The first step in our implementation is to select a common framework for handling place and distance. The mathematical model described so far is not dependent on any particular choice for a distance metric or on how places are labeled. However, as we move towards questions of implementation, we must select a way to identify real geographic locations with their mathematical representations and to select a method to calculate the distance between these places.

To locate places, we will solely use the latitude and longitude of the location, and the software tool will use exclusively latitude and longitude in decimal degrees. As our reference we will use the same reference as the U.S. Census Bureau Summary File 1 data (NAD-83).

Now suppose that $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$ are a pair of points, but now x_1 and y_1 are the longitudes and x_2 and y_2 are the latitudes of our points. To calculate the distance between these points, we will use the corresponding great-circle distance; in the software tool this will be calculated using the haversine formula

$$d(\mathbf{x}, \mathbf{y}) = 2\sin^{-1}\sqrt{\sin^2\left(\frac{x_2 - y_2}{2}\right) + \cos x_2 \cos y_2 \sin^2\left(\frac{x_1 - y_1}{2}\right)}.$$

Notice that the result of d is an angle; this is the central angle between the two rays from the center of the earth to the two points on the surface. It can be converted to approximate distance in miles by ensuring that d is measured in radians and then multiplying by the radius of the earth in miles.

The advantage of this approach is that this gives a good approximation to the straight-line linear distance and it is something that can be quickly and easily calculated from the known latitude and longitude of the point. On the other hand, this is still just an approximation to the linear distance; it does not account for a number features including differences in relative elevation and the fact that the earth is better approximated by a ellipsoid rather than a sphere. As discussed earlier, there are a number of reasonable choices of distance metric that could be made and the mathematical framework presented here can use any of them. On the other hand, a particular choice needs to be made for the software prototype, and this choice will be used in both the software as well as the mathematical examples that follow.

3.2.2 The distance decay

There are a number of reasonable choices that can be made for the distance decay function D. In our implementation, we assume that the distance decay function D follows a bivariate normal distribution. In the (current) absence of a consensus for one form of the distance decay function over another, we selected the bivariate normal because it is commonly used to describe two-dimensional diffusion processes that result from random walks.

Thus, we suppose that the offender's distance decay effect follows a bivariate normal distribution with mean $\mathbf{z}$ (the offender's anchor point) and covariance matrix $\sigma^2 I$, so that

$$D(\mathbf{x} \mid \mathbf{z}, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|\mathbf{x} - \mathbf{z}|^2}{2\sigma^2}\right).$$

The offender's travel distance is $r = |\mathbf{x} - \mathbf{z}|$; it has the density function

$$f(r \mid \sigma) = 2\pi r D(\mathbf{x} \mid \mathbf{z}, \sigma) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right).$$

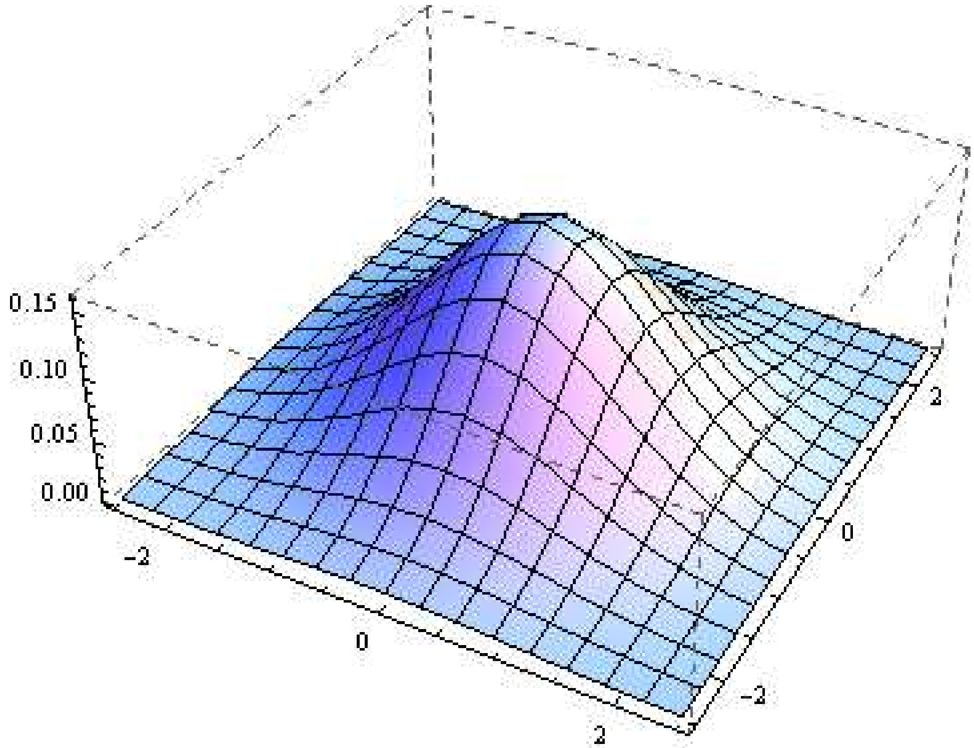


Figure 3.1: A graph of the two-dimensional distance decay distribution $D(\mathbf{x}\,|\,\mathbf{z},\alpha)$ where $\mathbf{z}$ is located at the origin and $\alpha = 1$.

Rather than use σ as the parameter, we would like to use α, the mean distance traveled by the offender. Now the average offense distance is given by

$$Er = \int_0^\infty r f(r\,|\,\sigma)\,dr = \int_0^\infty \frac{r^2}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)\,dr.$$

Integrating by parts, we find that

$$Er = \int_0^\infty \exp\left(-\frac{r^2}{2\sigma^2}\right)\,dr = \sqrt{\frac{\pi}{2}}\sigma$$

so we conclude $\alpha = \sqrt{\frac{\pi}{2}}\sigma$. Thus, the two-dimensional distribution of offense distances as a function of the average offense distance α is

$$D(\mathbf{x}\,|\,\mathbf{z},\alpha) = \frac{1}{4\alpha^2}\exp\left(-\frac{\pi}{4\alpha^2}|\mathbf{x}-\mathbf{z}|^2\right). \tag{3.11}$$

A graph of this function for $\alpha = 1$ is given in Figure 3.1; the x and y coordinates represent geographic points, while the corresponding height is the probability density that that location is selected.

We can also examine the probability density for the distances traveled; as a function of α it has the density

$$f(r\,|\,\alpha) = 2\pi r D(\mathbf{x}\,|\,\mathbf{z},\alpha) = \frac{\pi r}{2\alpha^2}\exp\left(-\frac{\pi r^2}{4\alpha^2}\right) \tag{3.12}$$

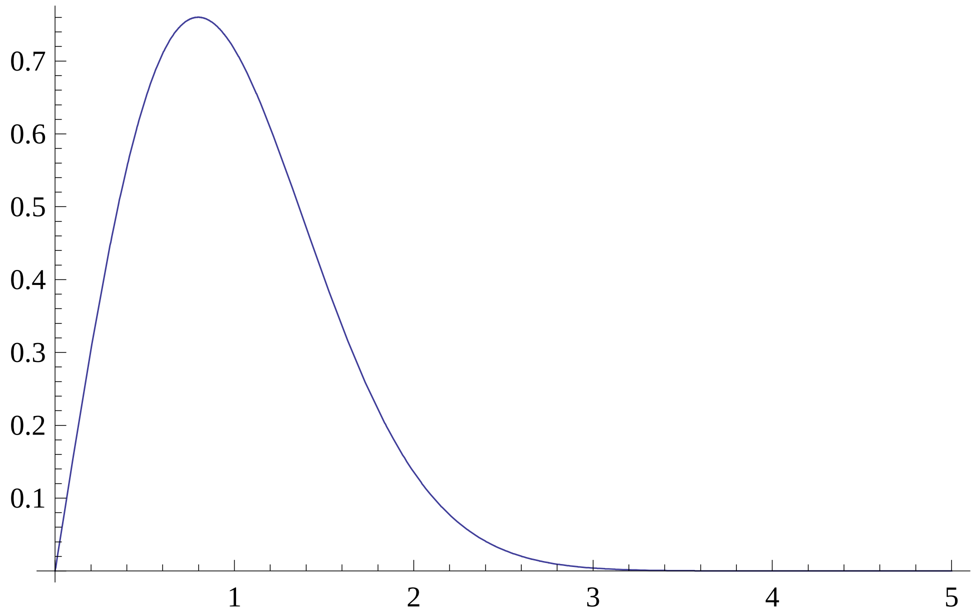


Figure 3.2: A graph of the one-dimensional distance decay distribution $f(r \,|\, \alpha)$ for $\alpha = 1$.

where again $r = |\mathbf{x} - \mathbf{z}|$ is the travel distance. This is a Rayleigh distribution in r, and it is graphed in Figure 3.2

It is important to be aware of the differences between the two-dimensional distribution given by $D(\mathbf{x} \,|\, \mathbf{z}, \alpha)$ and the corresponding one-dimensional distribution of distances $f(r \,|\, \alpha)$. First, though the two-dimensional distribution is (two-dimensional) normal, it is not the case that the distance distribution is normal; in fact the distances follow a Rayleigh distribution. More generally, consider the two dimensional distance decay distribution $D(\mathbf{x} \,|\, \mathbf{z}, \alpha)$, and for simplicity assume that we can write it in the form $D(r \,|\, \alpha)$ where $r = |\mathbf{x} - \mathbf{z}|$ is the distance between the offense site $\mathbf{x}$ and the anchor point $\mathbf{z}$. Then the distribution of the distances f is given by the one-dimensional distribution $f(r) = 2\pi r D(r \,|\, \alpha)$. Indeed, $f(r)$ is simply the the probability density that one of the sites at a distance r would be chosen namely $D(r \,|\, \alpha)$, multiplied by the number of sites at a distance r namely $2\pi r$, which is the circumference of a circle of radius r.

One place where these differences are particularly important is in the discussion of offender buffer zones. A buffer zone is an area near the offender's anchor point where they are less likely to offend due to *e.g.* fears that they would be recognized. Clearly the two-dimensional normal form for D shown in Figure 3.1 does not indicate the presence of a buffer zone; in fact it shows that the offender would prefer to offend at locations closer to their anchor point than locations farther away. Examining the distribution of distances in Figure 3.2 however, we see that the offender is less likely to offend in locations very near to their anchor point. This is not being caused by the existence of a buffer zone. Rather, though the offender is more likely to choose a location closer to their anchor point than farther away (Figure 3.1), the size of the region at a distance r from the anchor point decreases as r does.

More generally, we notice that the one-dimensional distribution $f(r \,|\, \alpha)$ satisfies $f(r \,|\, \alpha) \to 0$ as $r \to 0$ regardless of the form of D provided D remains finite.

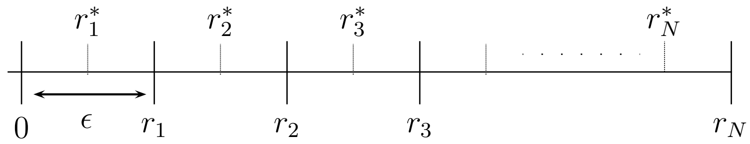


Figure 3.3: Creating distance bins

3.2.3 Estimating the prior distribution of average offense distances

We need to construct an estimate for $\pi(\alpha)$, which is the prior estimate for the distribution of average offense distances. It is fundamental to note that this is the distribution of the average offense distances *across offenders*. In particular, though this is related to the distribution of offense distances across offenses (obtained for example by examining crime statistics) the distribution here is across people. We explicitly allow for the possibility that this prior estimate may vary by region and offense type; however this prior is estimated before we include information from the geographic locations of the offense sites.

To perform this estimation, let us first assume that we know that the distribution of distances from home to offense sites across known offenses is given by the function $A(r)$. Practically, we are unlikely to know the exact form of the distribution $A(r)$, but we can estimate it from crime statistics. Indeed, let us suppose that we have a sample of S solved crimes, and that the distances from each offense site to the corresponding offender anchor point is given by $\{\rho_1, \rho_2, \dots, \rho_S\}$.

Choose a discretization size $\epsilon > 0$, then define $r_j = j\epsilon$ and $r_j^* = (j - \frac{1}{2})\epsilon$, to subdivide the real axis into a sequence of bins $[r_{j-1}, r_j)$ each with center r_j^*; see Figure 3.3. To estimate the value of A in the center of bin $[r_{j-1}, r_j)$, namely $A(r_j^*)$, we let a_j be the number of distances ρ_s in this bin,

$$a_j = \#\{s \,|\, r_{j-1} \le \rho_s < r_j\}. \tag{3.13}$$

Then we have the relationship

$$A(r_k^*)\epsilon \approx \frac{a_j}{S} \tag{3.14}$$

where both sides of approximate the probability that ρ lies in the bin $[r_{j-1}, r_j)$. Indeed, because $\{\rho_1, \rho_2, \dots, \rho_S\}$ is a sample of size S

$$\text{Prob}[r_{j-1} < \rho < r_j] \approx \frac{a_j}{S}$$

while

$$\text{Prob}[r_{j-1} < \rho < r_j] = \int_{r_{j-1}}^{r_j} A(\rho)\,d\rho = A(r_j^*)\epsilon + O(\epsilon^2)$$

from the midpoint rule.

Returning to our estimate of $\pi(\alpha)$, we begin with the fundamental relationship

$$A(r) = \int_0^\infty f(r \,|\, \alpha)\pi(\alpha)\,d\alpha \tag{3.15}$$

which states that the number of offenses at the distance r can be found by by multiplying the probability density that an offender with average distance to offend α chooses the offense distance

r by the probability that an offender actually has the offense distance α, and then integrating over all possible values of α. In particular, this accounts for two sources of variation- the variation in offense distances selected by one offender, and the variation in average offense distances across multiple offenders.

Consider the simple case where all offenders are considered to be identical; in this case there is a particular average offense distance α^* shared by all offenders, and the prior distribution $\pi(\alpha)$ is simply the Dirac distribution $\delta(\alpha - \alpha^*)$. Because $\delta(\alpha - \alpha^*) = 0$ for $\alpha \neq \alpha^*$ and because $\int_0^\infty \delta(\alpha - \alpha^*)\, d\alpha = 1$, we see that (3.15) reduces to

$$A(r) = f(r \,|\, \alpha^*)$$

so that the behavior of any single offender can be estimated from the aggregated data $A(r)$ by simply choosing the value of α^* that best fits the data.

Our approach is more general, and does not assume that all offender's exhibit the same distance decay behavior. As a consequence however, we still need to solve equation (3.15) for the unknown prior $\pi(\alpha)$. To do so, we proceed by collocation.

In particular, we know that offenders do not travel infinite distances to offend, so we choose a number N so large that $A(r) \approx 0$ for $r > \epsilon N$; then we want to choose $\pi(\alpha)$ so that

$$A(r_j^*) = \int_0^\infty f(r_j^* \,|\, \alpha)\pi(\alpha)\, d\alpha$$

for $j = 1, 2, \ldots, N$. The assumption $A(r) \approx 0$ for $r > \epsilon N$, also lets us conclude that $\pi(\alpha) \approx 0$ for $\alpha > \epsilon N$. Indeed, $A(r)$ is the measured number of crimes that occur at the distance r from home, while $\pi(r)$ is the density of offender's whose average offense distance is r.

To evaluate the integrals, define $\alpha_k = k\epsilon$, $\alpha_k^* = (k - \frac{1}{2})\epsilon$, and apply the midpoint rule to the integral

$$\begin{aligned}\int_0^\infty f(r \,|\, \alpha)\pi(\alpha)\, d\alpha &\approx \int_0^{\epsilon N} f(r \,|\, \alpha)\pi(\alpha)\, d\alpha \\ &\approx \sum_{k=1}^{N} f(r \,|\, \alpha_k^*)\pi(\alpha_k^*)\epsilon + O(\epsilon^2).\end{aligned}$$

Thus, for each $j, k \in \{1, 2, \ldots, N\}$, we have

$$A(r_j^*) \approx \epsilon \sum_{k=1}^{N} f(r_j^* \,|\, \alpha_k^*)\pi(\alpha_k^*)$$

We can then use (3.14) where a_j is defined by (3.13) to conclude that

$$a_j = S\epsilon^2 \sum_{k=1}^{N} f(r_j^* \,|\, \alpha_k^*)\pi(\alpha_k^*). \tag{3.16}$$

Now this equation holds regardless of the choice of D and f, but the expression

$$f(r \,|\, \alpha) = \frac{\pi r}{2\alpha^2} \exp\left(-\frac{\pi r^2}{4\alpha^2}\right)$$

from (3.12) lets us simplify this further giving us

$$
\begin{aligned}
a_j &= S\epsilon^2 \sum_{k=1}^{N} \frac{\pi r_j^*}{2(\alpha_k^*)^2} \exp\left(-\frac{\pi}{4}\frac{(r_j^*)^2}{(\alpha_k^*)^2}\right)\pi(\alpha_k^*) \\
&= S\epsilon^2 \sum_{k=1}^{N} \frac{\pi}{2}\frac{(j-\frac{1}{2})\epsilon}{(k-\frac{1}{2})^2\epsilon^2} \exp\left(-\frac{\pi}{4}\frac{(j-\frac{1}{2})^2\epsilon^2}{(k-\frac{1}{2})^2\epsilon^2}\right)\pi(\alpha_k^*) \\
&= \frac{\pi S\epsilon}{2} \sum_{k=1}^{N} \frac{(j-\frac{1}{2})}{(k-\frac{1}{2})^2} \exp\left(-\frac{\pi}{4}\frac{(j-\frac{1}{2})^2}{(k-\frac{1}{2})^2}\right)\pi(\alpha_k^*)
\end{aligned}
$$

Thus, if we define the matrix

$$
G = G_{jk} = \frac{\pi S\epsilon}{2}\frac{(j-\frac{1}{2})}{(k-\frac{1}{2})^2}\exp\left(-\frac{\pi}{4}\frac{(j-\frac{1}{2})^2}{(k-\frac{1}{2})^2}\right)
$$

and the vectors

$$
\mathbf{a} = (a_1, a_2, \dots, a_N) \tag{3.17}
$$

$$
\pi = (\pi(\alpha_1^*), \pi(\alpha_2^*), \dots, \pi(\alpha_N^*)) \tag{3.18}
$$

then we obtain the discrete linear system

$$
\mathbf{a} = G\pi. \tag{3.19}
$$

Tikhonov regularization

The direct solution of (3.19) for the unknown prior π is not practical. Indeed, because this results from the collocation of a Fredholm integral equation of the first kind (3.15), we would expect that the resulting linear system would be ill-posed- *c.f.* [19, §1.2] or [54, §1.1]. Using a computer algebra system or otherwise, one can verify that $\det G \approx 0$; moreover by examining the singular values of G, we obtain the results in Figure 3.4, which are characteristic of an ill-posed problem

Attempting to solve this linear system using traditional techniques like the pseudo-inverse are doomed to failure; see Figure 3.5 which illustrates what occurs when the attempt is made where $\mathbf{a}$ is calculated from a set of Baltimore County residential burglaries. Notice the the wide oscillations through positive and negative values over twelve orders of magnitude.

To see the fundamental issue, expand G via its singular value decomposition

$$
G = USV^\top
$$

where U and V are orthogonal matrices and $S = \operatorname{diag}(s_1, s_2, \dots, s_N)$ is the matrix of singular values [20, §5.4]. Then the pseudo-inverse solution $\pi^\dagger$ to $G\pi = \mathbf{a}$ is

$$
\pi^\dagger = VS^{-1}U^\top\mathbf{a}
$$

so that $\pi^\dagger$ satisfies

$$
\pi^\dagger = \sum_{i=1}^{N} \frac{\mathbf{u}_i^\top\mathbf{a}}{s_i}\mathbf{v}_i \tag{3.20}
$$

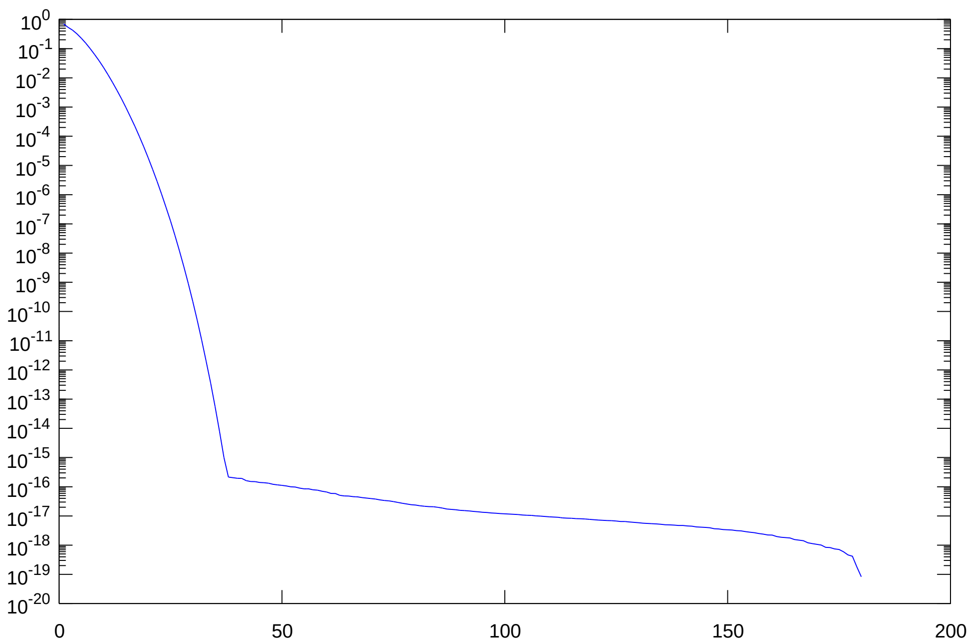


Figure 3.4: Singular values for the matrix G, when $N = 180$ and $\epsilon = 0.002$.

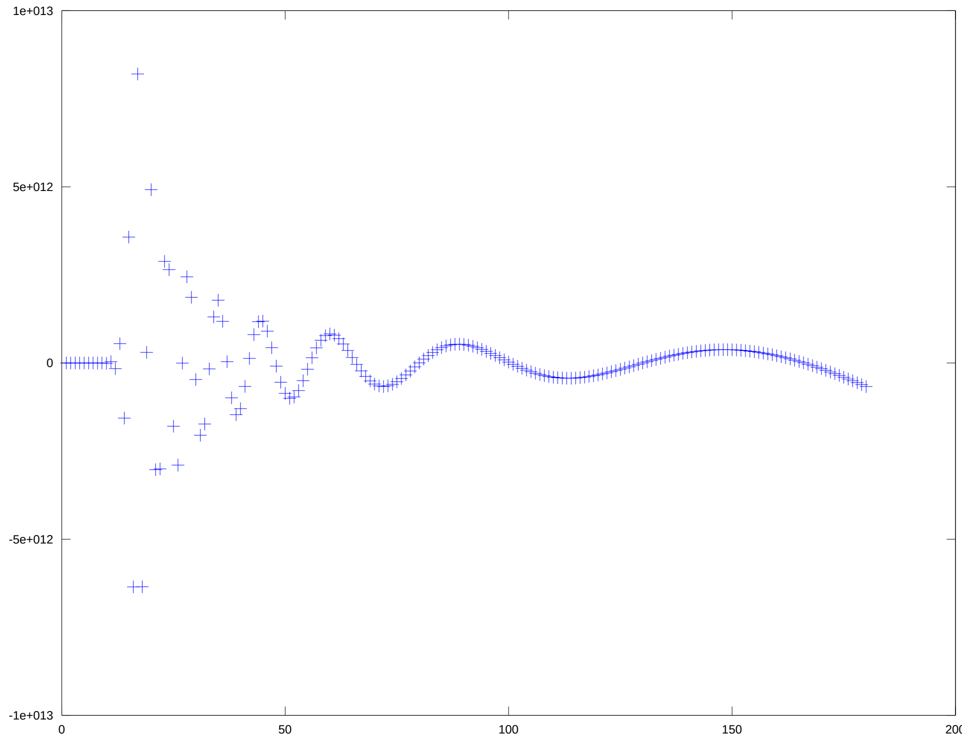


Figure 3.5: Attempt to solve (3.19) using the pseudo-inverse; the known data $\mathbf{a}$ is calculated from residential burglaries in Baltimore County, while $N = 180$ and $\epsilon = 0.002$.

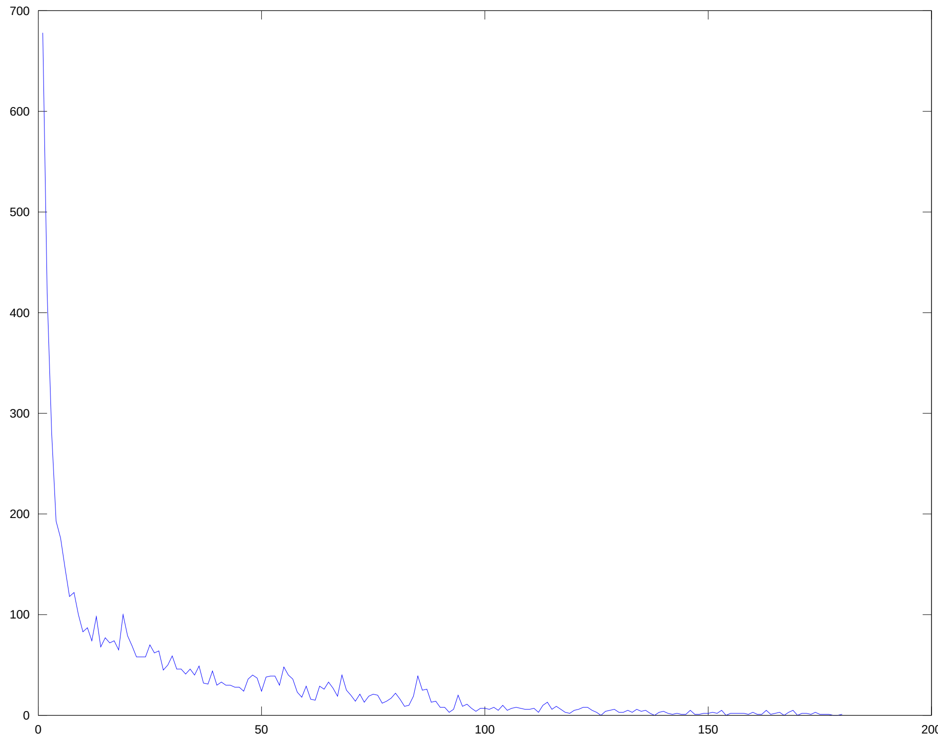


Figure 3.6: The vector $\mathbf{a}$ calculated from residential burglaries in Baltimore County, while $N = 180$ and $\epsilon = 0.002$.

where we represent U and V in terms of their columns as

$$U = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N), \qquad V = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N).$$

From this and Figure 3.4 the problem is apparent- when the singular values s_i are very small, their presence in the denominator results dramatically amplifies the component of the solution in the $\mathbf{v}_i$ direction, including any errors in those components. Indeed, if we examine the actual vector $\mathbf{a}$ for the set of Baltimore County residential burglaries and graph the result, we obtain Figure 3.6. We can clearly see the oscillations in the data and expect (correctly) that these are not significant, but are rather the result of various random fluctuations. Unfortunately, the these insignificant variations lead to nonzero values of $\mathbf{u}_i^\top \mathbf{a}$, some of which are then dramatically amplified by small singular values appearing in the denominator of (3.20). In some sense, the result in Figure 3.5 can be thought of as an attempt to match not only the information from Figure 3.6, but also the random noise.

One approach to this problem is to discard all singular values below a fixed threshold, but examination of the graph of singular values Figure 3.4 shows that this problem contains no natural cutoff threshold; rather the singular values decrease smoothly to $O(10^{-15})$; and then continue to decay more slowly.

Instead, we proceed via Tikhonov regularization. In particular we consider the regularized solutions

$$\pi_{\text{reg}}(\lambda) = \sum_{i=1}^{N} \frac{s_i^2}{s_i^2 + \lambda^2} \frac{\mathbf{u}_i^\top \mathbf{a}}{s_i} \mathbf{v}_i \tag{3.21}$$

instead of the pseudo-inverse solution (3.20). Here the factors

$$\phi(\lambda) = \frac{s_i^2}{s_i^2 + \lambda^2},$$

called filter factors, are chosen so that if $\phi \approx 1$ for $s_i \gg \lambda$, while $\phi \approx 0$ for $s_i \ll \lambda$; thus we are keeping the larger singular values, but discounting the smaller singular values, using λ as a cutoff. This process yields a regularized solution for every choice of parameter λ; we now need to select a method to choose it.

To choose λ, we first note that the pseudo-inverse solution $\pi^\dagger$ is the vector π that minimizes the functional

$$L(\pi) = \|G\pi - \mathbf{a}\|^2.$$

In the case where G has full rank or $\mathbf{a}$ lies in the range of G, this method returns a solution of $G\pi = \mathbf{a}$, while otherwise returning the value of π so that $G\pi$ is as close as possible to $\mathbf{a}$.

The Tikhonov regularized solution has a similar interpretation; the Tikhonov solution $\pi_{\text{reg}}(\lambda)$ with regularization parameter λ is the vector π that minimizes the functional

$$L_\lambda(\pi) = \|G\pi - \mathbf{a}\|^2 + \lambda\|\pi\|^2. \tag{3.22}$$

As a consequence, we see that the Tikhonov regularized solution is chosen to balance out the error obtained by fitting π to the data (the term $\|G\pi - \mathbf{a}\|^2$) with an estimate of the size of π (the term $\lambda\|\pi\|^2$).

If one graphs the value of $\log\|G\pi - \mathbf{a}\|$ versus $\log\|\pi\|$ one obtains a graph that has the general shape of an ‘L’; we have plotted this curve in Figure 3.7 where $\mathbf{a}$ is chosen from residential burglaries in Baltimore County.

The underlying cause of the L shape is that when λ is small, we are essentially obtaining the pseudo-inverse solution, and we have seen already that this results in a function with wild oscillations through many orders of magnitude, making $\lambda\|\pi\|^2$ large. In this case, we say that the problem is undersmoothed. On the other hand, if λ is large, the functional $L_\lambda(\pi)$ penalizes nonzero values of π, and thus π is pushed towards zero; this results in choices of π that do not fit the equation well, and so $\|G\pi - \mathbf{a}\|^2$ becomes large. In this case, we say that the problem is oversmoothed.

One method used to choose λ then is to find the point on the curve where the curvature is greatest; i.e. the vertex of the ‘L’. This method is called the L-curve method, and it is the method that is used in our program to determine the optimal value of the regularization parameter λ. This method is described in detail in [19, §4.6] and [54, §7.4].

To explain how we have implemented the L-curve method in the code, we begin by calculating the abscissa and ordinate of the L-curve, namely $\log\|G\pi - \mathbf{a}\|$ and $\log\|\pi\|$. From the expression

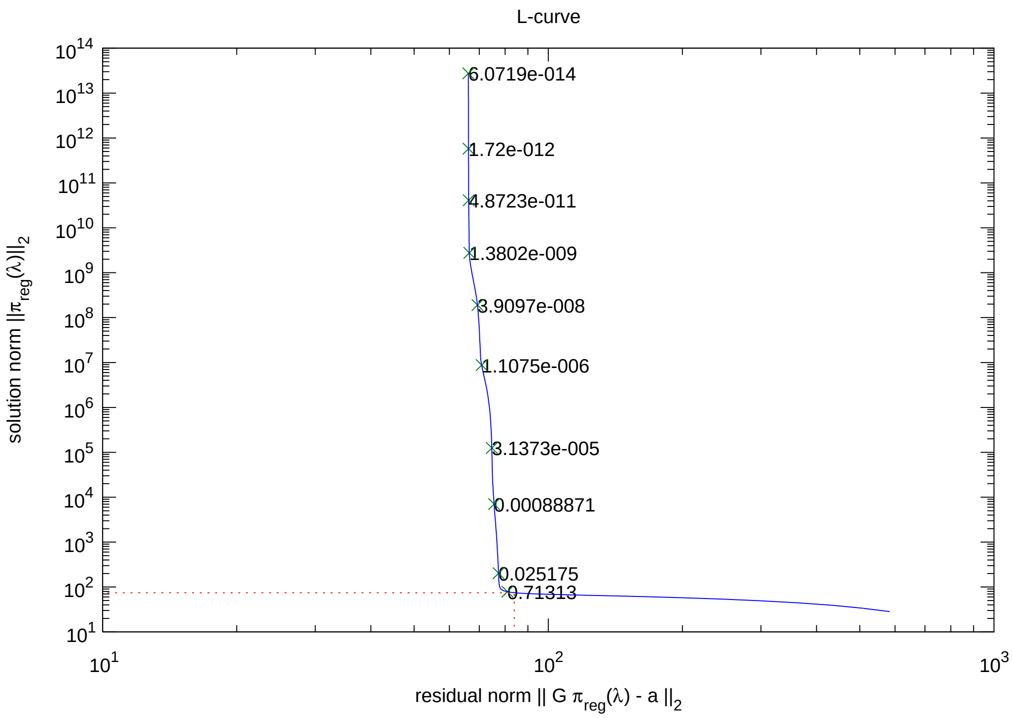


Figure 3.7: The L-curve, plotted using data for Baltimore County residential burglaries. Selected points on the L-curve are labeled with the corresponding value of λ. Here $N = 180$ and $\epsilon = 0.002$.

of the regularized solution (3.21) we find that

$$\begin{aligned} A\pi_{\text{reg}}(\lambda) &= A\sum_{i=1}^{N} \frac{s_i^2}{s_i^2+\lambda^2}\frac{\mathbf{u}_i^\top \mathbf{a}}{s_i}\mathbf{v}_i \\ &= \sum_{i=1}^{N} \frac{s_i^2}{s_i^2+\lambda^2}\frac{\mathbf{u}_i^\top \mathbf{a}}{s_i}A\mathbf{v}_i \\ &= \sum_{i=1}^{N} \frac{s_i^2}{s_i^2+\lambda^2}\frac{\mathbf{u}_i^\top \mathbf{a}}{s_i}s_i\mathbf{u}_i \\ &= \sum_{i=1}^{N} \frac{s_i^2}{s_i^2+\lambda^2}\mathbf{u}_i^\top \mathbf{a}\mathbf{u}_i \end{aligned}$$

where we have used the properties of the singular value decomposition $A\mathbf{v}_i = s_i\mathbf{u}_i$; [20, §5.4, Theorem 1]. On the other hand, because U is orthogonal, we can write

$$\mathbf{a} = \sum_{i=1}^{N}(\mathbf{u}_i^\top \mathbf{a})\mathbf{u}_i.$$

Combining these last two equations then, we see that

$$G\pi_{\text{reg}}(\lambda) - \mathbf{a} = -\sum_{i=1}^{N}(1-\phi_i(\lambda))\mathbf{u}_i^\top \mathbf{a}\mathbf{u}_i$$

for the filter factors $\phi(\lambda) = \frac{s_i^2}{s_i^2+\lambda^2}$.

From this, we find

$$\|G\pi_{\text{reg}}(\lambda) - \mathbf{a}\|^2 = \sum_{i=1}^{N}(1-\phi_i(\lambda))^2(\mathbf{u}_i^\top \mathbf{a})^2$$

and

$$\|\pi_{\text{reg}}(\lambda)\|^2 = \sum_{i=1}^{N}\frac{\phi_i^2(\lambda)}{s_i^2}(\mathbf{u}_i^\top \mathbf{a})^2.$$

To simplify the calculations as we proceed, we introduce some new notation. We begin by letting $\alpha = \lambda^2$, and defining

$$\begin{aligned} R(\alpha) &= \|G\pi_{\text{reg}}(\lambda) - \mathbf{a}\|^2 \qquad & T(\alpha) &= \|\pi_{\text{reg}}(\lambda)\|^2 \\ X(\alpha) &= \log R(\alpha) & Y(\alpha) &= \log T(\alpha) \end{aligned}$$

Our goal is to find the point on the curve $(X(\alpha), Y(\alpha))_{\alpha>0}$ where the curvature is largest. Now

$$\begin{aligned} R(\alpha) &= \|G\pi_{\text{reg}}(\lambda) - \mathbf{a}\|^2 = \sum_{i=1}^{N}(1-\phi_i(\lambda))^2(\mathbf{u}_i^\top \mathbf{a})^2 \\ T(\alpha) &= \|\pi_{\text{reg}}(\lambda)\|^2 = \sum_{i=1}^{N}\frac{\phi_i^2(\lambda)}{s_i^2}(\mathbf{u}_i^\top \mathbf{a})^2 \end{aligned}$$

so differentiating, we see that

$$R'(\alpha) = \sum_{i=1}^{N} -2(1-\phi_i(\alpha))\phi_i'(\alpha)(\mathbf{u}_i^\top \mathbf{a})^2$$
$$T'(\alpha) = \sum_{i=1}^{N} 2\frac{\phi_i^2(\alpha)}{s_i^2}\phi_i'(\alpha)(\mathbf{u}_i^\top \mathbf{a})^2.$$

Now because

$$\begin{aligned} 1-\phi_i(\alpha) &= 1 - \frac{s_i^2}{s_i^2+\alpha} = \frac{\alpha}{s_i^2+\alpha} = \frac{\alpha}{s_i^2}\frac{s_i^2}{s_i^2+\alpha} \\ &= \frac{\alpha}{s_i^2}\phi_i(\alpha) \end{aligned}$$

we see that we have the relationship

$$R'(\alpha) = -\alpha T'(\alpha). \tag{3.23}$$

With these preliminaries concluded, we start by examining the curvature κ of our curve

$$\kappa(\alpha) = \frac{X''Y' - X'Y''}{[(X')^2 + (Y')^2]^{3/2}}.$$

Then using (3.23), we see that

$$\begin{aligned} X' &= \frac{R'}{R} = \frac{-\alpha T'}{R} & Y' &= \frac{T'}{T} \\ X'' &= -\frac{T'}{R} - \frac{\alpha T''}{R} + \frac{\alpha T' R'}{R^2} & Y'' &= \frac{T''T - (T')^2}{T^2} \\ &= -\frac{T'}{R} - \frac{\alpha T''}{R} - \frac{\alpha^2 (T')^2}{R^2} & & \end{aligned}$$

Substituting these into our expression for curvature, we obtain the expression

$$\kappa = -\frac{RT(\alpha R + \alpha^2 T) + (RT)^2/T'}{(R^2 + \alpha^2 T^2)^{3/2}}$$

which is the expression [54, 7.32]. This algorithm is implemented in our program, in the method `CTikhonov::LCurvature`.

Now that we know how to calculate the curvature on the L-curve, we need to determine the point on the L-curve where the curvature is at its maximum. Our approach to this problem is to use the Golden section method. This is an elementary method used to solve constrained optimization problems, and is described in detail in [2, §8.1]. Nominally, the problem of determining the value of λ is not a constrained optimization problem, as any the curvature can be calculated for any positive λ. However, we incorporate two restrictions. Since λ is being used as a cutoff for the singular values, there is no need to look for values of λ larger than the largest singular value s_1. On the other hand, extremely small values of λ will result in significant numerical error due to the

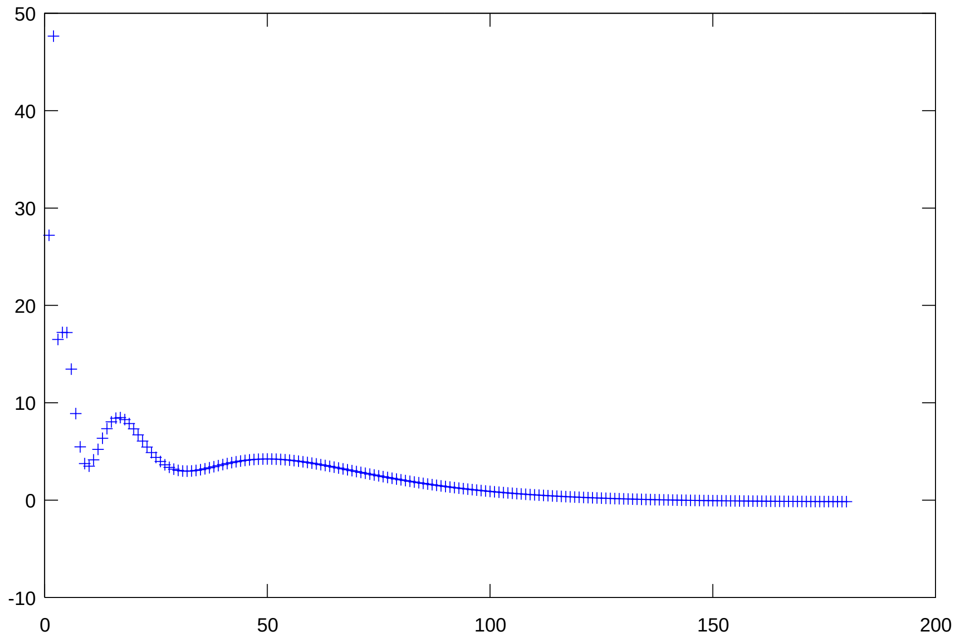


Figure 3.8: Tikhonov regularized solution estimate of π plotted using data from Baltimore County residential burglaries. Here $N = 180$ and $\epsilon = 0.002$.

limited precision of the computer. Thus, for the smallest allowable value of λ we use the larger of either the smallest singular value s_N or $16\delta s_1$, where δ is the minimum representable difference between two numbers in the computer.

Once a choice for λ has been made, we need to find the Tikhonov regularized solution which is the value of π that makes (3.22) as small as possible. Doing so for the Baltimore County residential burglary data, one obtains a graph like Figure 3.8 On its face, the graph appears reasonable; certainly much more reasonable than the graph of the non-regularized solution, Figure 3.5. Appearances however are misleading, as a closer look at the solution will show. Indeed, if we plot only the last 100 values of π, we obtain Figure 3.9. We immediately see the problem, as the graph clearly shows that there are components of π that are negative. Recalling our definition of the vector π from (3.18), we see that none of the components of π should be negative, as they are all simply the values of a probability distribution.

To proceed then, we minimize the functional $L_\lambda(\pi)$ not over all vectors π, but only over those vectors π all of whose components are nonnegative. To perform this process, we first rewrite the minimization of (3.22) in the slightly different form

$$L_\lambda(\pi) = \left\| \begin{pmatrix} G \\ \lambda I \end{pmatrix} \pi - \begin{pmatrix} \mathbf{a} \\ \mathbf{0} \end{pmatrix} \right\|^2 .$$

In this form, this is a standard last squares minimization problem with the linear inequality con-

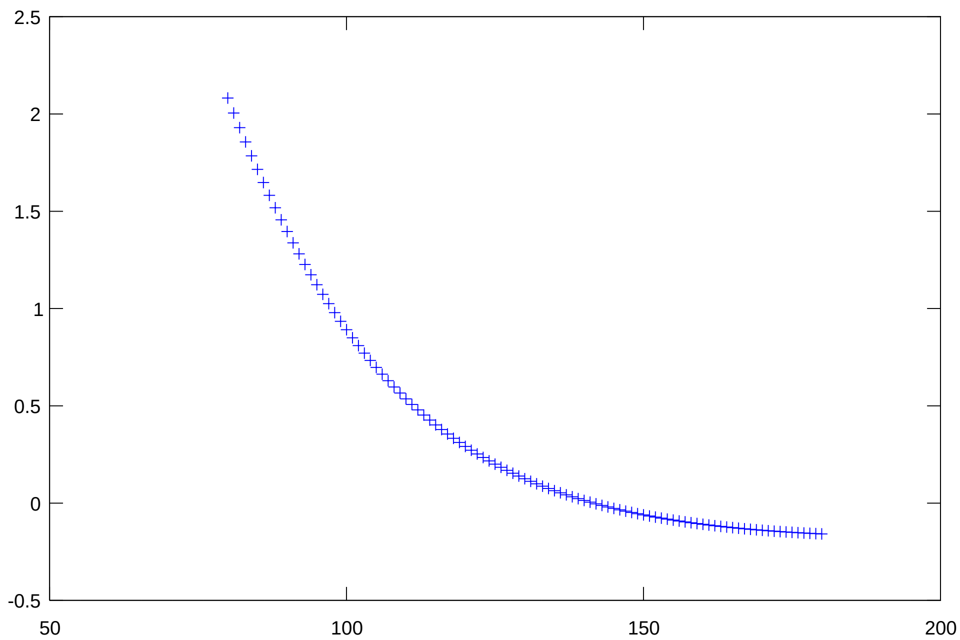


Figure 3.9: Entries 80-180 in the Tikhonov regularized solution estimate of π plotted using data from Baltimore County residential burglaries. Here $N = 180$ and $\epsilon = 0.002$.

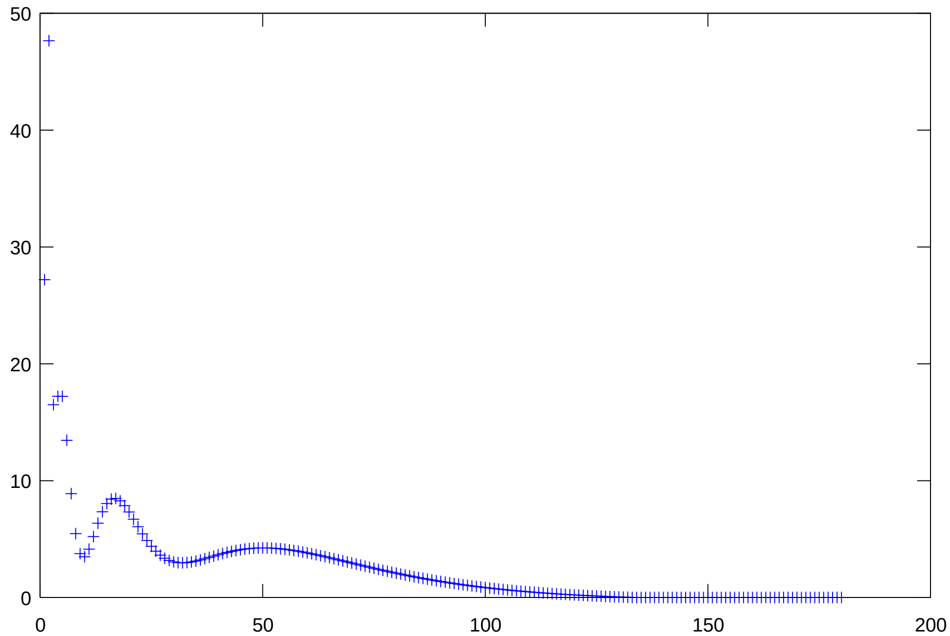


Figure 3.10: Tikhonov regularized solution with nonnegative entries to estimate π, plotted using data from Baltimore County residential burglaries. Here $N = 180$ and $\epsilon = 0.002$.

straint that all of the components of π are nonnegative. One standard algorithm for solving such minimization problems is provided in [26, Chp. 23]; this is the method that is implemented in the code.

When we solve this, we obtain the results in Figure 3.10 We shall discuss the significance of this graph (and others) from the point of view of what it tells us about offender behavior when we present our discussion of findings in Section 4.1.

Lastly, we note that the definition $\pi = (\pi(\alpha_1^*), \pi(\alpha_2^*), \dots, \pi(\alpha_N^*))$ from (3.18) tells us more than the fact that the components of π must be nonnegative; it also tells us that the components of π come from a probability distribution. In particular because π is a probability distribution, we know that

$$\int_0^\infty \pi(\alpha)\, d\alpha = 1.$$

Now if we once again apply the midpoint approximation for the integral, we see that

$$1 = \int_0^\infty \pi(\alpha)\, d\alpha \approx \epsilon \sum_{k=1}^{N} \pi(\alpha_k^*) = \epsilon \sum_{k=1}^{N} \pi_k. \tag{3.24}$$

Even though this condition on the components of π is not explicitly imposed by the method that we have implemented, we can check to see if it holds. Table 3.1 shows the resulting approximation to

Solution method	Estimate of $\int_0^\infty \pi(\alpha)\,d\alpha$
Pseudoinverse	-5896180
Tikhonov regularization	0.970644
Nonnegative least squares without regularization	1.074468
Nonnegative least squares with Tikhonov regularization	0.973170

Table 3.1: Estimates of $\int_0^\infty \pi(\alpha)\ d\alpha$ for various solution methods, calculated from Baltimore County residential burglary data with $N = 180$ and $\epsilon = 0.002$

$\int_0^\infty \pi(\alpha)\,d\alpha$ generated by (3.24) for various solution methods. We see that in each case we do have the appropriate rough approximation, save for the solution generated directly by the pseudoinverse, which we have already seen to be wildly inaccurate (Figure 3.5). However, no error analysis has yet been performed on these values. It is an interesting open question to see what better methods might be to perform the estimate of the function $\pi(\alpha)$.

3.2.4 The geographic triangulation

We now turn to the problem of representing the local geometry and geography. To do so, we start by realizing that there are three geographic regions that interest us

1. **The crime region**. We assume all of the known crimes are contained in a single geographic region, which we call the crime region. In particular, we assume that we have no information about whether or not the offender has committed additional series crimes outside the crime region. If an agency is investigating a crime series, and has no information about the series from other jurisdictions, then the crime region is just the jurisdiction of the investigating agency. If more than one agency is cooperating and sharing data, then the crime region will be the region formed by all of the jurisdictions sharing data.

2. **The home base region**. The home base region needs to be chosen sufficiently large to contain the anchor point of the offender; it must also contain the entire crime region. In our program, the home base region will be one or more county sized regions as selected by the program user.

3. **The search box**. This is the smallest rectangle that contains the crime region and the home base region.

The coarse mesh

As our first step in the geographic analysis, we will cover the search box with a coarse triangular mesh. Let the circumradius of the triangles in the coarse mesh be $R_{\text{coarse}} = R$. We assume that the grid contains:

- m rows of triangles, with m odd, and
- n columns of triangles, with n even.

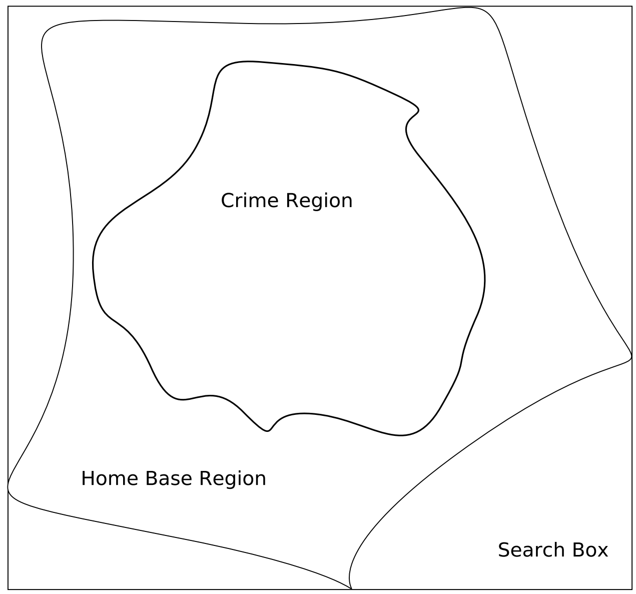


Figure 3.11: Relationships between crime region, home base region, and the search box

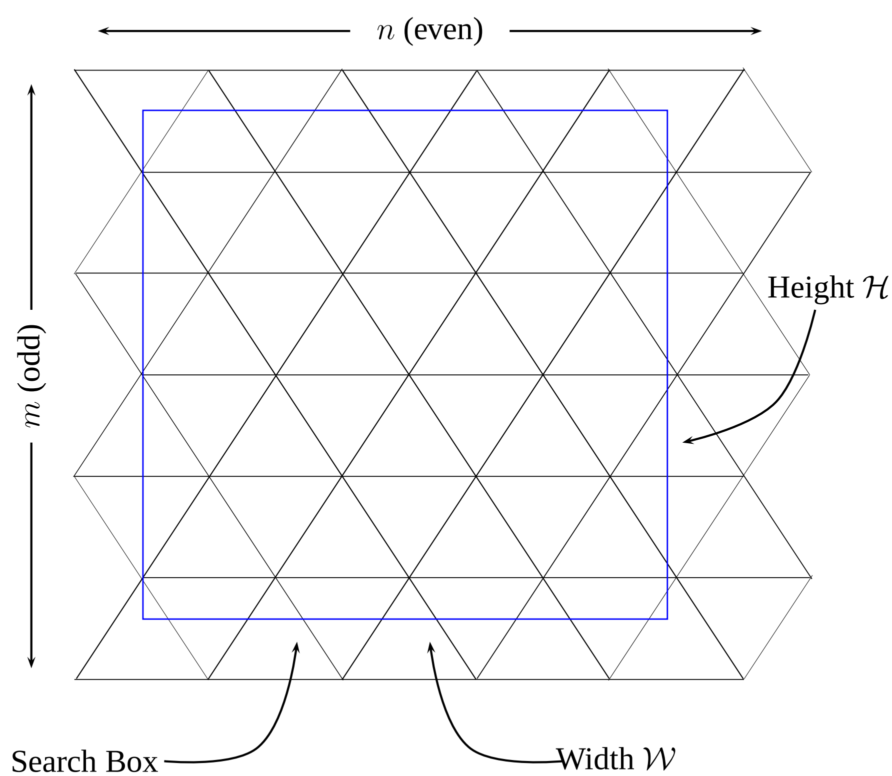


Figure 3.12: Coarse triangular mesh and the search box

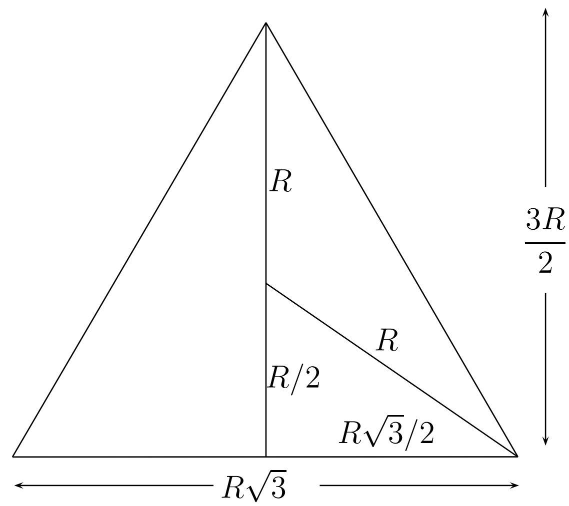


Figure 3.13: Dimensions of a typical triangle

To ensure that the coarse grid covers the search box, we place the top left corner of the search box at the centroid of the triangle in the first row and first column, as seen in Figure 3.12. Then, examining the size of each individual triangle, we see that to ensure that the search box remains entirely within the triangulated mesh, we need

$$(\tfrac{n}{2} - 1)R\sqrt{3} \geq \mathcal{W}$$
$$(m - 1)(3R/2) \geq \mathcal{H}$$

As a consequence, we choose

$$m \text{ odd, } m > 1 + \frac{2H}{3R} = 2 + \frac{2\mathcal{H}}{3R_{\text{coarse}}},$$
$$n \text{ even, } n > 2 + \frac{2W}{R\sqrt{3}} = 4 + \frac{2\mathcal{W}}{R_{\text{coarse}}\sqrt{3}}.$$

Area of a coarse mesh triangle In addition to the geometric dimensions of the coarse triangle shown in Figure 3.13, we will also need the area A of these triangles; for reference we simply note that

$$A = \frac{1}{2}\frac{3R}{2}R\sqrt{3} = \frac{3\sqrt{3}}{4}R^2.$$

Finding the orientation of the coarse mesh's triangles Label each triangle in the mesh by a pair of nonnegative integers (i, j), where this refers to the triangle in the i^{th} column and the j^{th} row; for consistency with our C++ code, we start these indices with zero, so $i \in \{0, 1, 2, \ldots, n-1\}$ and $j \in \{0, 1, \ldots, m-1\}$. We count the columns starting from the left edge working right, and we count the rows starting from the top and working downward.

Each triangle in the mesh has a side that is parallel to the x-axis. We say that the triangle points downward if the y-coordinate of the centroid is smaller that the y-coordinate of the side that is parallel to the x-axis; otherwise we say that the triangle points upward. We call this the orientation of the triangle.

Examining Figure 3.12, we see that

- $i + j =$ even if and only if the triangle points downward;
- $i + j =$ odd if and only if the triangle points upward.

Finding the centroids of the coarse mesh's triangles Suppose that the coordinates (x_0, y_0) of the top left corner of our search box are known. How can we find the coordinates of the centroids of all of the triangles in the coarse mesh?

Consider a triangle in the mesh at position (a, b) with center (x, y). Then choose (α, β), and ask what are the coordinates of the center of the triangle whose position in the mesh is $(a+\alpha, b+\beta) = (i, j)$. If $\alpha + \beta$ is even, then both the triangle (a, b) and the triangle $(a + \alpha, b + \beta)$ both have the same orientation, and so

$$\text{coordinates of } (a + \alpha, b + \beta) \text{ are } (x, y) + R\left(\alpha\tfrac{\sqrt{3}}{2}, \beta\left(\tfrac{-3}{2}\right)\right),$$

because each triangle has width $R\sqrt{3}/2$ and height $3R/2$ (see Figure 3.13) and because we count rows going downward.

On the other hand, if $\alpha + \beta$ is odd, then $\alpha + \beta - 1$ is even, and so

$$\text{coordinates of } (a + \alpha, b + \beta - 1) \text{ are } (x, y) + R\left(\alpha\tfrac{\sqrt{3}}{2}, (\beta - 1)\left(\tfrac{-3}{2}\right)\right).$$

Then, if $a+b$ is even, the original triangle pointed downward and so our new triangle points upward and hence

$$\text{coordinates of } (a + \alpha, b + \beta) \text{ are } (x, y) + R\left(\alpha\tfrac{\sqrt{3}}{2}, (\beta - 1)\left(\tfrac{-3}{2}\right)\right) + (0, -2R).$$

On the other hand, if $a + b$ is odd, then the original triangle pointed upward and our new triangle points downward; then

$$\text{coordinates of } (a + \alpha, b + \beta) \text{ are } (x, y) + R\left(\alpha\tfrac{\sqrt{3}}{2}, (\beta - 1)\left(\tfrac{-3}{2}\right)\right) + (0, -R).$$

With these calculations in hand, we can use the fact that the top left corner of the mesh (x_0, y_0) is the centroid of the triangle in position $(0, 0)$; thus the triangle is position (i, j) has centroid

- $(x_0, y_0) + R\left(i\frac{\sqrt{3}}{2}, j\left(\frac{-3}{2}\right)\right)$ if $i + j$ is odd, and
- $(x_0, y_0) + R\left(i\frac{\sqrt{3}}{2}, (j - 1)\left(\frac{-3}{2}\right) - 1\right)$ if $i + j$ is even.

These are the calculations that are used in our code, specifically in the constructor `CTriangulatedGeography::CTriangulatedGeography`.

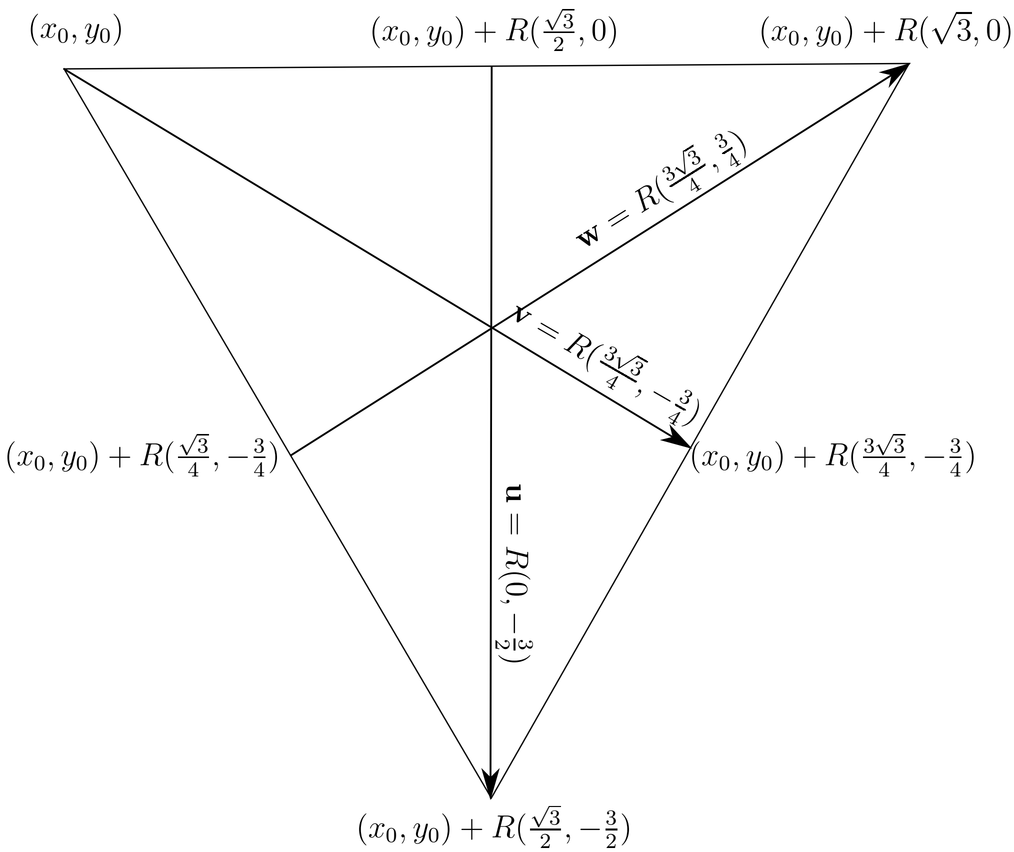


Figure 3.14: Vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ for a typical triangle in the coarse mesh

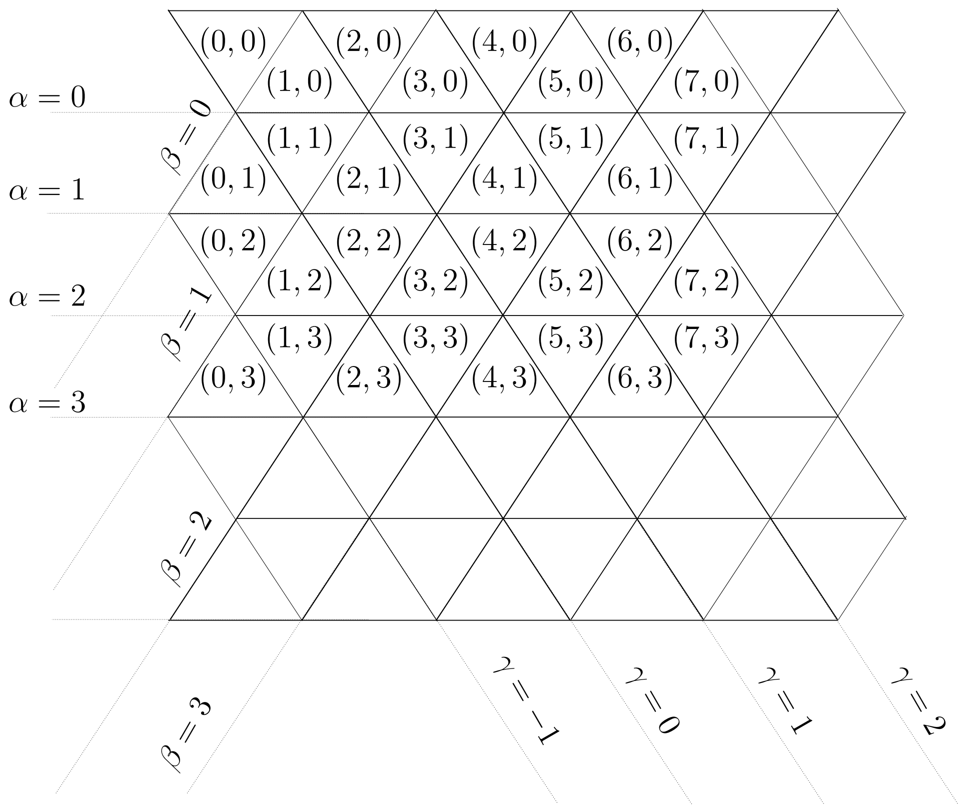


Figure 3.15: Coordinates in the coarse mesh

Determining the coarse mesh triangle containing a point Given a point with known coordinates (x, y), we would like to be able to determine the coordinates (i, j) of the coarse mesh triangle that contains that point. To do so, define the vector $\mathbf{r} = (x - x_0, y - y_0)$ and consider the vectors

$$\mathbf{u} = R(0, -\tfrac{3}{2}) \qquad \mathbf{v} = R(\tfrac{3\sqrt{3}}{4}, -\tfrac{3}{4}) \qquad \mathbf{w} = R(\tfrac{3\sqrt{3}}{4}, \tfrac{3}{4})$$

which are shown in Figure 3.14. Note that $|\mathbf{u}|^2 = |\mathbf{v}|^2 = |\mathbf{w}|^2 = 9R^2/4$.

Calculate the projection of $\mathbf{r}$ in the directions of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$, so that

$$\alpha = \left\lfloor \frac{\mathbf{r} \cdot \mathbf{u}}{|\mathbf{u}|^2} \right\rfloor, \qquad \beta = \left\lfloor \frac{\mathbf{r} \cdot \mathbf{v}}{|\mathbf{v}|^2} \right\rfloor, \qquad \gamma = \left\lfloor \frac{\mathbf{r} \cdot \mathbf{w}}{|\mathbf{w}|^2} \right\rfloor$$

where we have rounded each fraction to the largest integer lower than the given fraction. Then examining Figure 3.15, we see that the triplet (α, β, γ) then gives us our location within the mesh, and the coordinates (i, j) of the triangle are then found by simply calculating

$$i = \beta + \gamma \qquad j = \alpha.$$

Determining the coarse mesh triangle near a point Suppose that a point (x_0, y_0) is known, and we want to find all of the coarse mesh triangles within a distance ρ from (x_0, y_0). How we proceed depends on the relationship between ρ and R.

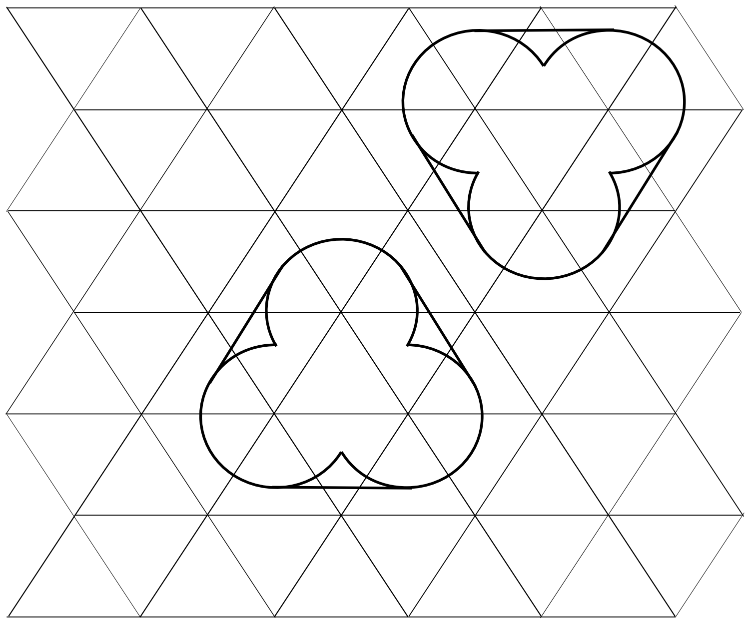


Figure 3.16: Determining the coarse mesh triangles near a point; $\rho < R$.

Small distances In the case where $\rho < R$, we begin by drawing circles of radius R centered at the three vertices of the triangle. Because every point in the triangle is a convex combination of the three vertices, it suffices to then take the convex hull of the resulting three circles; this is done in figure 3.16.

Suppose that (x_0, y_0) is in the coarse mesh triangle with coordinates (i_0, j_0). If this triangle is oriented upwards, then examining the figure, we see that all points within a distance $\rho < R$ of (x_0, y_0) lie in the triangles with coordinates:

- $(i_0 - 2, j_0)$, $(i_0 - 2, j_0 + 1)$
- $(i_0 - 1, j_0 - 1)$, $(i_0 - 1, j_0)$, $(i_0 - 1, j_0 + 1)$
- $(i_0, j_0 - 1)$, (i_0, j_0), $(i_0, j_0 + 1)$
- $(i_0 + 1, j_0 - 1)$, $(i_0 + 1, j_0)$, $(i_0 + 1, j_0 + 1)$
- $(i_0 + 2, j_0)$, $(i_0 + 2, j_0 + 1)$

while if the triangle containing (x_0, y_0) is oriented downwards, then all points within a distance $\rho < R$ of (x_0, y_0) lie in the triangles with coordinates:

- $(i_0 - 2, j_0 - 1)$, $(i_0 - 2, j_0)$
- $(i_0 - 1, j_0 - 1)$, $(i_0 - 1, j_0)$, $(i_0 - 1, j_0 + 1)$
- $(i_0, j_0 - 1)$, (i_0, j_0), $(i_0, j_0 + 1)$
- $(i_0 + 1, j_0 - 1)$, $(i_0 + 1, j_0)$, $(i_0 + 1, j_0 + 1)$
- $(i_0 + 2, j_0 - 1)$, $(i_0 + 2, j_0)$

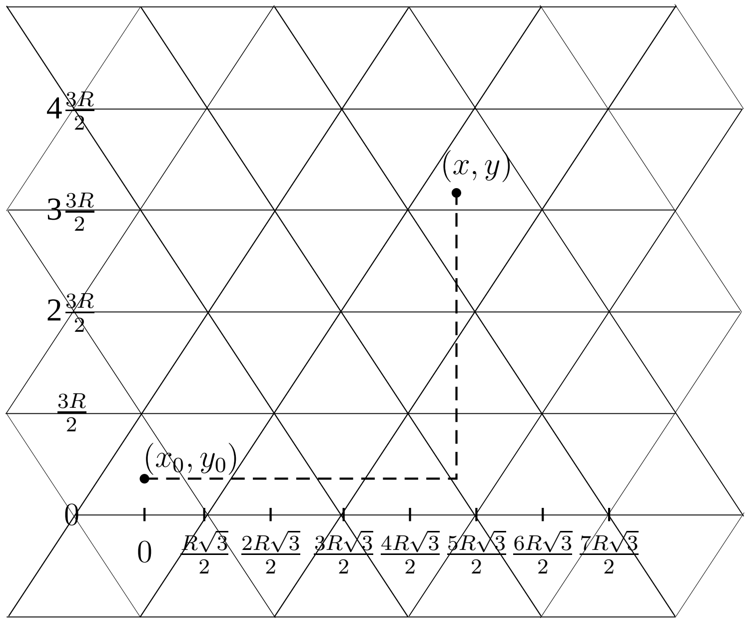


Figure 3.17: Distances in the coarse mesh

Large Distances To handle the situation where $\rho \geq R$, let us consider the situation where

- (x_0, y_0) is in the triangle (i_0, j_0),
- (x, y) is in the triangle (i, j), and
- $\mathrm{dist}((x_0, y_0), (x, y)) \leq \rho$.

Define the numbers $\Delta i = |i - i_0|$, $\Delta j = |j - j_0|$. Examining Figure 3.17, we can see that

$$(\Delta i - 2)\frac{R\sqrt{3}}{2} \leq |x - x_0| \leq (\Delta i + 2)\frac{R\sqrt{3}}{2},$$
$$(\Delta j - 1)\frac{3R}{2} \leq |y - y_0| \leq (\Delta j + 1)\frac{3R}{2}.$$

Thus

$$(\Delta i - 2)\frac{R\sqrt{3}}{2} \leq |x - x_0| \leq \mathrm{dist}((x_0, y_0), (x, y)) \leq \rho$$
$$(\Delta j - 1)\frac{3R}{2} \leq |y - y_0| \leq \mathrm{dist}((x_0, y_0), (x, y)) \leq \rho$$

so that

$$\Delta i \leq 2 + \frac{2\rho}{\sqrt{3}R} \qquad \Delta j \leq 1 + \frac{2\rho}{3R}.$$

Thus, we set

$$\Delta i = 2 + \left\lceil \frac{2\rho}{\sqrt{3}R} \right\rceil \qquad \Delta j = 1 + \left\lceil \frac{2\rho}{3R} \right\rceil$$

Then, given the triangle (i_0, j_0), all of the triangles (i, j) that contain a point of distance no more than ρ from triangle (i_0, j_0) satisfy

$$|i - i_0| \leq \Delta i \qquad |j - j_0| \leq \Delta j$$

Though fast to implement, this method is not sharp. For example, if $\rho < R$, then this method returns all triangles with $|i - i_0| \leq 3$ and $|j - j_0| \leq 2$, which is a larger set than what we returned above.

The fine mesh

Differing quantities of interest vary on very different distance scales. For example, the target attractiveness $G(\mathbf{x})$ from (3.10) may vary dramatically from one block to another. Other functions, like the normalization function $N(\mathbf{x})$ from (3.9) vary much more slowly. If we were to try to use a single mesh for all of our calculations, we would either lose accuracy because the mesh is too coarse for the rapidly varying functions, or we would lose performance because we would be calculating essentially the same slowly varying variables over and over.

Our solution is to use two different meshes, the coarse grid which we have already described, and another finer mesh, which we will obtain by subdividing the coarse mesh triangles.

Triangular coordinates To construct our fine mesh, we subdivide each triangle from the coarse mesh into N^2 subtriangles. We then need to determine the characteristics of each of these subtriangles from our knowledge of the characteristics of the original, coarse mesh triangle. To do so, we start by introducing the usual set of triangular coordinates. To do so, label the vertices as $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$. Then a point $\mathbf{x}$ in the triangle has triangular coordinates (α, β, γ) if and only if

$$\mathbf{x} = \alpha \mathbf{v}_1 + \beta \mathbf{v}_2 + \gamma \mathbf{v}_3$$

where $\alpha + \beta + \gamma = 1$.

We then split the collection of subtriangles into those with the same orientation as the parent triangle and those with the opposite orientation as the parent triangle. See Figures 3.18 and 3.20; note that the orientation of the parent triangle is irrelevant though the parent triangle in these figures is shown with an upward orientation for definiteness in the figure and the following discussion.

Subtriangles with the same orientation as the parent We begin by considering the subcollection of subtriangles with the same orientation as the parent. Examining Figure 3.18, we see that there are $\sum_{i=1}^{N} i = \frac{1}{2}N(N+1)$ such subtriangles. We index each subtriangle by the triangular coordinates of the bottom left corner of the subtriangle. For any i, j, k satisfying $i+j+k = N$, $j \geq 1$ and $i, k \geq 0$, the point $(i/N, j/N, k/N)$ is the bottom left corner of a triangle in our subcollection. This condition on i, j, k is equivalent to the requirements

- $0 \leq i \leq N - 1$,
- $1 \leq j \leq N - i$, and
- $k = N - i - j$

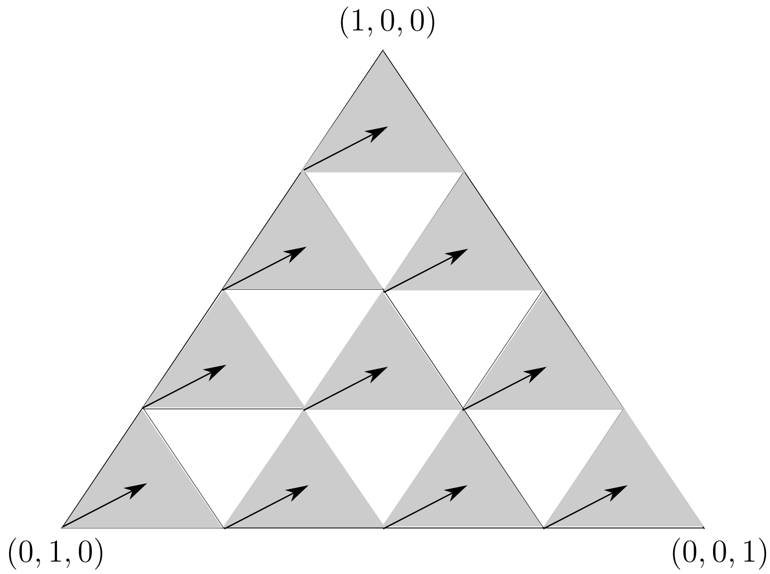


Figure 3.18: Subdivided coarse triangle- subtriangles with the same orientation

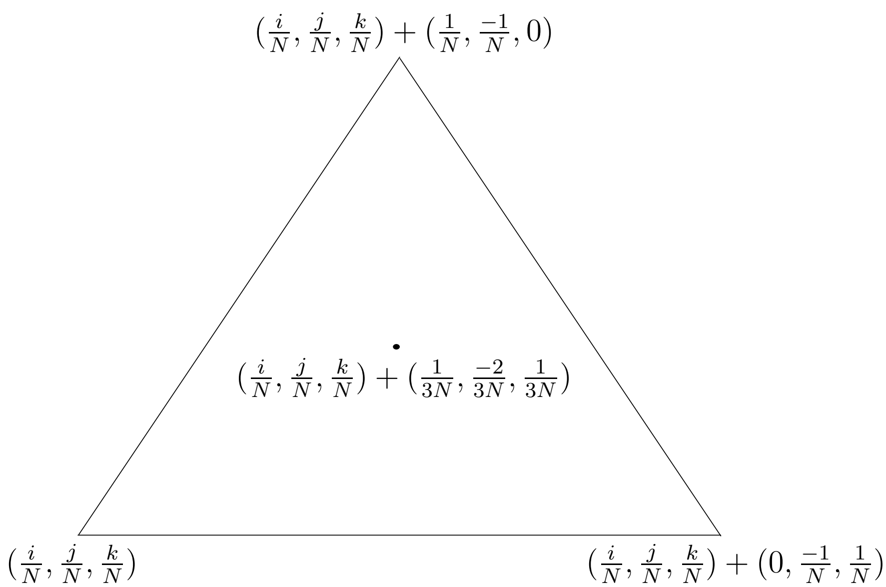


Figure 3.19: Calculating the centroid of a subtriangle

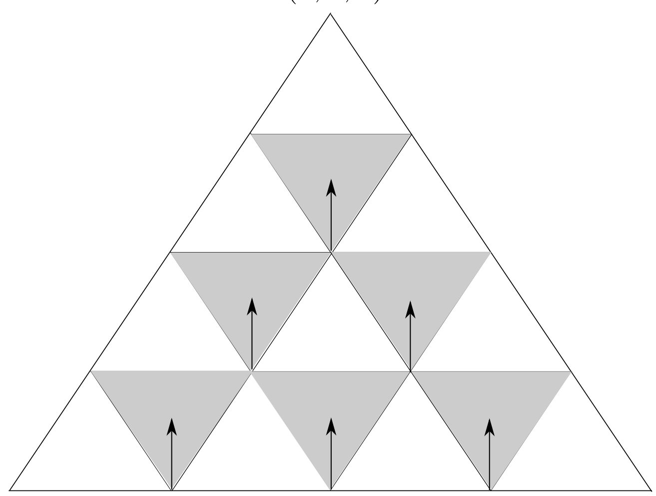


Figure 3.20: Subdivided coarse triangle- subtriangles with the opposite orientation

We also notice that the vector from the bottom left corner to the centroid of the subtriangle is, in triangular coordinates, the vector $(\frac{1}{3N}, \frac{-2}{3N}, \frac{1}{3N})$; this follows immediately from Figure 3.19. Thus, the centroid of the triangle with bottom left corner $(i/N, j/N, k/N)$ is $(\frac{i}{N}, \frac{j}{N}, \frac{k}{N}) + (\frac{1}{3N}, \frac{-2}{3N}, \frac{1}{3N})$.

Subtriangles with the opposite orientation as the parent Examining figure 3.20, we see that there are $\sum_{i=1}^{N-1} i = \frac{1}{2}(N-1)N$ subtriangles whose orientation is opposite to that of the parent. We index these subtriangles by the triangular coordinates of their bottom vertex. Then, for any i, j, k satisfying $i + j + k = N$, $j, k \geq 1$, $i \geq 0$, the point $(i/N, j/N, k/N)$ is the bottom vertex of a triangle in our subcollection. This condition on i, j, k is equivalent to the requirements

- $0 \leq i \leq N-2$,
- $1 \leq j \leq N-i-1$, and
- $k = N-i-j$.

We also notice that the vector from the bottom vertex to the centroid of the subtriangle is, in triangular coordinates, the vector $(\frac{-2}{3N}, \frac{1}{3N}, \frac{1}{3N})$; this follows immediately from Figure 3.21. Thus, the centroid of the triangle with bottom vertex $(i/N, j/N, k/N)$ is $(\frac{i}{N}, \frac{j}{N}, \frac{k}{N}) + (\frac{-2}{3N}, \frac{1}{3N}, \frac{1}{3N})$.

Cartesian Coordinates Now that we have determined the triangular coordinates for the centroids of all of the subtriangles, we need to determine their corresponding Cartesian coordinates. To do so, we begin by finding Cartesian coordinates of the point with triangular coordinates (α, β, γ) with $\alpha + \beta + \gamma = 1$ for a parent triangle with centroid (x_0, y_0) and circumradius R.

Examining Figure 3.22, we see clearly that the result will depend on the orientation of the parent triangle. In the case where the parent triangle has an upward orientation (as was the default shown in Figures 3.18 and 3.20) we see that we have the following relationship between the Cartesian and triangular coordinates of the vertices of the parent triangle

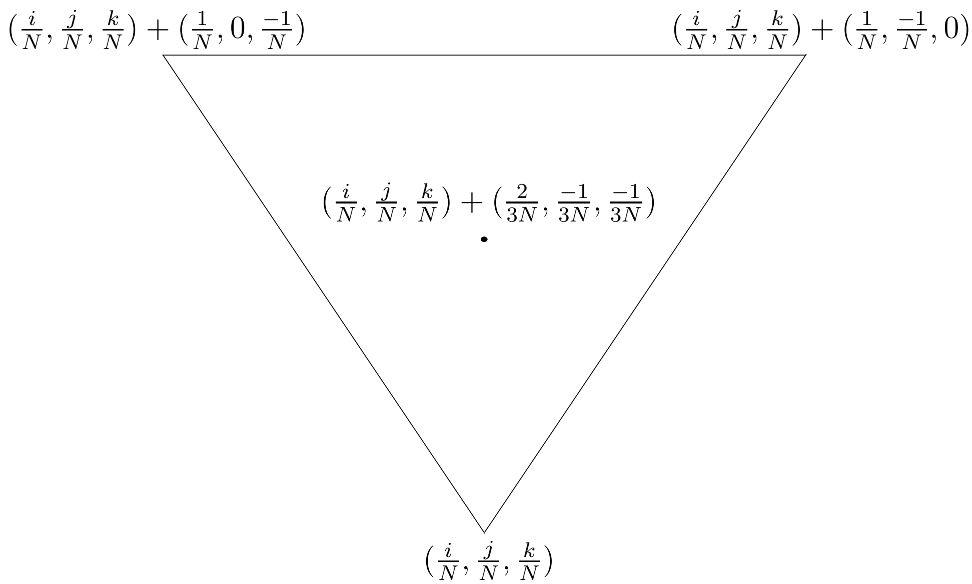


Figure 3.21: Calculating the centroid of a subtriangle

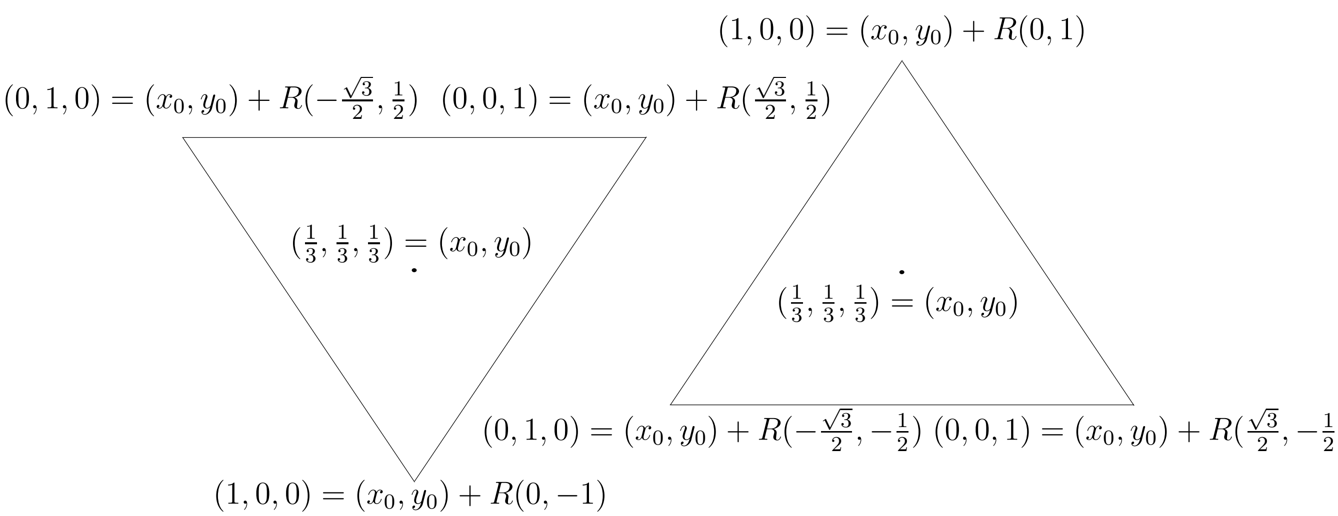


Figure 3.22: Converting from triangular coordinates to Cartesian coordinates. Left: Parents with a downward orientation. Right: Parents with an upward orientation

Parent oriented upwards	
Cartesian Coordinate	Triangular Coordinate
$(1, 0, 0)$	$(x_0, y_0) + R(0, 1)$
$(0, 1, 0)$	$(x_0, y_0) + R(-\frac{\sqrt{3}}{2}, -\frac{1}{2})$
$(0, 0, 1)$	$(x_0, y_0) + R(\frac{\sqrt{3}}{2}, -\frac{1}{2})$

As a consequence, the Cartesian coordinates for the point with triangular coordinates (α, β, γ) in a parent triangle with upward orientation are

$$(\alpha, \beta, \gamma) \mapsto (x_0, y_0) + R(-\tfrac{\sqrt{3}}{2}\beta + \tfrac{\sqrt{3}}{2}\gamma, \alpha - \tfrac{1}{2}\beta - \tfrac{1}{2}\gamma).$$

On the other hand, if the parent triangle has a downward orientation, then we have the following

Parent oriented downwards	
Cartesian Coordinate	Triangular Coordinate
$(1, 0, 0)$	$(x_0, y_0) + R(0, -1)$
$(0, 1, 0)$	$(x_0, y_0) + R(-\frac{\sqrt{3}}{2}, \frac{1}{2})$
$(0, 0, 1)$	$(x_0, y_0) + R(\frac{\sqrt{3}}{2}, \frac{1}{2})$

As a consequence, the Cartesian coordinates for the point with triangular coordinates (α, β, γ) in a parent triangle with downward orientation are

$$(\alpha, \beta, \gamma) \mapsto (x_0, y_0) + R(-\tfrac{\sqrt{3}}{2}\beta + \tfrac{\sqrt{3}}{2}\gamma, -\alpha + \tfrac{1}{2}\beta + \tfrac{1}{2}\gamma).$$

3.2.5 Evaluating the probability density $P(\mathbf{z})$

Our goal is to determine the probability density $P(\mathbf{z})$ that the offender's anchor point is located at the point $\mathbf{z}$. Of course, this function is defined for every value of $\mathbf{z}$. Rather than calculate this for every value of $\mathbf{z}$ however, we will only calculate it for the points $\mathbf{z}$ that are the centroid of a triangle in our fine mesh.

3.2.6 The Normalization Function

The normalization function $N(\mathbf{z}, \alpha)$ defined by (3.9)

$$N(\mathbf{z}, \alpha) = \left[\iint_{\mathbf{R}^2} D(\mathbf{x} \,|\, \mathbf{z}, \alpha) G(\mathbf{x}) \, dx^{(1)} dx^{(2)} \right]^{-1}. \tag{3.25}$$

is effectively impossible to evaluate analytically, and very difficult to do so numerically. However, it does possess some properties that will enable us to evaluate it more efficiently. As we will be focusing in the integral rather than on its inverse N, we adopt the notation

$$I(\mathbf{z}, \alpha) = \iint_{\mathbf{R}^2} D(\mathbf{x} \,|\, \mathbf{z}, \alpha) G(\mathbf{x}) \, dx^{(1)} dx^{(2)}$$

in what follows.

The Tail

First, we notice thanks to the definition of $G(\mathbf{x})$ in (3.10), there is an absolute constant Γ so that

$$0 \leq G(\mathbf{x}) \leq \Gamma$$

for all $\mathbf{x} \in \mathbf{R}^2$. We also note that the graph of $f(r \mid \alpha)$ has a single positive maximum after which it decays rapidly; see Figure 3.2. In particular, we note that

$$\max f(r \mid \alpha) = \frac{1}{\alpha}\sqrt{\frac{\pi}{2e}} \approx \frac{0.760173}{\alpha},$$
$$\arg\max f(r \mid \alpha) = \alpha\sqrt{\frac{2}{\pi}} \approx 0.797885\alpha.$$

To actually evaluate $I(\mathbf{z}, \alpha)$, we begin by replacing the region of integration by the region $[|\mathbf{x} - \mathbf{z}| < R]$ for some R. The resulting error is

$$\begin{aligned}
\left| \iint_{\mathbf{R}^2} D(\mathbf{x} \mid \mathbf{z}, \alpha) G(\mathbf{x}) d\mathbf{x} - \iint_{[|\mathbf{x}-\mathbf{z}|<R]} D(\mathbf{x} \mid \mathbf{z}, \alpha) G(\mathbf{x}) d\mathbf{x} \right| \\
&= \iint_{[|\mathbf{x}-\mathbf{z}|\geq R]} D(\mathbf{x} \mid \mathbf{z}, \alpha) G(\mathbf{x}) d\mathbf{x} \\
&\leq 2\pi\Gamma \int_R^\infty \frac{r}{4\alpha^2} \exp\left(-\frac{\pi r^2}{4\alpha^2}\right) \, dr \\
&\leq \Gamma \exp\left(-\frac{\pi R^2}{4\alpha^2}\right)
\end{aligned}$$

The exact value of $I(\mathbf{z}, \alpha)$ is unknown, so we are unable to estimate the relative error; however if $G(\mathbf{x}) \approx \lambda\Gamma$ for all $\mathbf{x}$, then the integral is approximately $\lambda\Gamma$; thus we use the factor $\exp\left(-\frac{\pi R^2}{4\alpha^2}\right)$ as a proxy for the relative error.

In particular, if we want out estimate for the relative error caused by neglecting the tail to satisfy

$$\exp\left(-\frac{\pi R^2}{4\alpha^2}\right) \leq 10^{-k}$$

then

$$\frac{\pi R^2}{4\alpha^2} \geq k \ln 10$$

so we choose

$$R \geq 2\sqrt{\frac{\ln 10}{\pi}}\sqrt{k}\alpha \approx 1.71223\sqrt{k}\alpha.$$

Reasonable values are

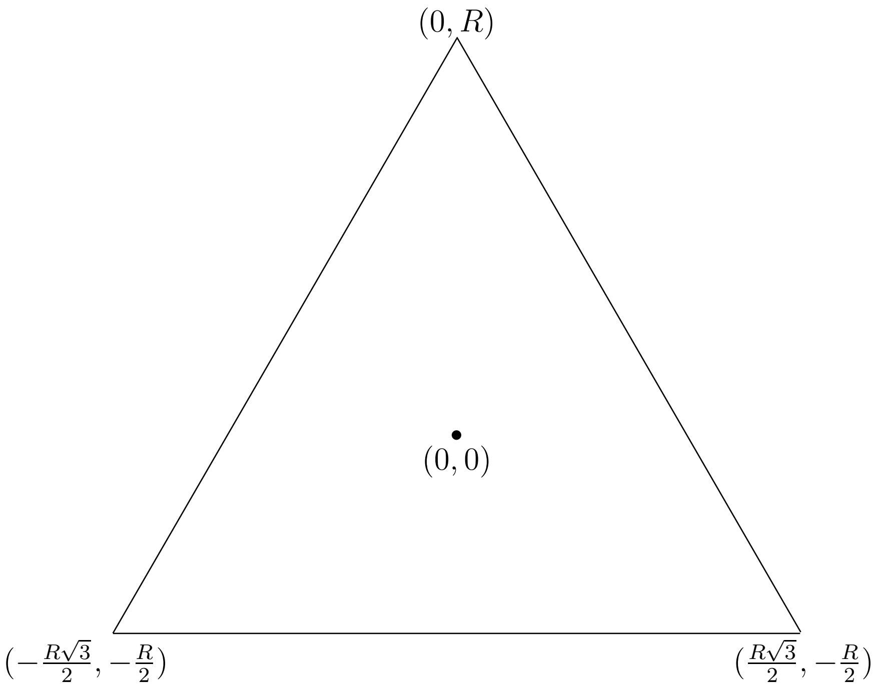


Figure 3.23: Discretization of an integral

10^{-k}	R_{crit}
10^{-1}	1.71223α
10^{-2}	2.42146α
10^{-3}	2.96567α
10^{-4}	3.42447α
10^{-5}	3.82687α
10^{-6}	4.19410α
10^{-7}	4.53014α
10^{-8}	4.84293α

The discretization

To evaluate the portion of the integral excluding the tail, *i.e.* the portion of the integral for which $|\mathbf{x} - \mathbf{z}| < R$, we use the two-dimensional midpoint rule.

The two-dimensional midpoint rule For simplicity, we start by describing how to evaluate the integral of an arbitrary function $f(x, y)$ over an equilateral triangle with centroid at the origin and circumradius R, as seen in figure 3.23

Let T be our equilateral triangle; then for any $(x, y) \in T$, there is a point $(x_0, y_0) \in T$ so that Taylor's Theorem will let us write

$$f(x,y) = f(0,0) + \frac{\partial f}{\partial x}(0,0)x + \frac{\partial f}{\partial y}(0,0)y + \frac{1}{2}\left(\frac{\partial^2 f}{\partial x^2}(x_0,y_0)x^2 + 2\frac{\partial^2 f}{\partial x \partial y}(x_0,y_0)xy + \frac{\partial^2 f}{\partial y^2}(x_0,y_0)y^2\right)$$

We now integrate both sides in (x, y) over T.

A simple symmetry argument shows that

$$\iint_T \frac{\partial f}{\partial x}(0,0)x\; dx\; dy = 0$$

On the other hand

$$\begin{aligned}
\iint_T \frac{\partial f}{\partial y}(0,0)y\; dx\; dy &= 2f_y(0,0)\int_{-R/2}^{R}\int_0^{-\frac{\sqrt{3}}{3}(y-R)} y\; dx\; dy \\
&= 2f_y(0,0)\int_{-R/2}^{R} -\frac{\sqrt{3}}{3}y(y-R)\; dy \\
&= -\frac{2\sqrt{3}}{3}f_y(0,0)\left[\frac{1}{3}y^3 - \frac{1}{2}Ry^2\right]_{-R/2}^{R} \\
&= -\frac{2\sqrt{3}}{3}f_y(0,0)\left[\frac{1}{3}R^3 + \frac{1}{3}\left(\frac{R}{2}\right)^3 - \frac{1}{2}RR^2 + \frac{1}{2}R\left(\frac{R}{2}\right)^2\right]_{-R/2}^{R} \\
&= -\frac{2\sqrt{3}}{3}f_y(0,0)\left[\frac{1}{3} + \frac{1}{24} - \frac{1}{2} + \frac{1}{8}\right]R^3 = 0.
\end{aligned}$$

Thus,

$$\iint_T f(x,y)\; dx\; dy = A(T)f(0,0) + E$$

where $A(T)$ is the area of triangle T, and the error E satisfies

$$E = \frac{1}{2}\iint_T \left(\frac{\partial^2 f}{\partial x^2}(x_0,y_0)x^2 + 2\frac{\partial^2 f}{\partial x \partial y}(x_0,y_0)xy + \frac{\partial^2 f}{\partial y^2}(x_0,y_0)y^2\right)$$

so that

$$|E| \le \frac{1}{2}\|f_{xx}\|_\infty \iint_T x^2\; dx\; dy + \|f_{xy}\|_\infty \iint_T |xy|\; dx\; dy + \frac{1}{2}\|f_{yy}\|_\infty \iint_T y^2\; dx\; dy$$

Now direct evaluation shows us that

$$\begin{aligned}
\iint_T x^2\; dx\; dy &= \frac{3\sqrt{3}}{8}R^4 \\
\iint_T y^2\; dx\; dy &= \frac{3\sqrt{3}}{8}R^4 \\
\iint_T |xy|\; dx\; dy &= \frac{15}{32}R^4
\end{aligned}$$

and consequently

$$|E| \le \left(\frac{15}{16} + \frac{3\sqrt{3}}{4}\right)\left\|D^2 f\right\|_\infty R^4$$

We will then use the midpoint approximation

$$\iint_T f(x,y)\,dx\,dy = \left(\frac{3\sqrt{3}}{4}R^2\right) f(0,0) + O(R^4)$$

to evaluate our integral.

Application to the normalization function Now

$$I(\mathbf{z},\alpha) = \iint_{\mathbf{R}^2} D(\mathbf{x}\,|\,\mathbf{z}|,\alpha)G(\mathbf{x})\,dx^{(1)}\,dx^{(2)} = \iint_{\mathbf{R}^2} \frac{1}{4\alpha^2}\exp\left(-\frac{\pi}{4\alpha^2}|\mathbf{x}-\mathbf{z}|^2\right) G(\mathbf{x})\,dx^{(1)}\,dx^{(2)}$$

Let Δ be a collection of equilateral triangles whose closures cover the plane and whose interiors are disjoint. Then, we can rewrite the integral above as

$$\iint_{\mathbf{R}^2} D(\mathbf{x}\,|\,\mathbf{z}|,\alpha)G(\mathbf{x})\,dx^{(1)}\,dx^{(2)} = \sum_{T\in\Delta}\iint_T \frac{1}{4\alpha^2}\exp\left(-\frac{\pi}{4\alpha^2}|\mathbf{x}-\mathbf{z}|^2\right) G(\mathbf{x})\,dx^{(1)}\,dx^{(2)}.$$

Using the midpoint method, we then obtain the approximation

$$\iint_{\mathbf{R}^2} D(\mathbf{x}\,|\,\mathbf{z}|,\alpha)G(\mathbf{x})\,dx^{(1)}\,dx^{(2)} \approx \frac{3\sqrt{3}}{16}\frac{1}{\alpha^2}\sum_{T\in\Delta} R_T^2 \exp\left(-\frac{\pi}{4\alpha^2}|\mathbf{x}_T-\mathbf{z}|^2\right) G(\mathbf{x}_T)$$

where R_T is the circumradius and x_T is the centroid of the triangle T.

Selecting the mesh

We begin by selecting a circumradius of the coarse mesh triangles, R_{coarse}. Based on our analysis of the tail, we see that relative error in replacing the integral over $\mathbf{R}^2$ with the integral over the smaller region $[|\mathbf{x}-\mathbf{z}| < 3\alpha]$ is roughly $e^{-9\pi/4} \approx 0.000851$. Thus, we shall drop from the mesh Δ any triangle T for which $|\mathbf{x}_T - \mathbf{z}| > 3\alpha$

We also see that the relative error in replacing the integral over $\mathbf{R}^2$ with the integral over the even smaller region $[|\mathbf{x}-\mathbf{z}| < 2\alpha]$ is roughly $e^{-\pi} \approx 0.0432$. Because this is not insignificant, we will retain all of the triangles for which $2\alpha < |\mathbf{x}_T - \mathbf{z}| \le 3\alpha$.

Finally, our analysis of the tail tells us that the vast majority of the contributions to our integral come from the region $|\mathbf{x}_T - \mathbf{z}| \le 2\alpha$. For this reason, we subdivide all of the coarse triangles in this region into their corresponding fine subtriangles before evaluating the integral.

This is the process that is used in the code to evaluate $I(\mathbf{z},\alpha)$.

Evaluating $I(\mathbf{z},\alpha)$ as $\mathbf{z}$ varies

Although we have outlined how we are able to evaluate the integral $I(\mathbf{z},\alpha)$, we do not want to do so for every potential value of $\mathbf{z}$. As described in Section 3.2.5, we only need to evaluate $P(\mathbf{z})$ and hence $I(\mathbf{z})$ for points $\mathbf{z}$ that are the centroids of triangles in the fine mesh. However, this turns out to be impractical, as the computation time to evaluate $I(\mathbf{z})$ even once is significant. Testing has shown that this process, simplified as it was in the previous discussion, is still by far the most computationally expensive portion of the algorithm.

Rather than use this algorithm at the centroid of every fine triangle, we instead use this method only to calculate the values of $I(\mathbf{z})$ on the vertices of the coarse triangles and interpolate into the fine triangles within.

Suppose we are given a coarse triangle T and the values of I at the vertices of T; and want to use interpolation to approximate the values of I in the interior of T. Let $\mathbf{c}$ be the center and R the circumradius of the coarse triangle T; then the vectors from $\mathbf{c}$ to the vertices of T are

$$\begin{aligned}
\mathbf{v}_1 &= \pm R\left\langle 0, 1\right\rangle \\
\mathbf{v}_2 &= \pm R\left\langle \frac{\sqrt{3}}{2}, -\frac{1}{2}\right\rangle \\
\mathbf{v}_3 &= \pm R\left\langle -\frac{\sqrt{3}}{2}, -\frac{1}{2}\right\rangle
\end{aligned}$$

where the $+$ sign is chosen when the triangle is oriented upwards and the $-$ sign when the triangle is oriented downwards. The vertices of T are then $\{\mathbf{c} + \mathbf{v}_i\}_{i=1}^{3}$.

We then examine the known values of I on the vertices of T, naming them

$$\begin{aligned}
I_1 &= I_1(\alpha) = I(\mathbf{c} + \mathbf{v}_1, \alpha), \\
I_2 &= I_2(\alpha) = I(\mathbf{c} + \mathbf{v}_2, \alpha), \\
I_3 &= I_3(\alpha) = I(\mathbf{c} + \mathbf{v}_3, \alpha).
\end{aligned}$$

Now select a point $\mathbf{z} \in T$, and let $\mathbf{x} = \mathbf{z} - \mathbf{c}$; then

$$\begin{aligned}
\mathbf{x} &= t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + t_3\mathbf{v}_3 \\
\mathbf{z} &= \mathbf{c} + t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + t_3\mathbf{v}_3
\end{aligned}$$

where

$$t_i = \frac{1}{3} + \frac{2}{3}\frac{\mathbf{x} \cdot \mathbf{v}_i}{R^2}.$$

Indeed

$$\begin{aligned}
\mathbf{z} &= \mathbf{c} + t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + t_3\mathbf{v}_3 \\
&= \mathbf{c} + \tfrac{1}{3}(\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3) + \tfrac{2}{3}\tfrac{1}{R^2}\left\{(\mathbf{x} \cdot \mathbf{v}_1)\mathbf{v}_1 + (\mathbf{x} \cdot \mathbf{v}_2)\mathbf{v}_2 + (\mathbf{x} \cdot \mathbf{v}_3)\mathbf{v}_3\right\} \\
&= \mathbf{c} + \tfrac{2}{3}\tfrac{1}{R^2}\left\{(\mathbf{x} \cdot \mathbf{v}_1)\mathbf{v}_1 + (\mathbf{x} \cdot \mathbf{v}_2)\mathbf{v}_2 + (\mathbf{x} \cdot \mathbf{v}_3)\mathbf{v}_3\right\} \\
&= \mathbf{c} + \tfrac{2}{3}\left\{\left\langle 0, x_2\right\rangle + \left(\tfrac{\sqrt{3}}{2}x_1 - \tfrac{1}{2}x_2\right)\left\langle \tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}\right\rangle + \left(-\tfrac{\sqrt{3}}{2}x_1 - \tfrac{1}{2}x_2\right)\left\langle -\tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}\right\rangle\right\} \\
&= \mathbf{c} + \tfrac{2}{3}\left\{\left\langle 0, x_2\right\rangle + \left\langle \tfrac{3}{2}x_1, \tfrac{1}{2}x_2\right\rangle\right\} \\
&= \mathbf{c} + \mathbf{x} = \mathbf{z}.
\end{aligned}$$

as required.

Thus, for any point $\mathbf{z} \in T$, we calculate the numbers t_i and use the linear approximation

$$I(\mathbf{z}, \alpha) \approx t_1 I_1 + t_2 I_2 + t_3 I_3.$$

Evaluating $I(\mathbf{z}, \alpha)$ as α varies

The approximation of $I(\mathbf{z}, \alpha)$ by linear interpolation within a coarse triangle does not always produce reasonable results; in fact the accuracy of the approximation deteriorates as $\alpha \downarrow 0$.

Recall that α is the average distance the offender is willing to travel and that the dependence on $\mathbf{z}$ of the integrand is through (3.11),

$$D(\mathbf{x} \,|\, \mathbf{z}, \alpha) = \frac{1}{4\alpha^2} \exp\left(-\frac{\pi}{4} \left[\frac{|\mathbf{x} - \mathbf{z}|}{\alpha} \right]^2 \right).$$

Examining this, we see that if $\mathbf{z}_1$ and $\mathbf{z}_2$ are far apart relative to α, then $|\mathbf{x} - \mathbf{z}_1|/\alpha$ and $|\mathbf{x} - \mathbf{z}_2|/\alpha$ are very different, and so $I(\mathbf{z}_1, \alpha)$ and $I(\mathbf{z}_2, \alpha)$ are likely different.

Since the coarse triangles have circumradius $R = R_{\text{coarse}}$, if $R \ll \alpha$, then the variation in the integrand over T is small, and the interpolation should provide reasonable approximations; on the other hand if $R \gg \alpha$, then the variation of the integrand over T is large, and we expect poor approximations. This has been observed in numerical results.

To proceed, we write the integral as

$$\begin{aligned} I(\mathbf{z}, \alpha) &= \iint_{\mathbf{R}^2} D(\mathbf{x} \,|\, \mathbf{z}, \alpha) G(\mathbf{x}) dx^{(1)} dx^{(2)} \\ &= \frac{1}{4\alpha^2} \iint_{\mathbf{R}^2} \exp\left(-\frac{\pi}{4} \frac{|\mathbf{x} - \mathbf{z}|^2}{\alpha^2} \right) G(\mathbf{x}) dx^{(1)} dx^{(2)} \end{aligned}$$

Set $\xi = \frac{1}{2\alpha}(\mathbf{x} - \mathbf{z})$, then $d\xi = \frac{1}{4\alpha^2} d\mathbf{x}$ so

$$I(\mathbf{z}, \alpha) = \iint_{\mathbf{R}^2} \exp(-\pi |\xi|^2) G(\mathbf{z} + 2\alpha\xi)\, d\xi.$$

From this we can clearly see that $I(\alpha) \to G(\mathbf{z})$ as $\alpha \downarrow 0$.

Continuing our analysis, we can replace G by its Taylor series; then

$$I(\mathbf{z}, \alpha) = \iint_{\mathbf{R}^2} \exp(-\pi |\xi|^2) \Big\{ G(\mathbf{z}) + 2\alpha\, DG(\mathbf{z}) \cdot \xi + 4\alpha^2 \xi^\top D^2 G(\mathbf{z}) \xi + O(\alpha^3) \Big\}\, d\xi.$$

Note that the second term vanishes; indeed

$$\iint_{\mathbf{R}^2} \exp(-\pi |\xi|^2) DG(\mathbf{z}) \cdot \xi \, d\xi = \int_0^{2\pi} \int_0^\infty e^{-\pi r^2} |DG(\mathbf{z})| r \cos\theta \cdot r \, dr \, d\theta = 0$$

where θ is the angle measured from the direction $DG(\mathbf{z})/|DG(\mathbf{z})|$. Thus, we can write

$$I(\mathbf{z}, \alpha) \approx G(\mathbf{z}) + c_2(\mathbf{z}) \alpha^2 + O(\alpha^3)$$

for small α.

With this in mind, we wish to approximate $I(\mathbf{z}, \alpha)$ for α near 0 by interpolation. We suppose that we already have approximations for $I(\mathbf{z}, \alpha)$ for $\alpha = \alpha_1 < \alpha_2 < \cdots < \alpha_N$, but that the accuracy of the approximations diminishes as $\alpha \downarrow 0$.

We begin by choosing K so large that our approximations are reasonable at α_k for $k \geq K$. In particular, we choose K so large that $\alpha_K^* \geq 2R$, that is

$$\alpha_K^* = (K - \tfrac{1}{2})\epsilon \geq 2R$$

so we want

$$K \geq \tfrac{1}{2} + 2(R/\epsilon)$$

To proceed, we use the Hermite approximation

$$A(\alpha) = \omega_0 + \omega_1(\alpha - \alpha_K) + \omega_2(\alpha - \alpha_K)^2 + \omega_3\alpha(\alpha - \alpha_K)^2$$

so that

$$A'(\alpha) = \omega_1 + 2\omega_2(\alpha - \alpha_K) + \omega_3(\alpha - \alpha_K)^2 + 2\omega_3\alpha(\alpha - \alpha_K).$$

The coefficients ω_0, ω_1, ω_2 and ω_3 are chosen so that

$$\begin{aligned} A(0) &= I(\mathbf{z}, 0) & \qquad A(\alpha_K) &= I(\mathbf{z}, \alpha_K) \\ A'(0) &= \frac{\partial I}{\partial \alpha}(\mathbf{z}, 0) & \qquad A'(\alpha_K) &= \frac{\partial I}{\partial \alpha}(\mathbf{z}, \alpha_K) \end{aligned}$$

From our Taylor series approximations near $\alpha = 0$, we concluded

$$I(\mathbf{z}, 0) = G(\mathbf{z}) \qquad \frac{\partial I}{\partial \alpha}(\mathbf{z}, 0) = 0.$$

The value of $I(\mathbf{z}, \alpha_K)$ we have from our approximations, while we use

$$\frac{\partial I}{\partial \alpha}(\mathbf{z}, \alpha_K) \approx \frac{I(\mathbf{z}, \alpha_{K+1}) - I(\mathbf{z}, \alpha_K)}{\epsilon}.$$

Now

$$\begin{aligned} A(0) &= \omega_0 - \alpha_K\omega_1 + \alpha_K^2\omega_2 & \qquad A(\alpha_K) &= \omega_0 \\ A'(0) &= \omega_1 - 2\alpha_K\omega_2 + \alpha_K^2\omega_3 & \qquad A'(\alpha_K) &= \omega_1 \end{aligned}$$

so we set

$$\begin{aligned} \omega_0 &= I(\mathbf{z}, \alpha_K) \\ \omega_1 &= \tfrac{1}{\epsilon}\left[I(\mathbf{z}, \alpha_{K+1}) - I(\mathbf{z}, \alpha_K)\right] \\ \omega_2 &= \tfrac{1}{\alpha_K^2}\left[G(\mathbf{z}) - \omega_0 + \alpha_K\omega_1\right] \\ \omega_3 &= \tfrac{1}{\alpha_K^2}\left[2\alpha_K\omega_2 - \omega_1\right] \end{aligned}$$

This is the method that is used in the code to evaluate the integrals $I(\mathbf{z}, \alpha)$.

3.3 Reprise of mathematical assumptions and computational techniques

Assumption	Comments	Location
The offender chooses offense locations according to a probability distribution $P(\mathbf{x}\,\|\,\mathbf{z},\alpha)$ that depends only on the offender's anchor point $\mathbf{z}$ and the average distance α the offender is willing to travel to offend.	This is fundamental to our approach, though the mathematical framework allows for the possibility that P depends on other variables.	page 17
The offender's choices of crime sites are mathematically independent of one another	This is a required mathematical simplification for (3.4); without it one must also model the nature and form of the effect of the choice of one crime site has on the offender's selection of subsequent crime sites.	page 19
The offender's distance decay function is independent of the offender's anchor point	This is a required mathematical simplification for (3.2); without it one must also model the nature and form of the dependence of the average offense distance on the location of the offender's anchor point.	page 19
The probability density P has the form $P(\mathbf{x}\,\|\,\mathbf{z},\alpha) = D(d(\mathbf{x},\mathbf{z}),\alpha)G(\mathbf{x})N(\mathbf{z},\alpha)$	Other reasonable forms for this probability density exist and can be treated via these mathematical techniques; this is the selection that was implemented in the software prototype.	page 20
The offender's two-dimensional distance decay function is a normal distribution $D(\mathbf{x}\,\|\,\mathbf{z},\alpha) = \frac{1}{4\alpha^2}\exp\left(-\frac{\pi}{4\alpha^2}\|\mathbf{x}-\mathbf{z}\|^2\right)$	This is the choice made in the software prototype. Other reasonable choices exist, like the negative exponential.	pages 21, 26
The likelihood that a particular location will be the location of a future offense can be estimated from the locations of historical crimes of the same type as the offense.	In the prototype software, the analyst needs to select the historically similar crimes.	page 22
The distribution of anchor points follows local population density for people with the same age / sex / race or ethnic group of the offender	The prototype software uses U.S. Census data to perform the calculations.	page 23
The prior distribution of average offense distances $\pi(\alpha)$ can be calculated by using historical data and solving (3.15)	This is difficult to solve in practice; see computational techniques.	page 28

Table 3.2: Summary of mathematical assumptions in the model

Technique	Location
Equation (3.15) which is used to determine the estimate for the distribution of average offense distances across offenders is solved by using collocation to convert it to a linear system, which is then solved via Tikhnonv regularization.	Subsection 3.2.3
The continuous functions in the model are estimated by calculating their values on a triangular grid that overlays the geographic region under consideration. The grid is kept at two levels of detail, a coarse grid and a fine sub-grid.	Subsections 3.2.4 and 3.2.5
The integral defining the normalization function $N(\mathbf{z}, \alpha)$ is evaluated at each point in the coarse grid using the midpoint method, where the precise mesh used is an adaptive combination of elements from the coarse mesh and the fine mesh depending on the value of α. These values are interpolated into the fine mesh via geometric interpolation and Hermite approximation in α.	Subsection 3.2.6

Table 3.3: Summary of computational techniques in the software

3.4 The software

3.4.1 Overview

We have developed software that implements these mathematical algorithms. The software was built in two parts:

- A program called Profiler that performs all of the mathematical analysis, and
- A program called ProfilerGUI with which the user interacts.

The software was developed in this fashion for a number of reasons. First and foremost, this approach allows this scientific work to be kept separate from user interface issues. In particular, because the scientific analysis is performed in a separate, stand-alone program, it is possible for this tool to be incorporated into other software suites that have compatible open source licenses. In addition, this approach lets us work separately on the user interface and the scientific analysis tools.

To use the software, the user starts ProfilerGUI; a screen shot is contained in Figure 3.24. The user interface then prompts the user to enter the required data for the program. When this is complete, ProfilerGUI writes a plain text parameter file that contains all of the user's selections. ProfilerGUI then calls the program Profiler, and passes it the name of the parameter file it just wrote. Profiler performs its analysis; as it returns information this is noted by ProfilerGUI and reported to the user. It total, the program provides

- A map of the proposed search area, in .kml format,
- A map of the target attractiveness function $G(\mathbf{x})$, in .kml format, and
- A map of the population density of the search area that matches the provided demographic information, again in .kml format.

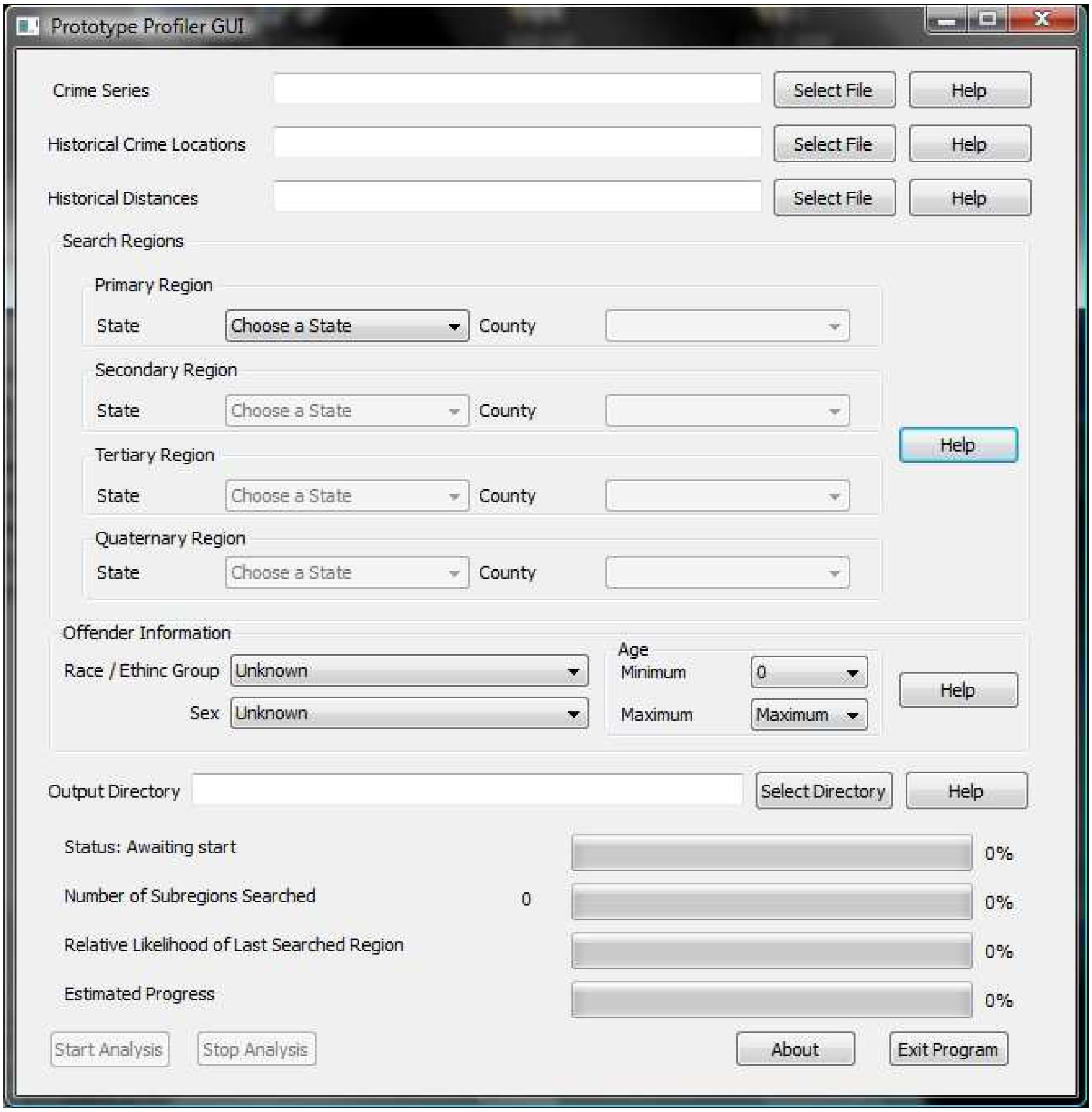


Figure 3.24: Screenshot of the Program

Because the analysis process can take some time- a few hours is typical, the program also provides the analyst with its partial results as the analysis continues.

The output maps are provided in .kml format; this is an open standard. Files in this format can be rendered by a number of free programs, including

- Google Earth,
- ESRI's ArcGIS Explorer, and
- NASA's World Wind.

All of the maps in this report are screen shots taken from Google Earth.

In this section, we will describe first how the program is installed, including how to obtain all of the required files. We will then provide step-by-step instructions on how the program is meant to be used. Finally, we will take a brief look at the source code for the two software tools, and explain how they have been designed and built.

3.4.2 Installing the software

The software tool comes packaged as a compressed .zip archive. Once downloaded, the user simply needs to uncompress the archive to create a folder called Profiler; see Figure 3.25 as an example.

At this point, the tool is ready to run; just use the ProfilerGUI program provided in the root directory. No other installation tasks need to be performed.

3.4.3 Using the software

Detailed instructions on the use of the program have been provided with the software in the form of a .pdf sideshow that illustrates the use of the program on a series of convenience store robberies in Baltimore County. In this report, we shall briefly summarize the process.

Selecting the crime series

The first step is to select the crime series under analysis. The program requires a plain text file that contains the locations of the crime series. Each line of this file contains the longitude and latitude of one element of the crime series, separated by one or more spaces. Both longitude and latitude are specified in decimal degrees, and need to be separated by one or more spaces. The file name can be entered either by directly typing its name into the Crime Series box or by pressing the corresponding Select File button to obtain a dialog box that allows the user to simply click on the desired file.

Selecting the historical crime locations

Next, the analyst needs to provide a plain text file that contains the locations of a representative set of crimes of the same type as the series under consideration. This file has the same structure as the crime series file, save that each line now contains the longitude and latitude of a single historical crime.

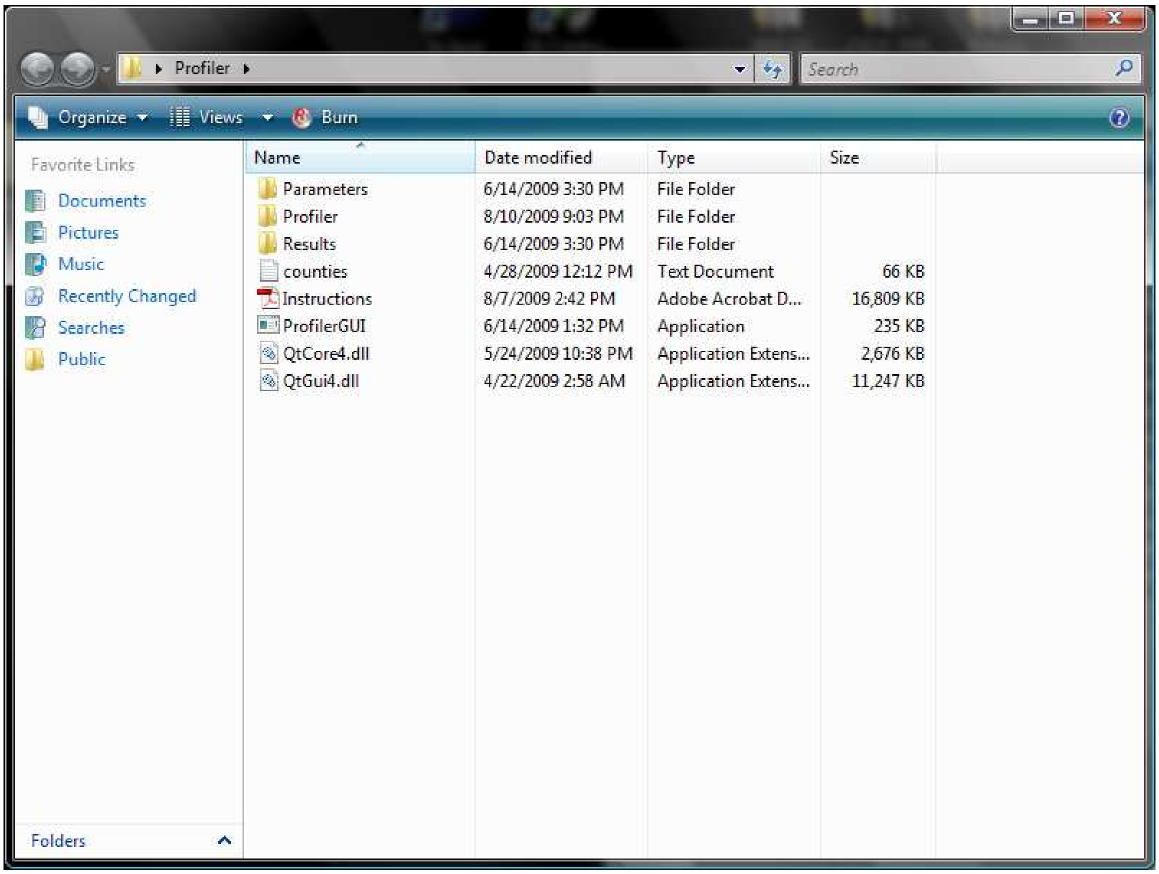


Figure 3.25: Unpacking the software archive

This is the information that the algorithm uses to generate its estimate for $G(\mathbf{x})$. The crimes used in the historical crime locations may be solved or unsolved, and need not be from series crimes. Not every crime of the same type needs to be included, just enough to generate a representative sample. As part of the analysis, the program will generate a map of the resulting target attractiveness $G(\mathbf{x})$; it is important that this map be checked for reasonableness.

In particular, the bandwidth used to generate the map of historical crimes is twice the mean nearest neighbor distance between offenses. Thus, if too few historical crime sites are included, then the mean nearest neighbor distance will be large, and so the graph will be very broad and smooth- perhaps broader and smoother than the analyst feels is appropriate. As a consequence, when deciding what historical crimes to use, the analyst needs to balance how well the historical crimes match the series versus the need to have a representative sample that generates a reasonable map of historical crimes.

It is also important to note that regions where the historical crime locations map is zero are considered to be regions where there can be no known crimes in the series known to the analyst. A consequence of this is that if an analyst is investigating a series that crosses multiple jurisdictions, then the analyst needs to have the corresponding historical information for all of these jurisdictions.

Selecting the historical distances

To create the estimate of the prior distribution of offender average offense distances, the program requires a plain text file that contains the locations of a representative set of solved crimes together with the location of the offender's home base. This file has the same structure as the crime series file, save that each line now contains the longitude and latitude of a crime site together with the longitude and latitude of the offender's home base.

The solved crimes used to generate the historical distances do not need to be of exactly the same type as the series under consideration. They only need to provide a reasonable basis to estimate the distance decay function for offenders in the jurisdiction.

Selecting the search region

The search region is composed of one or more county-sized regions. The list of available county-sized regions is taken from the U.S. Census. The analyst simply needs to select the state and corresponding county sized region(s) from the available drop boxes. Up to four different regions can be selected.

When selecting the search region, it is important to note that the algorithm will assume that the offender is not located outside of the search region. In particular, the search area chosen at this stage must be sufficiently large to encompass the anchor point of any potential offender. Failure to do so will result in search areas that are strongly biased.

To perform its analysis, the program needs to possess the necessary U.S. Census data for the selected regions. When a region is selected, the program will check to see if it has the necessary data files. If it does not, then a message will be displayed, explaining the issue. The box tells the user the name of the files that are needed, provides a link to the Census web site from which the files can be downloaded, and indicates the locations where the downloaded data files need to be stored. It is impractical to distribute all of the required data files directly with the program due to their large size. Indeed, if we were to include the necessary census data for all fifty states, the result would be many hundreds of times the size of the program.

Selecting the offender information

If the age, sex, or race / ethnic group of the offender is known to the analyst, this information can be included. When the program develops its prior estimate for the distribution of potential offender anchor points $H(\mathbf{z})$ before the information from the crime series is taken into account, it uses population density data from the U.S. Census for this purpose. Because the Census provides block level data subdivided by age, sex, and race /ethnic group, we can use that information when developing this prior estimate. Entering the available data is done by simply choosing from the appropriate drop boxes; the default assumption is that no demographic information is available about the offender.

Program results

As we have already noted, the program produces a number of maps as its output, including a map of the search area, a map of the target attractiveness function, a map of the population density, and interim maps of the search area as the analysis proceeds. All of these results are stored in the directory that the user specifies here, either by directly entering the name of the directory, or by using the dialog box obtained by pressing the Select Directory button.

The analysis

Once all of the necessary data has been entered, the Start Analysis button will be enabled; once pressed the program will begin its work. First the program will read the necessary data files and initialize internal data structures; as this process occurs, the status bar will keep the user informed. When the initialization is complete, the program will begin by working one subregion at a time and calculating the probability that the offender's anchor point lies in that subregion. The total number of subregions searched is recorded as a fraction of the total number of subregions in the search area, as well as the relative likelihood of the last searched subregion as a fraction of the maximum likelihood so far encountered.

Now an exhaustive search of every possible subregion would significantly increase computation time to no real benefit- given that the search area is the size of one or more counties, most of the subregions are unlikely to contain the offender's anchor point. Instead, the search proceeds by spiraling out from the crime series, and continuing until it can make a complete circuit where the the likelihood that any of the subregions in the circuit contains the offender's anchor point is less than 0.5% of the most likely subregion. The Estimated Progress bar tracks how close the program has come to making a complete circuit where all of the subregion likelihoods have remained below this threshold. If during this process the program encounters a subregion where the likelihood is larger than this threshold, then it must begin a new circuit at this point and so the Estimated Progress bar will drop back to 0%.

Running time

The software prototype is computationally intensive, and the process can take quite some time to complete. In Chapter 4, we apply the software prototype to a series of 6 convenience store robberies in Baltimore County. When the prototype is run on a new Windows Vista system (Core 2 Duo T6400, 4 GB memory) for this series, the program needs roughly an hour and a half to complete. The same series when analyzed on an older Windows XP system (Pentium D 805, 2GB memory) has a run time of roughly two hours; older systems have even longer run times.

Despite the long run times for the software prototype, it is designed to be run while other applications also run on the system without using all of the computer's resources. Thus, the machine should not "freeze" or become unresponsive while the prototype is running, and other applications can be used concurrently with the prototype.

3.4.4 Internal structure of the software

Profiler

Profiler is a program written entirely in C++ to implement our mathematical algorithms. It is a console based program that is usually run in concert with ProfilerGUI, but can also be run separately as a stand-along program.

Using Profiler as a stand-alone program To run Profiler as a stand-alone program, it needs to be run from the command line with the name of a plain text file that contains the parameters necessary for the analysis; if no parameter is passed the program looks for the parameter file with name `./Parameters/Parameters.txt`.

A typical parameter file looks like the following

```
Triangle Circumradius = 0.01
Crime Series Data File Name = ./Parameters/series.txt
Historical Data File Name = ./Parameters/history.txt
Historical Distances File Name = ./Parameters/historical_distances.txt
Number of regions = 2
State = MD
County Code = 005
State = MD
County Code = 510
Race / Ethnic group = Asian alone
Sex = Female
Minimum Age = 0
Maximum Age = Maximum
Results Directory = ./Results
```

An explanation of the individual parameters follows

- **Triangle Circumradius**: The geographic region under study is subdivided into a mesh of equilateral triangles (the coarse mesh), and each of these triangles is then subdivided into a number of subtriangles, resulting in the fine mesh. The size of the triangles in the coarse mesh is the input variable Triangle Circumradius.

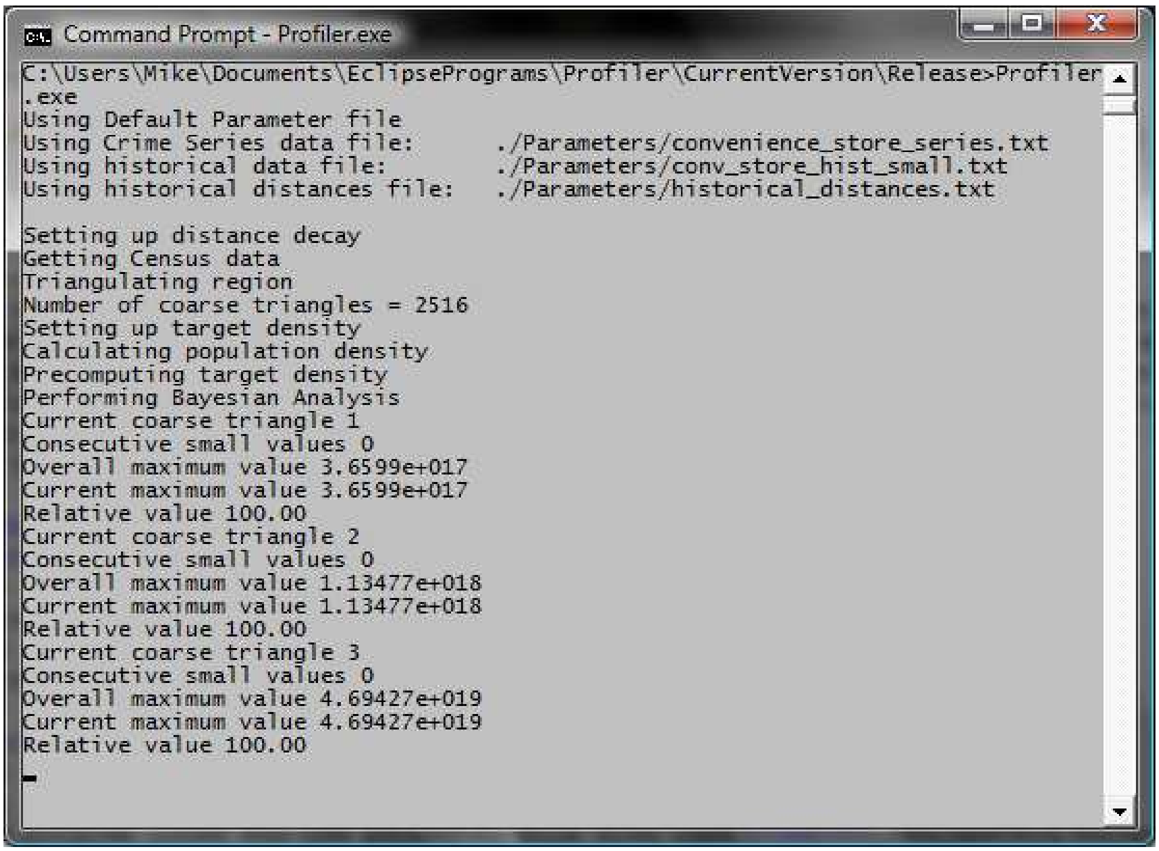


Figure 3.26: Screenshot of Profiler running as a stand-alone program

- **Crime Series Data File Name**: This is the name of a plain text file that contains the longitude and latitude of each element of the crime series.

  1. Each line of the file should contain the latitude and longitude of a single crime location.
  2. The latitude and longitude should be separated by one or more blank spaces, with longitude first, and latitude last.
  3. The latitude and longitude should be specified in decimal degrees.
  4. The file should contain no other data; in particular it should not end with a blank line.

- **Historical Data File Name**: This is the name of a plain text file that contains the longitude and latitude of historical crimes of the same type as the series under study.

  1. The historical data is used to generate a map of the relative likelihood that a particular location is the site of a crime of the same type as the that series under study. The data set should be sufficiently large for this purpose.
  2. The historical data does not need to consist of solved crimes, nor does the historical data need to consist of series crimes.
  3. Each line of the file should contain the latitude and longitude of a single crime location.
  4. The latitude and longitude should be separated by one or more blank spaces, with longitude first, and latitude last.
  5. The latitude and longitude should be specified in decimal degrees.
  6. The file should contain no other data; in particular it should not end with a blank line.

- **Historical Distances File Name**: This is the name of a plain text file that contains the longitude and latitude of both the locations of a crime and the home base of the offender.

  1. The historical data is used to generate a graph of crime frequency versus distance.
  2. The historical data should be from crimes similar to the series under consideration.
  3. Each line should contain the data from a single historical crime, and should contain in order the longitude and latitude of the crime site, then the longitude and latitude of the home base of the offender.
  4. The latitudes and longitudes should be separated from each other by one or more blank spaces.
  5. The latitudes and longitudes should be specified in decimal degrees.
  6. The file should contain no other data; in particular it should not end with a blank line.

- **Number of regions**: This is the number of county-equivalent regions that are necessary to contain both the all of the elements of the crime series as well as the largest possible search area

- **State**: Two letter abbreviation for a state containing a county-sized region to be searched.

- **County Code**: This is the corresponding three digit U.S. Census Bureau County Code[1]; this does not include the associated two digit state code.
- **Race / Ethnic group**: This is the Race or Ethnic group of the offender; the classification scheme is the same as that which is used by the U.S. Census Bureau in Summary File 1[2].
- **Sex** This is the sex of the offender.
- **Minimum Age** This is a lower estimate for the possible age of the offender. Not every choice of age is valid; the value needs to be the lower bound for an age range for which the U.S. Census records block level population data; see Summary File 1.
- **Maximum Age** This is a upper estimate for the possible age of the offender. Not every choice of age is valid; the value needs to be the upper bound for an age range for which the U.S. Census records block level population data.
- **Results Directory**: This is the name of the directory that the program will use to store its results.

Compiling Profiler This program was compiled with gcc(3.4.5)/MinGW; it also requires the Lapack++ libraries (2.5.1). It is important to use the a recent version of Lapack++. There is a much older version (1.1a) of Lapack++ available at the NIST web site; it is not suitable.

Brief overview of the Profiler source code Profiler was developed using a strict object-oriented approach; its functions were split into twelve distinct classes:

- Input and output classes
  - `CInput` is used to handle all of the input to the program; it reads the parameter file, as well as any auxiliary data files and performs all validation.
  - `COutput` is used to generate all of the maps developed by the code in `.kml` format.
  - `CCensus` is used to read data from U.S. Census data files.
- Auxiliary classes for data
  - `CBlock` is an auxiliary class that encapsulates a U.S. census block.
  - `CPt` is an auxiliary class that encapsulates a geographic point.
  - `CTriangle` is an auxiliary class that represents a triangular geographic region.
- Classes for geography
  - `CTriangulatedGeography` is the main class that handles the geography. The geographic region under study is covered by two triangular meshes, a coarse mesh that is used for functions that vary slowly and are difficult to compute- like $N(\mathbf{z},\alpha)$, and a much finer mesh for functions that vary much more rapidly- like $G(\mathbf{x})$ and $H(\mathbf{z})$, which can change dramatically from block to block.

[1] `http://www.census.gov/datamap/fipslist/AllSt.txt`
[2] `http://www.census.gov/Press-Release/www/2001/sumfile1.html`

- Functions in the mathematical model
  - `CDistanceDecay` is the class that contains the data for the distance decay function D; it also contains the results for the prior distribution of offender average offense distance $\pi(\alpha)$.
  - `CNormalizationFunction` is the class that calculates the values of the normalization function $N(\mathbf{z}, \alpha)$.
  - `CTargetDensity` is the class that contains the values for the target density function $G(\mathbf{x})$.
  - `CPopulation` is the class that contains the data for $H(\mathbf{z})$.
  - `CTikhonov` is the class that is used to perform the regularization process, it is an auxiliary class for `CDistanceDecay`, and reports its results back to that class for storage.

Now we briefly describe each class used by Profiler.

CBlock This class is a data structure to hold the elements of a U.S. Census Bureau block that are relevant for the code. This data is all available in Summary File 1. In particular, the class records

- The U.S. Census Bureau logical record number of the block,
- The land area of the block, in square meters,
- The latitude and longitude of the block (in degrees), and the
- Population of the block.

The population recorded in `CBlock` is not necessarily the total population of the block, but rather the population for the combination of age, sex, and race / ethnic group under consideration.

CCensus This class that contains all of the census data for the county or counties that contain the potential search area.

The Census Bureau Summary file 1 has block level population data for each county in the country, sorted by age, sex, and race / ethnic group. For each state, we require three files:

- STgeo.uf1
- ST00005.uf1
- ST00006.uf1

where ST is replaced by the corresponding two letter state abbreviation. These files must be included in the same directory as the program executable; they can be directly downloaded from the U.S. Census site.

The demographic information- age, sex, and race/ethnic group follow the Census Bureau standard definitions and groupings.

Internally, the class stores an array of blocks (type `CBlock`) which contain the location, land area, and population (in the specified demographic) of each block.

This data is be used by `CPopulation` to construct the prior distribution for anchor points $H(z)$.

CDistanceDecay This class handles both the explicit form of the distance decay function as well as the creation of the prior estimate for the distance decay function.

All distances are measured in decimal degrees. The motivation for this choice is the fact that since the input latitude and longitude will be in degrees, it makes more sense to keep the same units throughout.

The implementation of the distance between two points is actually handled by the `CPt::S2DistanceTo` function, and work on the assumption that the Earth's surface is spherical. In particular, no correction factors for the ellipsoidal nature of Earth have (yet) been incorporated.

The primary mathematical quantities that this class needs to store are the arrays

- α_k, the values of α at which the prior π is recorded,
- $\pi_k \approx \pi(\alpha_k)$, the values of the prior estimate for the distance decay function, and
- $\Delta\alpha_k \equiv \alpha_{k+1} - \alpha_k$, the increments in α

To evaluate the later integrals, only the nonzero values of π_k are needed; hence the class stores the values of the triples $(\alpha_k, \pi_k, \Delta\alpha_k)$ for only the non-zero values of π_k.

This data structure also allows for the possibility of adaptive integration in α by adjusting the step sizes $\Delta\alpha_k$, but right now all of these values are constant.

CInput This is the class used to set up and handle all of the input and output files for the program. The constructor is passed the program's arguments; it uses these to determine the location of the parameter file, which it then opens. The constructor then reads the parameter file, and stores all of the necessary results in class variables.

The class provides a family of functions that provide to access the parameters, which are stored in class variables.

CNormalizationFunction This class handles all of the calculations necessary to evaluate $N(\mathbf{z}, \alpha)$.

It has one public member function, which is used to actually calculate the integral. It starts with a value $\mathbf{z}$ and an array $\{\alpha_1, \alpha_2, \ldots, \alpha_K\}$, then returns the array of values

$$\{N^{-1}(\mathbf{z}, \alpha_1), N^{-1}(\mathbf{z}, \alpha_2), \ldots, N^{-1}(\mathbf{z}, \alpha_K)\},$$

where $N^{-1}(\mathbf{z}, \alpha)$ is defined by

$$N^{-1}(\mathbf{z}, \alpha) = \iint_{\mathbf{R}^2} D(d(\mathbf{x}, \mathbf{z}))G(x)\, dx^{(1)}\, dx^{(2)}$$

This is done by adaptive integration and uses the private function `CNormalizationFunction::GetOneMeshValue` The private function chooses a triangulation of the region of integration that depends on the values of α_i and then performs the integration.

COutput This class is used to generate and write all of the output maps produced by the analysis engine. To create an instance, you must specify the output type and the file name of the output. Right now, only .kml output formats are supported. The constructor creates the file with the provided name in the directory recorded in the `COutput::resultsDirectory`. The constructor also writes the required elements of the .kml header.

Function calls can then be made to `COutput::DrawPoint` and `COutput::DrawGraph` so that data can be added to the output. Multiple maps can be overlaid in the same output file; the altitude of each map is stored in the class variable `COutput::altitude`, and this value is increased by 50 each time `COutput::DrawGraph` is called.

The destructor ∼`COutput` writes the KML footer and closes the file.

CPopulation This class stores the population data for the search region. This is used as a proxy for an estimate of the prior distribution of offender anchor points.

Right now, this is calculated only from US Census data; in the future we would like to use other data sets (e.g. distributions of other offenders) to generate this distribution. This can be done most simply by overloading the constructor.

CPt This class forms a data structure for a two-dimensional point. Implemented operations include the standard vector operations (addition, scalar multiplication), the (Euclidean) dot product and Spherical distance (`CPt::S2DistanceTo`).

CTargetDensity This class is used as a data structure to hold the target density function $G(\mathbf{x})$. In particular, this class calculates and stores the value of $G(\mathbf{x})$ at the centers of the fine triangles in the mesh; these values can be retrieved via `CTargetDensity::getFineValue`.

The class also calculates and stores the value of $G(\mathbf{x})$ for each coarse triangle. Rather than simply calculate the value of $G(\mathbf{x})$ at the center point of the coarse triangle, it instead uses the averages of the values of G at all of the fine triangle contained within the coarse triangle. The reason for this behavior is that the underlying function G varies on a very fine scale. It makes sense that to calculate the value of G on the fine mesh, because that is the finest mesh in the problem. When looking for the value of G at for a coarse triangle, if we simply calculated the value of G at the center of the triangle, then we run the risk that a small change in the location of the center of the coarse triangle can result in a large change in the value returned. Instead, we return the average value to try to smooth our the results.

The values of $G(\mathbf{x})$ are calculated using a kernel density estimation process. In particular, suppose that the historical crimes have been committed at the locations $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N$. Consider the truncated quartic kernel function $K(\mathbf{x}|\lambda)$ with bandwidth λ

$$K(\mathbf{x}|\lambda) = \begin{cases} \dfrac{3}{\pi\lambda^6}(|\mathbf{x}|^2 - \lambda^2)^2 & \text{if } |\mathbf{x}| \le \lambda \\ 0 & \text{if } |\mathbf{x}| \ge \lambda \end{cases}$$

The bandwidth λ is chosen to be twice the mean nearest neighbor distance between historical crimes, and

$$G(\mathbf{x}) \propto \sum_{i=1}^{N} K(\mathbf{x} - \mathbf{c}_i|\lambda).$$

CTikhonov This class is used solely to compartmentalize the process of calculating the non-negative Tikhonov regularized solution to a linear equation. It uses the mathematical algorithms described in Section 3.2.3.

CTriangle This is a data structure for a triangle with sides parallel to the x-axis. It retains the center of the triangle, its circumradius, and its orientation, where the orientation determines if the centroid is above or below the side parallel to the x-axis.

CTriangulatedGeography This is a data structure that contains all of the underlying geographic structure of the program. The class contains two primary data structures- a mesh of coarse triangles and a mesh of fine triangles. Each of the triangles in the coarse mesh are of type `CTriangle`, meaning that they are equilateral and have one side parallel to the x-axis. This coarse grid covers the entire geographic region under consideration, including the crimes, the historical crimes, and the region to be searched.

Each coarse triangle is subdivided into subtriangles, also of class `CTriangle`.

The coarse grid and the fine grid are entirely created by the constructor; the primary public methods return various grids, sub-grids, and individual triangles.

ProfilerGUI

ProfilerGUI is a program written in C++ to provide a simple interface to allow an analyst to interact with the Profiler program. ProfilerGUI requires the presence of the program Profiler in a subdirectory of the same name. When ProfilerGUI runs, it takes the data provided by the user, writes a parameter file, calls Profiler, and then interprets the results returned by Profiler.

Compiling ProfilerGUI This program was developed with the GPL version of Qt Creator 1.1.1, based on Qt 4.5.1. It requires no additional libraries beyond those included in Qt Creator.

Brief overview of the ProfilerGUI source code This program is designed to take full advantage of the signals and slots mechanisms provided by Qt. In particular, this allows us to create an event-driven program where actions taken in the user interface (*e.g.* pressing a button or editing a box) cause code to be executed.

The program contains only two classes- `MainDlg` and `CStatus`, and nearly all of the work occurs in `MainDlg`. This is the class that handles the program's primary dialog. It is responsible for

- Displaying the dialog (and all sub-dialogs, including file/directory selection and help dialogs)
- Recording all of the entered data
- Starting the profiler (analysis) program
- Reading the data returned by the profiler program, and updating the user on the status of the analysis.

The class `CStatus` is simply a data structure containing five boolean variables where each variable represents the status of one of the required data elements-

- The crime series,
- The history file,
- The historical distances file,
- The search region, and
- The results directory.

When true they indicate that the GUI program has the necessary data for that data element.

The public function `CStatus::Prepared` returns true if all of the status variables are true; otherwise it returns false. It is used to determine if the program is ready to allows the analysis component to run.

Integrating Profiler and ProfilerGUI When complete, the software package contains

- ProfilerGUI.exe
- The file counties.txt, which is a modified version of a list of county codes provided by the U.S. Census.
- Two .dll files from the Qt library:
  - QtCore4.dll
  - QtGui4.dll
- The analysis program (Profiler.exe) and its associated files in the Profiler subdirectory. In particular, the subdirectory needs to contain
  - Profiler.exe
  - Three lapack++ library files:
    - libblas32.dll
    - liblapack32.dll
    - liblapackpp-15.dll

Chapter 4

Results

4.1 Discussion of findings

The purpose of our project was to develop a new mathematical approach for the geographic profiling problem and to implement that new algorithm in software, and we have done so. To illustrate how our model and tools work in practice, let us consider a series of convenience store robberies that occurred in Baltimore County in May 2008.

Basic information about the crime series is provided in Table 4.1, and the crime locations are mapped in Figure 4.1. Note that crimes 1, 4 and 6 all occurred at the same location marked as Crime Site # 6 on the map. To give this information to our program, we simply provide it with a plain text file containing the latitude and longitude of each crime site.

As we begin our analysis, we start with the elementary observation that not every location can be the site of a convenience store robbery- after all not every location is the site of a convenience store. Instead, experience tells us that convenience stores tend to be concentrated at or near major intersections. Moreover, some convenience stores are more likely to be the site of a robbery than others. To account for these geographic facts, our algorithm requires the locations of a representative sample of the locations of convenience store robberies throughout the jurisdiction. A simple graph of the locations of convenience store robberies in Baltimore County is provided in Figure 4.2. Even though this figure does not include the locations of the major roadways, the target pattern clearly shows where many of them lie.

Date	Time	Location		Target
		Latitude	Longitude	
March 8	12:30 pm	-76.71350	39.29850	Speedy Mart
March 19	4:30 pm	-76.74986	39.31342	Exxon
March 21	4:00 pm	-76.76204	39.34100	Exxon
March 27	2:30 pm	-76.71350	39.29850	Speedy Mart
April 15	4:00 pm	-76.73719	39.31742	Citgo
April 28	5:00 pm	-76.71350	39.29850	Speedy Mart

Table 4.1: A series of convenience store robberies in Baltimore County

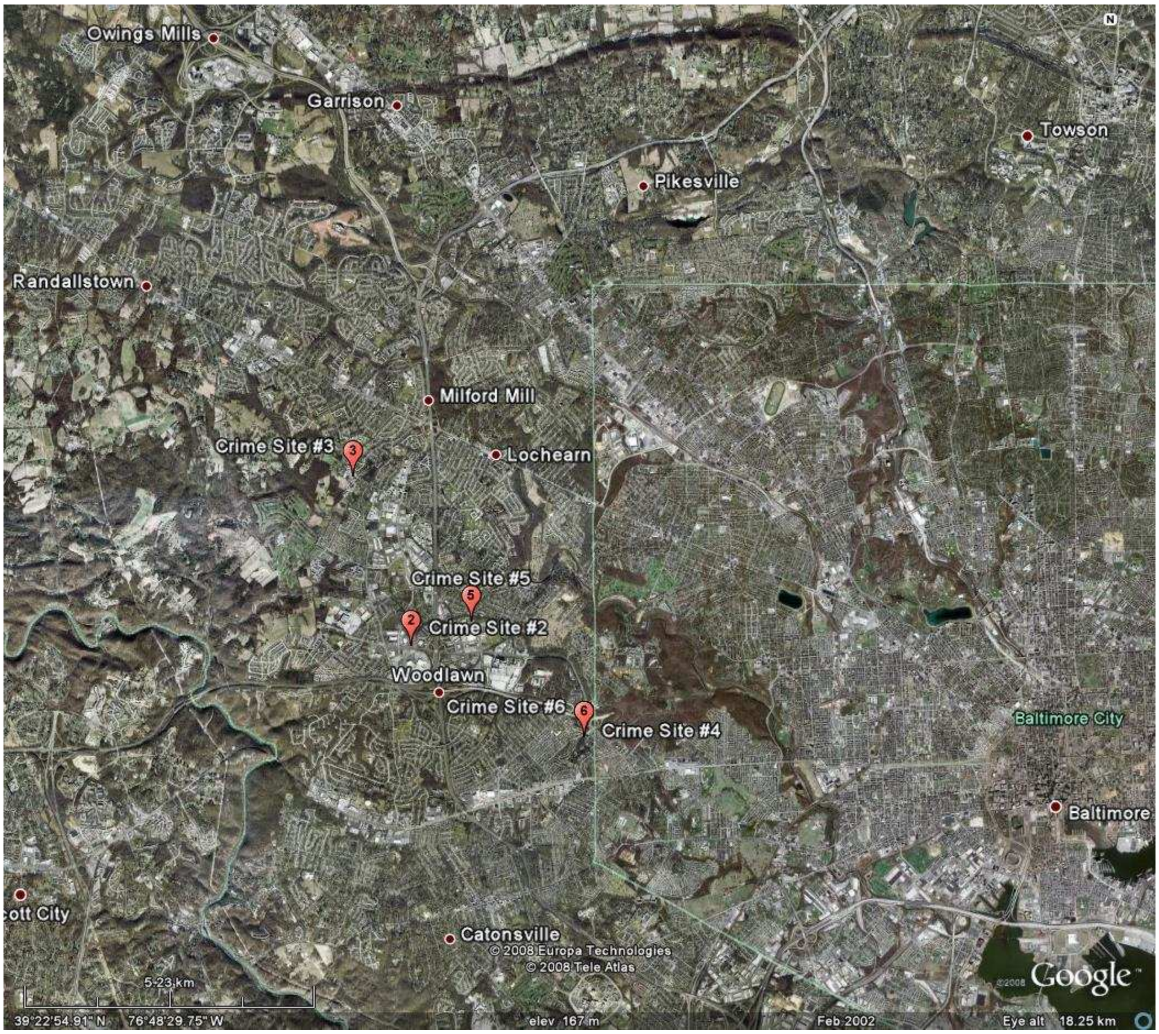


Figure 4.1: Convenience store robbery series in Baltimore County

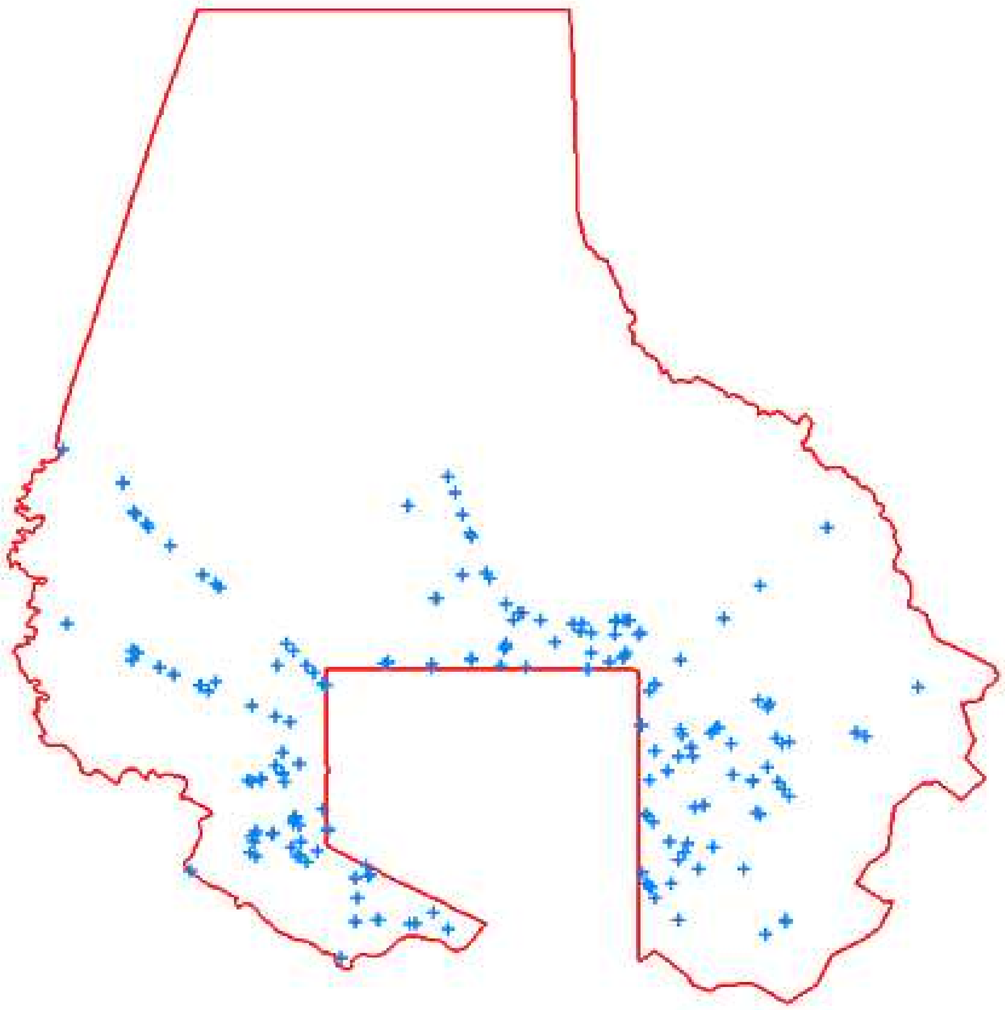


Figure 4.2: Convenience store robberies in Baltimore County

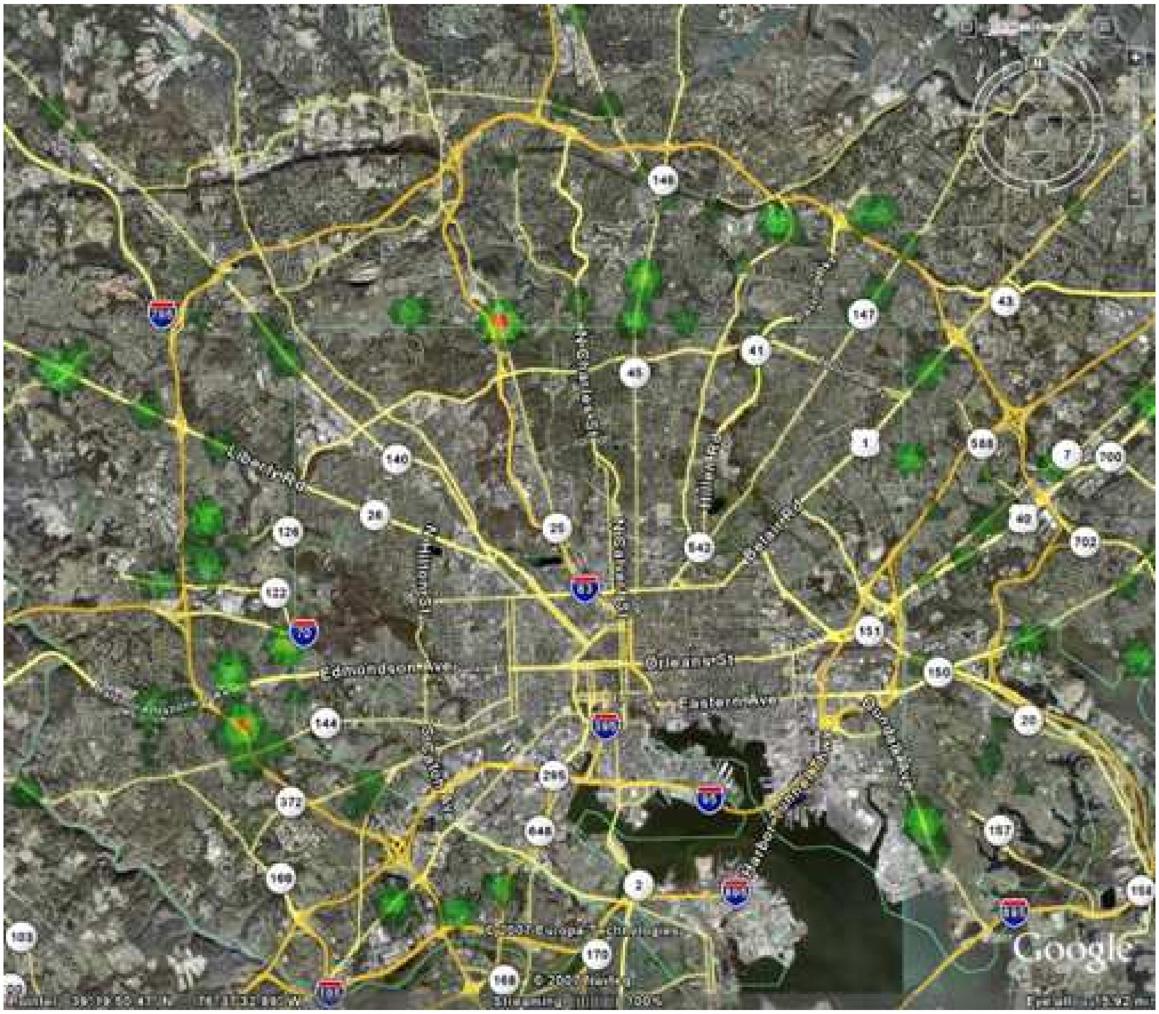


Figure 4.3: Target attractiveness for convenience store robberies in Baltimore County

Another important geographic factor in the analysis of this series is the presence of the jurisdictional boundary. Baltimore County and Baltimore City are separate jurisdictions, and the elements of the crime series under consideration lie very close to this boundary. This particular series was identified by Baltimore County, and though there are no known elements of the series in the city, this cannot be ruled out.

Thus, we provide the program with another plain text file, this one containing the locations of 449 convenience store robberies within the county. The program then calculates the resulting target attractiveness function $G(\mathbf{x})$, and returns it to the analyst; this map is shown in Figure 4.3

Our goal is to obtain an estimate of the location of the offender's anchor point; if we assume that the anchor point is a residence, then we immediately know that there are locations where the anchor point cannot be located. For example, looking at the crime series map (Figure 4.1), we see that east of the crime series locations is a large park, near the crime sites is a large interstate overpass and a commercial area including offices and a number of malls, and to the west is a wooded area. Since none of these areas contain residences, none of these areas are likely to contains the offender's

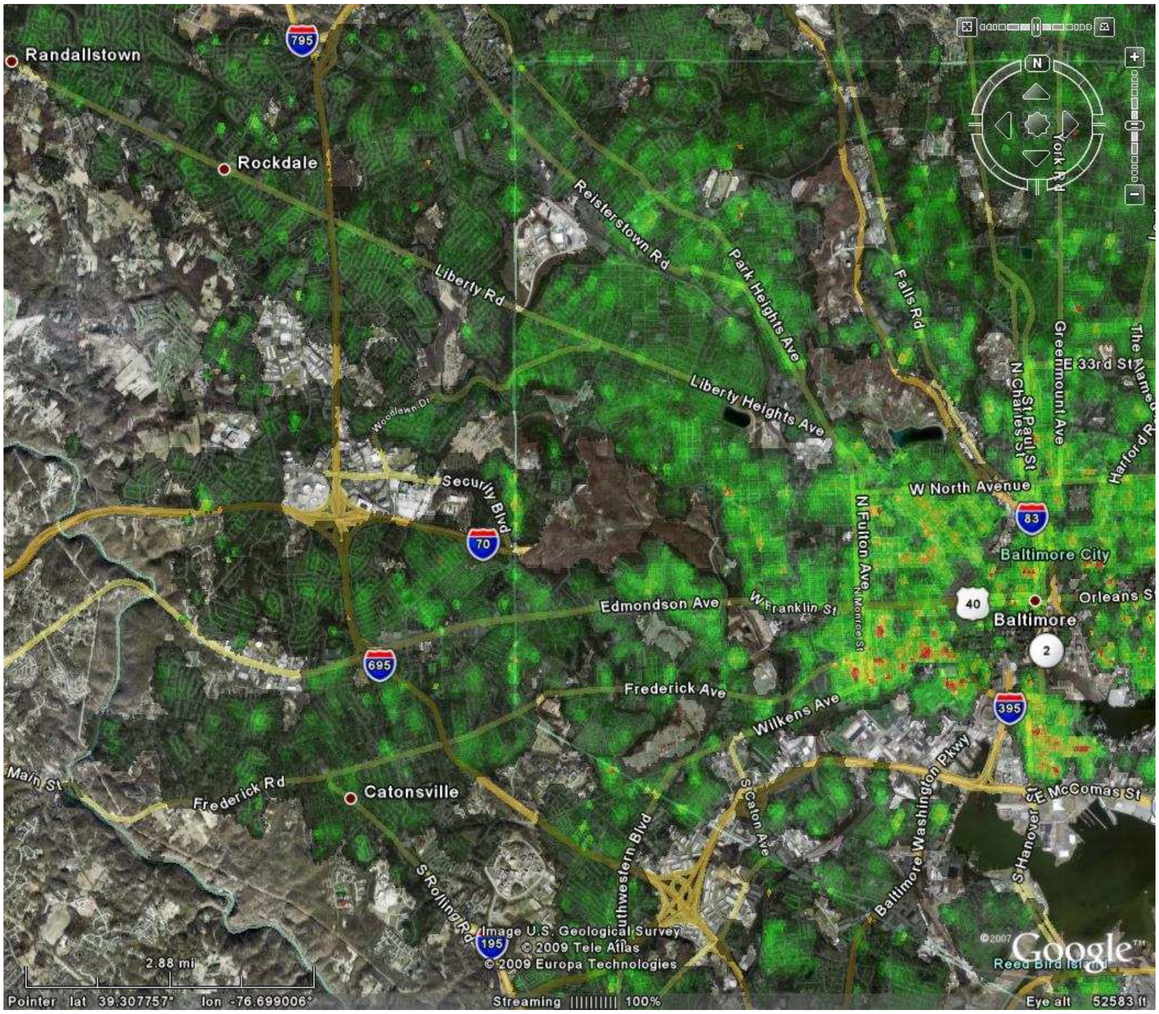
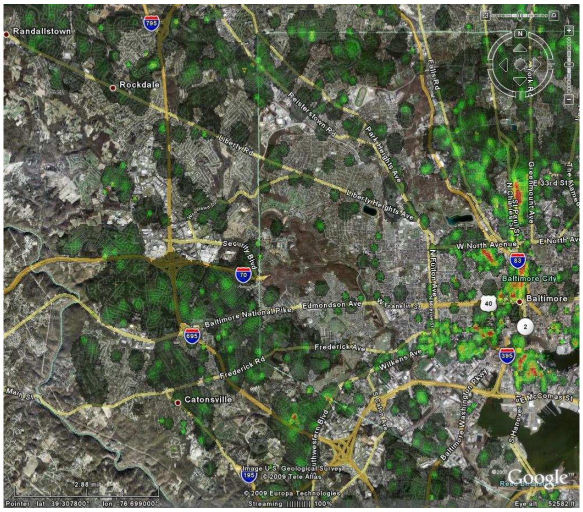


Figure 4.4: Population density and proposed anchor point density near the crime series

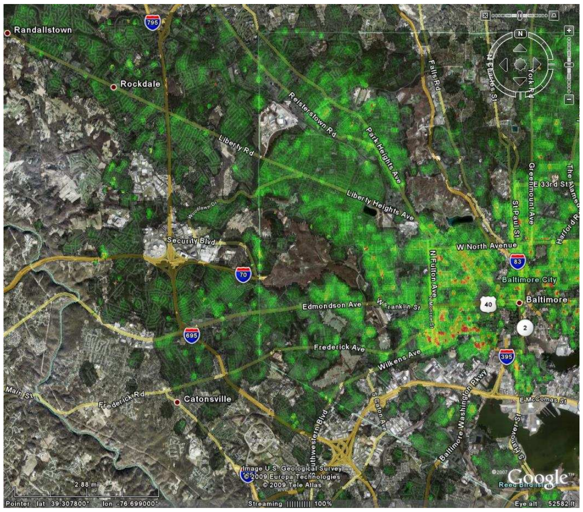
residence. To handle this geographic information, our method uses local population density from U.S. Census data as a proxy for anchor point density. In particular, our existing software is able to read the raw data from the census and use a modified kernel density parameter estimation technique to generate a map of potential anchor points; it is shown in Figure 4.4.

One advantage of using census data is that population data is available at the block level sorted by age, sex, and race or ethnic group. Thus, if demographic information about the offender is available, then it too can be incorporated into our analysis. For example, notice the differences in the population density maps for different racial / ethnic groups shown in Figure 4.5.

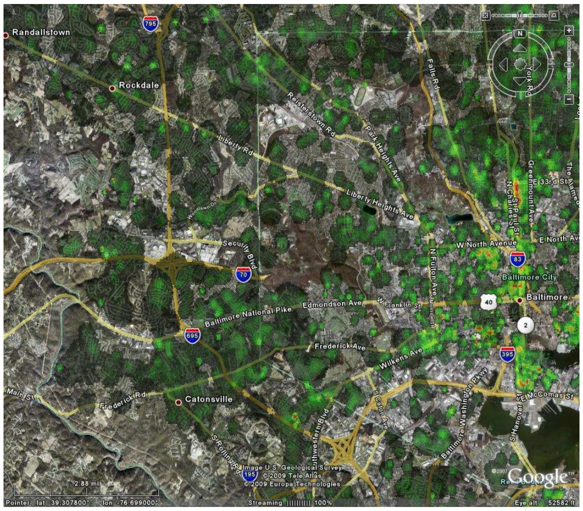
Another fundamental factor of interest is the distance decay function of an offender. Our mathematical method does not make an *a priori* choice of the offender's distance decay function. Instead, our method assumes that different offenders have different average offense distances. We provide the program with the locations of both the offense site and the home of 751 solved robberies. The program then performs the analysis described in Section 3.2.3, and generates the distribution of average offense distances across offenders that is shown in Figure 3.10.



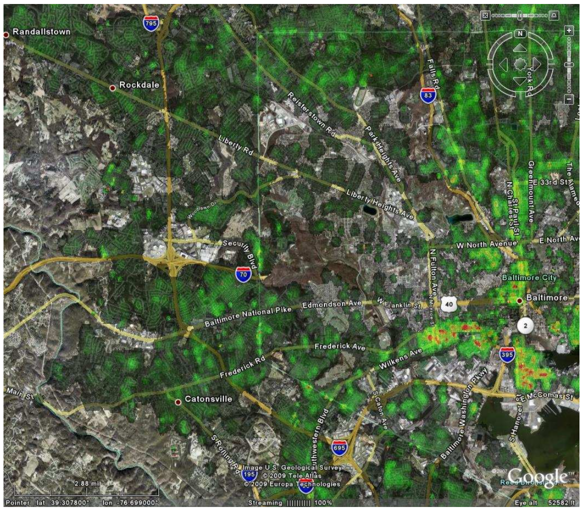
(a) Asian



(b) Black



(c) Hispanic



(d) White

Figure 4.5: Population density for different race / ethnic groups

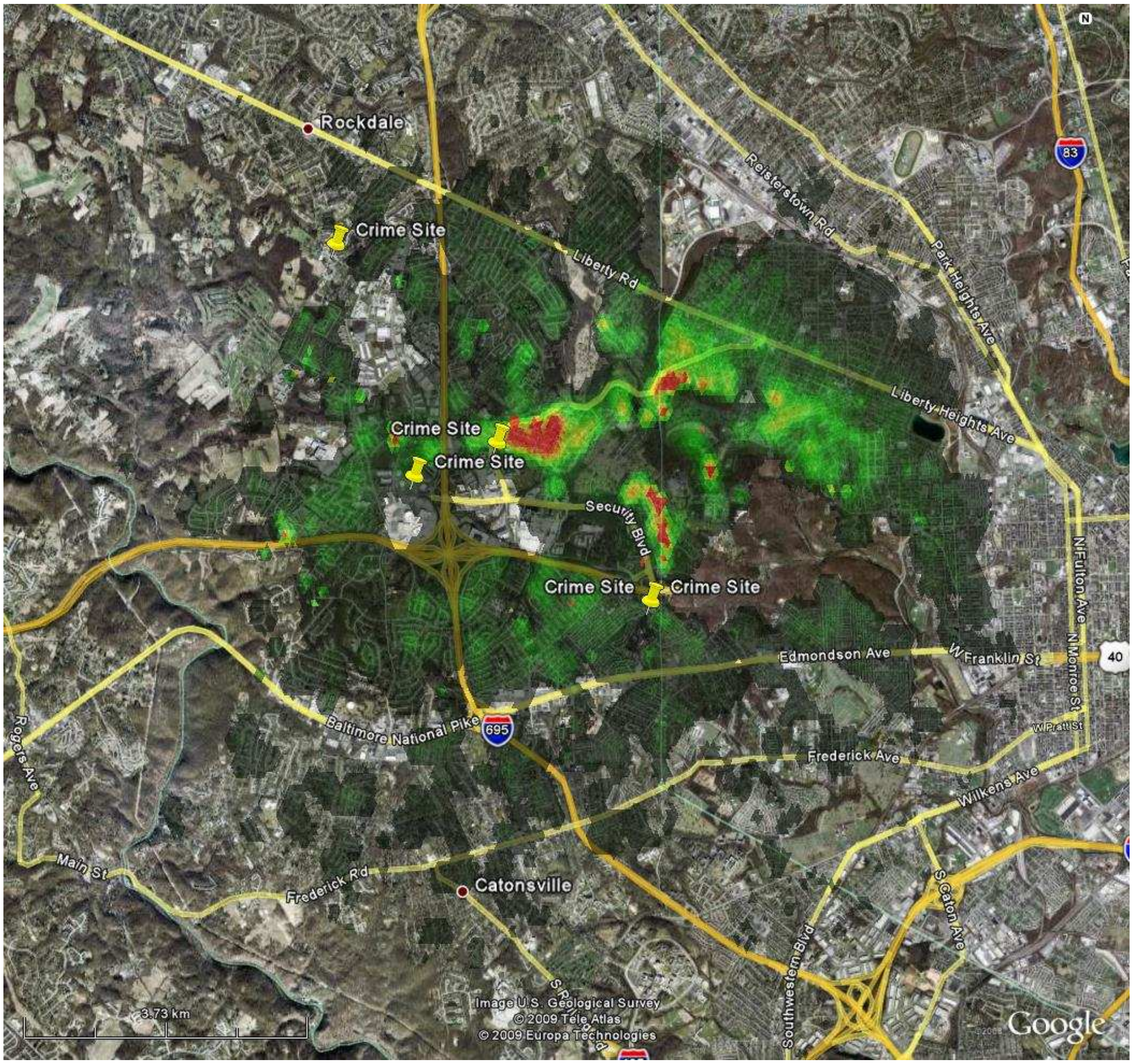


Figure 4.6: Proposed search area

When our software prototype is run on this series, we obtain the search area in Figure 4.6 Notice that the calculated search area avoids parks and commercial areas, while at the same time following the local road network. Note also that the regions considered most likely to contain the offender's residence do not lie in the region bounded by the crime sites, but rather lie towards the center of the city; compare this with the comments about the convex hull effect made by Levine [28].

The algorithm that generates the search area uses the local population density, so it knows that there are more potential offenders nearer to the city center than farther away. In addition, the algorithm is also aware of the jurisdictional boundaries; in particular it takes into account the possibility that elements of the series may have been committed in the city but that these elements are unknown to the analyst. Together, these facts suggest that the offender is more likely to live closer to the city, and the calculated search area agrees.

The potential impact of demographic information about the offender is clearly illustrated in

Figure 4.7 which show what the search area would be if we had information about the race or ethnic group of the offender. It is clear to see that the search areas have significant differences.

4.2 Implications for policy and practice

Our primary motivation for this project is to improve the operational efficiency of law enforcement agencies as they search for serial offenders by providing them with better and more sophisticated tools. This we have done; the software we have developed in now freely available on our website at `http://pages.towson.edu/moleary/Profiler.html`.

We presented our first functional prototype software package at the NIJ Conference in June 2009. The program is now being used by both the Los Angeles Police Department and the Baltimore County Police Department, both of whom are examining the effectiveness and usefulness of the tool.

It should be noted however, that we have not yet made a study of the effectiveness of the tool or of the mathematical algorithms that it contains.

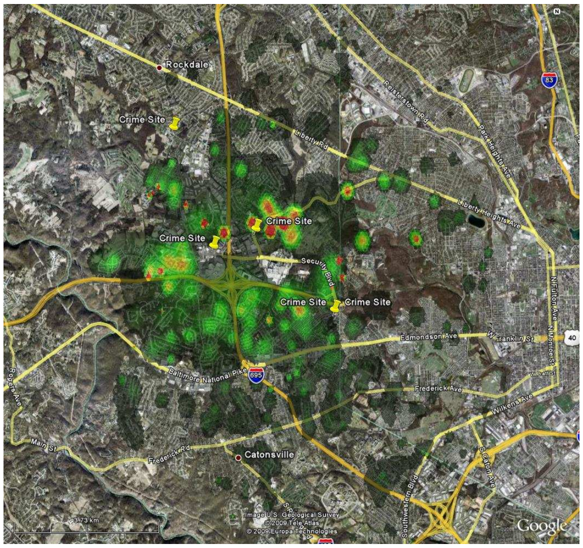
On the other hand, by making our mathematics, algorithms and code widely and publicly available, we also hope that we can provide valuable insights to other researchers. We also have written a manuscript that describes the mathematical techniques that we have used; this is currently under review at the Journal of Investigative Psychology and Offender Profiling.

4.3 Implications for further research

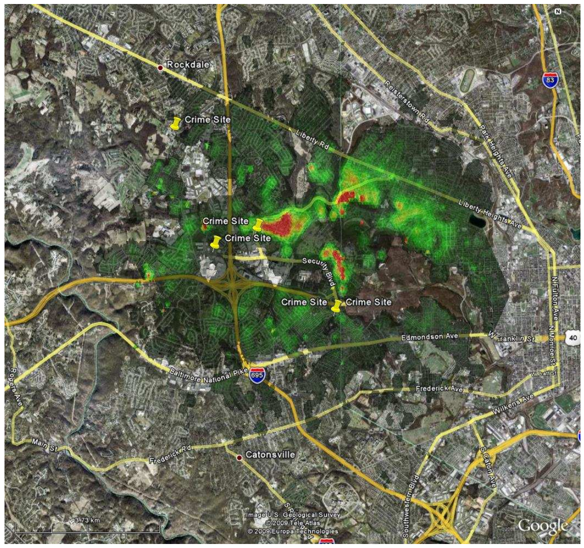
There are a number of important areas in which research in this area should be continued. First, the effectiveness of the our tool needs to be measured. The first prototype version was just released in June; it will take some time to collect the necessary data to see if this will result in any significant improvement over the current generation of techniques for geographic profiling.

Second, we need to remove some of the limitations that currently exist in the software prototype. For example, Phil Canter of the Baltimore County Police Department has indicated that the software prototype would be much more useful if the range of output file formats was increased. In particular, he has asked us to develop the capability to return results in either a plain text format, or even better in a Shapefile format so that the results can be better integrated into their existing GIS infrastructure. Sean Malinowski, of the Los Angeles Police Department asked if it would be possible to use geographical information that describes the distribution of known offenders in place of Census data to generate the estimate of the prior distribution of offenders. We hope to begin work on both of these suggestions soon so that we can get an improved tool into the hands of these agencies.

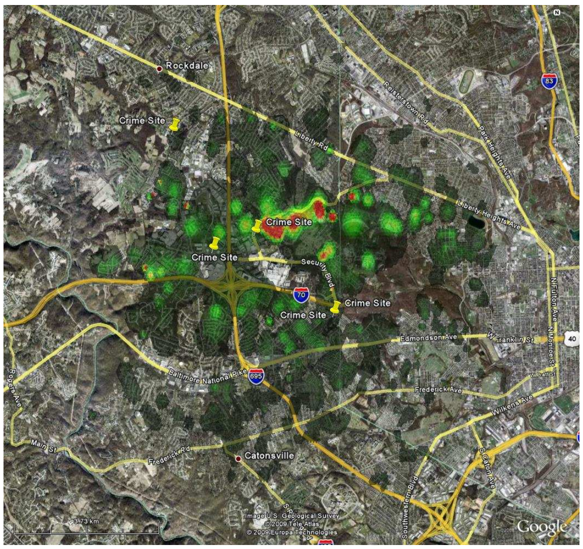
The mathematical framework we have developed for the geographic profiling problem is quite broad, and there are a number of reasonable forms for many of the distributions that are used in the model. To name just one example, our software tool posits that the distance decay component D follows a bivariate normal distribution, but clearly this is not the only reasonable alternative. We need to go back and start examining and evaluating different options for all of these distributions, and comparing them and their effectiveness. The tools of model selection and multimodel inference have yet to be applied to this problem, and may be able to generate some new insights.



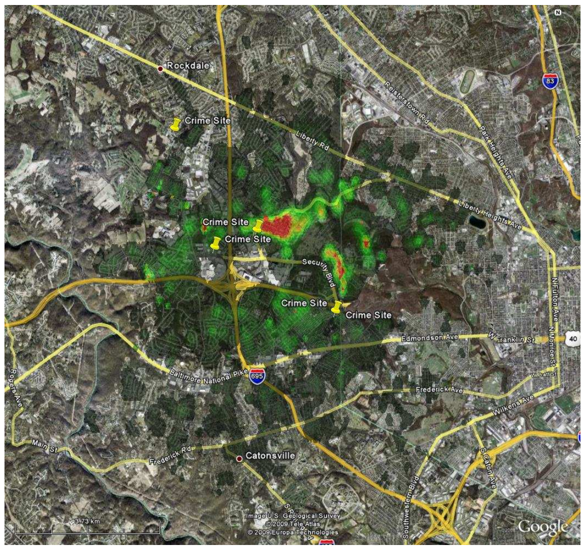
(a) Asian



(b) Black



(c) Hispanic



(d) White

Figure 4.7: Proposed search area for offender of different race / ethnic groups

Finally, the mathematical model of offender behavior described in $P(\mathbf{x} \mid \mathbf{z}, \alpha)$ remains quite simple. A new preprint by Mohler and Short [36] shows how they have developed an approach to geographic profiling where the simple probabilistic models for the distance decay component of offender behavior are replaced by the results of random walks. This is an innovative new idea that deserves significant additional study.

Bibliography

[1] Martin Baxter and Andrew Rennie, *Financial calculus*, Cambridge University Press, 1996.

[2] Mokhtar S. Bazaraa, Hanif D. Sherali, and C.M. Shetty, *Nonlinear programming*, John Wiley and Sons, 2006.

[3] Edward J. Beltrami, *Mathematical models in the social and biological sciences*, Jones and Bartlett Publishers, 1993.

[4] Craig Bennell, Brent Snook, and Paul Taylor, *Geographic profiling- The debate continues*, Blue Line Magazine (2005), 34–36.

[5] Craig Bennell, Brent Snook, Paul Taylor, Shevaun Corey, and Julia Keyton, *It's no riddle, choose the middle*, Criminal Justice and Behavior **34** (2007), no. 1, 119–132.

[6] Wim Bernasco and Paul Nieuwbeerta, *How do residential burglars select target areas? A new approach to the analysis of criminal location choice*, British Journal of Criminology **45** (2005), no. 3, 296–315.

[7] Patricia L. Brantingham and Paul J. Brantingham, *Residential burglary and urban form*, Urban Studies **12** (1975), no. 3, 273–284.

[8] ______, *Nodes, paths and edges: considerations on the complexity of crime and the physical environment*, Journal of Environmental Psychology **13** (1993), 3–28.

[9] Marilyn A. Brown, *Modelling the spatial distribution of suburban crime*, Economic Geography (1982), 247–261.

[10] Thiess Buettner and Hannes Spengler, *Local determinants of crime: Distinguishing between resident and non-resident offenders*, Darmstadt Discussion Papers in Economics 120, Institut fr Volkswirtschaftslehre (Department of Economics), Technische Universitt Darmstadt (Darmstadt University of Technology), February 2003, Online at `http://ideas.repec.org/p/tud/ddpiec/120.html`.

[11] David Canter, Toby Coffey, Malcolm Huntley, and Christopher Missen, *Predicting serial killers' home base using a decision support system.*, Journal of Quantitative Criminology **16** (2000), no. 4, 457–478.

[12] David Canter and P. Larkin, *The environmental range of serial rapists*, Journal of Environmental Psychology **13** (1993), 63–69.

[13] Bradley P. Carlin and Thomas A. Louis, *Bayes and empirical Bayes methods for data analysis*, 2nd ed., Chapman & Hall / CRC, 2000.

[14] George Casella and Roger L. Berger, *Statistical inference*, 2nd ed., Duxbury, 2002.

[15] Spencer Chainey, *Methods and techniques for understanding crime hot spots*, Mapping Crime: Understanding Hot Spots, The National Institute of Justice, Washington DC, August 2005, pp. 15–34.

[16] Warren John Ewens, *Mathematical population genetics*, Springer Verlag, 2004.

[17] Katarina Fritzon, *An examination of the relationship between distance travelled and motivational aspects of firesetting behaviour*, Journal of Environmental Psychology **21** (2001), no. 1, 45–60.

[18] Elizabeth Groff and Nancy G. La Vigne, *Mapping an opportunity surface of residential burglary*, Journal of Research in Crime and Delinquency **38** (2001), no. 3, 257–278.

[19] Per Christian Hansen, *Rank-deficient and discrete ill-posed problems*, SIAM, 1998.

[20] David Kincaid and Ward Cheney, *Numerical analysis*, second ed., Brooks Cole, 1996.

[21] Richard N. Kocsis, Ray W. Cooksey, Harvey J. Irwin, and Greg Allen, *A further assessment of circle theory for geographic psychological profiling*, The Australian and New Zealand Journal of Criminology **35** (2002), no. 1, 43–62.

[22] Richard N. Kocsis and Harvey J. Irwin, *An analysis of spatial patterns in serial rape, arson, and burglary: The utility of the circle theory of environmental range for psychological profiling*, Psychiatry, Psychology and Law **4** (1997), 195.

[23] Peter J. Van Koppen and Jan W. de Keijser, *Desisting distance decay: On the aggregation of individual crime trips*, Criminology **35** (1997), 505.

[24] Peter J. Van Koppen and Robert W. J. Jansen, *The road to robbery. Travel patterns in commercial robberies*, British Journal of Criminology **38** (1998), no. 2, 230–246.

[25] Manne Laukkanen and Pekka Santtila, *Predicting the residential location of a serial commercial robber*, Forensic Science International **157** (2006), no. 1, 71–82.

[26] Charles L. Lawson and Richard J. Hanson, *Solving least squares problems*, SIAM, 1995.

[27] James L. LeBeau, *The methods and measures of centrography and the spatial dynamics of rape*, Journal of Quantitative Criminology **3** (1987), no. 2, 125–141.

[28] Ned Levine, *The evaluation of geographic profiling software: Response to Kim Rossmo's critique of the NIJ methodology*, Online at `http://www.nedlevine.com/ResponsetoKimRossmoCritiqueoftheGPEvaluationMethodology.May82005.doc`, (Accessed July 2009), 2005.

[29] ______, *CrimeStat: A spatial statistics program for the analysis of crime incident locations (v. 3.2)*, Online at `http://www.icpsr.umich.edu/crimestat`, (Accessed July 2009), 2009.

[30] ______, *Update notes to version 3.2. CrimeStat: A spatial statistics program for the analysis of crime incident locations*, Online at `http://www.icpsr.umich.edu/crimestat`, (Accessed July 2009), 2009.

[31] Ned Levine and Richard Block, *Bayesian journey to crime estimation: An improvement in geographic profiling methodology*, Professional Geographer, to appear.

[32] Manuel J.J. López, *The spatial behavior of residential burglars*, Proceedings of the 5th Space Syntax Symposium (A. van Nes, ed.), Techne Press, 2005, pp. 423–435.

[33] Yongmei Lu, *Getting away with the stolen vehicle: an investigation of journey-after-crime*, The Professional Geographer **55** (2003), no. 4, 422–433.

[34] Jacek Malczewski, Anneliese Poetz, and Luigi Iannuzzi, *Spatial analysis of residential burglaries in London, Ontario*, The Great Lakes Geographer **11** (2004), no. 1, 15–27.

[35] Rebecca Meaney, *Commuters and marauders: An examination of the spatial behavior of serial criminals*, Journal of Investigative Psychology and Offender Profiling (2004), 121–137.

[36] George O. Mohler and Martin B. Short, *Geographic profiling from kinetic models of criminal behavior*, Online at `http://www.math.ucla.edu/~gmohler/geopro2.pdf`, (Accessed July 2009).

[37] Denise R. Osborn and Andromachi Tseloni, *The distribution of household property crimes*, Journal of Quantitative Criminology **14** (1998), no. 3, 307–330.

[38] Jerry H. Ratcliffe, *A temporal constraint theory to explain opportunity-based spatial offending patterns*, Journal of Research in Crime and Delinquency **43** (2006), no. 3, 261–291.

[39] George F. Rengert, Alex R. Piquero, and Peter R. Jones, *Distance decay reexamined*, Criminology **37** (1999), 427–445.

[40] Tom Rich and Michael Shively, *A methodology for evaluating geographic profiling software*, Tech. report, Abt Associates, Cambridge, MA, 2004.

[41] Kim Rossmo, *Geographic profiling*, CRC Press, 2000.

[42] ______, *An evaluation of NIJ's evaluation methodology for geographic profiling software*, Online at `http://www.txstate.edu/gii/documents/ResponsetoNIJGPEvaluationMethodology.doc`, (Accessed July 2009), 2005.

[43] ______, *Geographic heuristics or shortcuts to failure?: Response to Snook et. al.*, Applied Cognitive Psychology **19** (2005), 651–654.

[44] Kim Rossmo and Scot Filer, *Analysis versus guesswork*, Blue Line Magazine (2005), 24,26.

[45] Kim Rossmo, Scot Filer, and Carl Sesley, *Geographic profiling debate- round four. The big problem with Bennell, Snook and Taylor's research*, Blue Line Magazine (2005), 28–29.

[46] Bernard Silverman, *Density estimation for statistics and data analysis*, CRC Press, 1986.

[47] Brent Snook, *Individual differences in distance travelled by serial burglars*, Journal of Investigative Psychology and Offender Profiling **1** (2004), no. 1, 53–66.

[48] Brent Snook, David Canter, and Craig Bennell, *Predicting the home location of serial offenders: A preliminary comparison of the accuracy of human judges with a geographic profiling system*, Behavioral Sciences & the Law **20** (2002), 109–118.

[49] Brent Snook, Paul Taylor, and Craig Bennell, *Geographic profiling: The fast, frugal, and accurate way*, Applied Cognitive Psychology **18** (2004), no. 1, 105–121.

[50] ______, *Man versus machine. The case of geographic profiling*, Blue Line Magazine (2005), 56.

[51] ______, *Shortcuts to geographic profiling success: A reply to Rossmo*, Applied Cognitive Psychology **19** (2005), 655–661.

[52] Brent Snook, Michele Zito, Craig Bennell, and Paul J. Taylor, *On the complexity and accuracy of geographic profiling strategies*, Journal of Quantitative Criminology **21** (2005), no. 1, 1–26.

[53] Andromachi Tseloni, Karin Wittebrood, Graham Farrel, and Ken Pease, *Burglary victimization in England and Wales, the United States, and the Netherlands*, British Journal of Criminology **44** (2004), no. 1, 66–91.

[54] Curtis R. Vogel, *Computational methods for inverse problems*, SIAM, 2002.

[55] Fahui Wang and W. William Minor, *Where the jobs are: Employment access and crime patterns in Cleveland*, Annals of the Association of American Geographers **92** (2002), no. 3, 435–450.

[56] Janet Warren, Roland Reboussin, Robert R. Hazelwood, Andrea Cummings, Natalie Gibbs, and Susan Trumbetta, *Crime scene and distance correlates of serial rape*, Journal of Quantitative Criminology **14** (1998), no. 1, 35–59.

[57] R. Clyde White, *The relation of felonies to environmental factors in Indianapolis*, Social Forces **10** (1931), 498–509.

[58] Paul Wiles and Andrew Costello, *The road to nowhere: the evidence for travelling criminals*, Tech. report, Home Office Research Study 207, 2000, Online at `http://www.homeoffice.gov.uk/rds/pdfs/hors207.pdf`, accessed July 2009.

[59] Paul Wilmott, *Derivatives*, John Wiley and Sons, 1998.

[60] Paul Wilmott, Sam Howison, and Jeff Dewynne, *The mathematics of financial derivatives*, Cambridge University Press, 1995.

[61] B. J. Worton, *Kernel methods for estimating the utilization distribution in home-range studies*, Ecology **70** (1989), no. 1, 164–168.

Dissemination of Research Findings

Articles

- Mike O'Leary. *The Mathematics of Geographic Profiling*, under review at the Journal of Investigative Psychology and Offender Profiling

Presentations

- *A new software tool for geographic profiling*, The NIJ Conference, Arlington, VA, June 2009
- *Mathematical models for the geographic profiling problem*, Center for Evidence Based Crime Policy, George Mason University, March 2009
- *Mathematical models for the geographic profiling problem*, Georgetown University Mathematics Department Colloquium, March 2009
- *Determining the Optimal Search Area for a Serial Criminal*, Joint Mathematics Meetings, Washington DC, January 2009.
- *Advances in Geographic Profiling*, 31st Applied Geography Conference, Wilmington DE, October 2008.
- *Determining the Optimal Search Area for a Serial Criminal*, INFORMS National Meeting, Washington DC, October 2008.
- *The Mathematics of Geographic Profiling*, The NIJ Conference, Arlington VA, July 2008
- *The Mathematics of Geographic Profiling*, Spring Meeting, Geospatial Technology Working Group, New Orleans LA, April 2008.
- *The Mathematics of Geographic Profiling*, Center for Army Analysis, Fort Belvoir, April 2008

List of Supplemental Material

- The software tool, as a .zip archive
- Instructions for the use of the software tool, as a .pdf slide show
- The source code for the Profiler program
- Browseable documentation for the source code of the Profiler program in web / html format
- Printable documentation for the source code of the Profiler program in .pdf format
- The source code for the ProfilerGUI program.
- Browseable documentation for the source code of the ProfilerGUI program in web / html format
- Printable documentation for the source code of the ProfilerGUI program in .pdf format