# Identifying Fixations in Gaze Data via Inner-Density and Optimization

Andrew C. Trapp[†,‡], Wen Liu[‡], Soussan Djamasbi[†]

[†]Robert A. Foisie Business School, [‡]Data Science Program
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA USA

atrapp@wpi.edu, wliu3@wpi.edu, djamasbi@wpi.edu

**Abstract:** Eye tracking is an increasingly common technology with a variety of practical uses. Eye-tracking data, or gaze data, can be categorized into two main events: *fixations* represent focused eye movement, indicative of awareness and attention, whereas *saccades* are higher velocity movements that occur between fixation events. Common methods to identify fixations in gaze data can lack sensitivity to peripheral points, and may misrepresent positional and durational properties of fixations. To address these shortcomings, we introduce the notion of *inner-density* for fixation identification, which concerns both the duration of the fixation, as well as the proximity of its constituent gaze points. Moreover, we demonstrate how to identify fixations in a sequence of gaze data by optimizing for inner-density. After decomposing the clustering of a temporal gaze data sequence into successive regions (chunks), we use nonlinear and linear 0–1 optimization formulations to identify the densest fixations within a given data chunk. Our approach is parametrized by a unique constant that adjusts the degree of desired density, allowing decision makers to have fine-tuned control over density during the process. Computational experiments on real datasets demonstrate the efficiency of our approach, and its effectiveness in identifying fixations with greater density than existing methods, thereby enabling the refinement of key gaze metrics such as fixation duration and fixation center.

**Keywords:** Eye-Tracking; Gaze Data; Fixation Identification; Fixation Inner-Density; Mixed-Integer Nonlinear Optimization

## 1. Introduction

Interest in understanding the behavior and movement of eyes has long existed, and is proliferating with the availability of low-cost eye-tracking devices that have ever-increasing capabilities. Indeed, it is estimated that eye-tracking mechanisms will be standard options for laptop computers in the near future (Djamasbi 2014). A tracking device is able to record

eye-tracking, or *gaze*, data, of a subject that is presented a visual stimulus. Many purposes for studying gaze exist, including understanding of the human visual system (Radvay et al. 2007), diagnosis of psychological disorders (Cockerham et al. 2009), analysis of marketing techniques (Wedel and Pieters 2008), design of products (Goldberg et al. 2002), and web experience (Djamasbi et al. 2010), among others.

Essential to many eye-movement behavior applications is a precise understanding of the recorded gaze data from eye-tracking devices (Nyström and Holmqvist 2010). This understanding comes from the translation of raw, longitudinal gaze data into distinct eye-movement, or *oculomotor*, events. This process is known as *fixation identification* (Salvucci and Goldberg 2000), and it separates gaze data into two primary event types: fixations and saccades. *Fixations* are pauses over informative regions of interest, where cognitive processing is believed to occur, whereas *saccades* are rapid movements between fixations, used to recenter the eye on a new location (Salvucci and Goldberg 2000, Blignaut 2009). Fixations are the primary unit of analysis for attention and awareness studies. Fixations characterize attention because they represent effort in maintaining a relatively stable gaze to take foveal snapshots of an object for subsequent processing by the brain (Djamasbi 2014).

To date, computational analysis has enabled a great deal of progress towards translating gaze data into fixations. Primary existing methods for identifying fixations use either *gaze location* (e.g., I-DT filter) or *velocity* metrics (e.g., I-VT filter). Methods based on the former typically use a constant area size as the threshold for grouping consecutive gaze points into a fixation, while the latter use a fixed velocity threshold to separate fixations from saccades. While these existing approaches are relatively simple to implement and generally effective, they can lead to issues with precision because they are prone to including points on the fringe of tolerance settings, thereby skewing summary fixation metrics (further discussed in Section 2).

Our work makes two novel contributions to address these shortcomings. The first is the identification of fixations via *inner-density*, which carries two characterizations of cognitive effort: the duration of a fixation, as well as its proximal compactness. It has been shown that fixation duration is a reliable measure of attention (Djamasbi 2014), and proximal compactness of individual gaze points in a fixation represent a person's focused attention and increased levels of information processing (Shojaeizadeh et al. 2016). Fixations with greater inner density tend to exclude peripheral gaze points, thereby improving the accuracy of traditional fixation metrics.

While there is great potential to use inner-density for refining gaze data, there are no known studies that use the concept to identify fixations, let alone optimization-based approaches. Our second contribution is a *computational* approach to identify the *densest* fixa-

tions from gaze data. Given the impressive progress of modern optimization technology (for one such review in the context of data analysis see, e.g., Bertsimas and King 2015), exact methods that provide a performance guarantee on solution quality are now a reality; that is, given a dense fixation, optimization methods can prove *no denser fixation exists*. This is incredibly important when exact, rather than approximate, oculomotor event identification is desirable, or even essential.

In this paper, we develop and compare two density-based optimization approaches to identify fixations in gaze data. We are unaware of any other studies that explicitly consider inner-density as a criterion for fixation detection, i.e., emphasizing fixations with many points in a relatively compact region. Neither are we aware of any study that uses optimization-related technology to detect fixations. We decompose the broader problem of identifying all fixations in a sequence of raw gaze data points, into smaller, manageable chunks that are in particular amenable to exact solution via optimization-based approaches. We develop two new mathematical formulations, a binary integer nonlinear program that is subsequently linearized and a mixed 0–1 integer linear program, each of which optimize for density while adhering to necessary temporal and consistency restrictions. Further, we share a similar viewpoint as that of Nyström and Holmqvist (2010), who advocate for leaving as few subjective settings to the end-user as possible, in that we introduce a single, powerful parameter $\alpha$ that influences the outcome of whether each individual gaze point is included in a fixation. This enables researchers to make fine-tune adjustments to the inner-density, as desired, to study focused attention at a more refined scale. By decomposing the entire gaze point sequence by velocity, and subsequently optimizing for density on individual chunks, essentially a dynamic dispersion threshold, we exploit the attractive properties of two leading fixation detection methods, the I-VT and I-DT filters.

The remainder of this paper is organized in the following manner. In Section 2 we provide background material related to gaze data, including existing processing methods, as well as our proposed characterization of the inner-density of a single fixation. In Section 3 we highlight our technical developments, including a formal problem description and related analysis of the problem, as well as a decomposition strategy that breaks the overall problem into more easily processed components, a novel constraint set that ensures fixations contain temporally consecutive gaze points, and the integration of fine-grained density control through parametrization. We then provide two mathematical programs together with an algorithmic approach that utilize these developments to identify the densest fixations in gaze data. Section 4 contains computational experiments and related discussion, where we use real gaze datasets to test our approaches. The paper concludes with a summary of our work, including future directions, in Section 5.

# 2. Background

Gaze data has a particular structure, and must be reliably processed to generate meaningful information. Prior to proposing fixation *inner-density* and its associated optimization as a novel approach to measure information processing behavior, we review existing methods.

## 2.1 Common Fixation Identification Methods Based on Position and Velocity

The process of fixation identification separates gaze data into distinct oculomotor events (e.g., fixations and saccades). The gaze data we consider results from user interaction with 2D static stimuli, e.g. visual computer displays, as a major focus of behavioral research is to understand user interaction with static screen based technologies. This gaze data is recorded in two dimensions $(x, y)$ for every discrete time point $t$. Hence each 2D data point is an $(x, y, t)$ triplet. Each time-series sequence $\mathcal{S}$ of consecutive discrete $(x, y, t)$ gaze data points can be computationally separated into constituent fixations. Common sampling rate frequencies range from 30 Hz to 300 Hz, though some eye-tracking devices can record at levels exceeding 1,000 Hz (Holmqvist et al. 2011). Once gaze data has been computationally processed into its fundamental oculomotor events, each event can be characterized using summary statistics, for example the duration and center (centroid) of the event. Figure 1 depicts approximately 10,000 gaze points in a segment of a real, raw gaze data sequence in $(x, y, t)$ space, which arises from a task of reading on a 2D static computer display stimulus. The problem of interest is to separate this gaze data into distinct fixations.

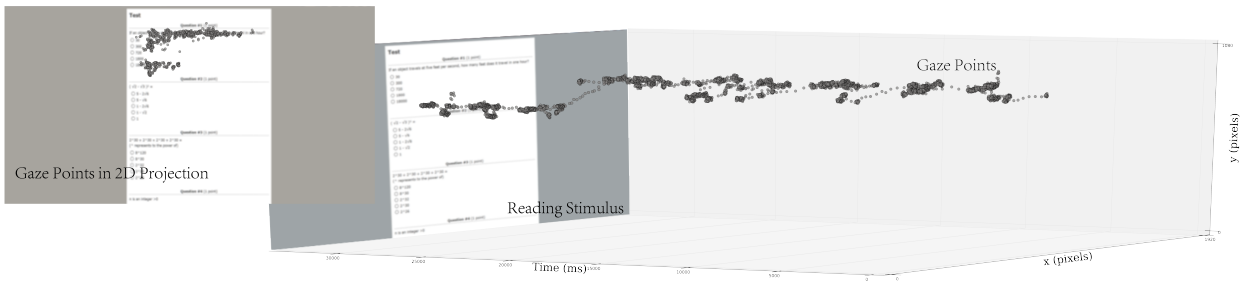


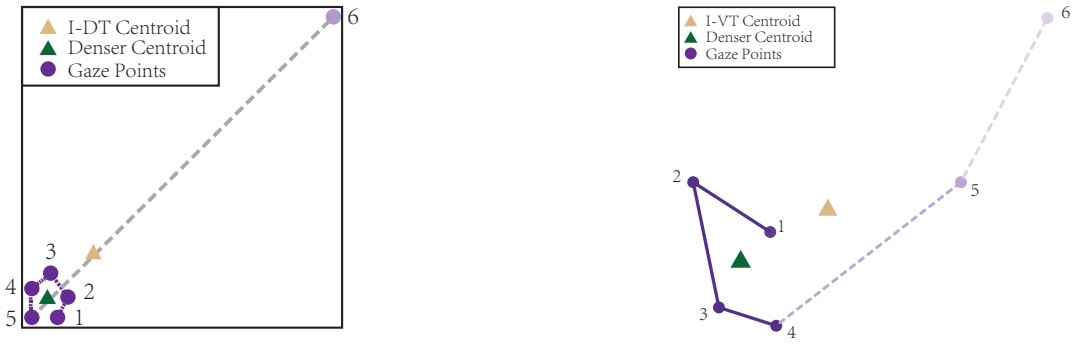

Figure 1: Raw $(x, y, t)$ gaze data depicted in three dimensions, as recorded by a typical eye-tracking device.

Two primary methods exist to analyze and process gaze data: those based on gaze-point *position*, such as the I-DT, and those based on gaze-point *velocity*, such as the I-VT (for in-depth descriptions of these approaches, see, e.g., Salvucci and Goldberg 2000, Komogortsev et al. 2010). It is widely accepted that all existing event detection methods

have flaws (Nyström and Holmqvist 2010). This is due in part to the arbitrary and somewhat interpretive nature of classifying gaze data points into representative events. Estivill-Castro (2002) contends that the reason there are so many ways to identify fixations (clusters) is because the notion cannot precisely be defined; rather, it is in the eye of the beholder. Even so, there are basic criteria, many used by existing approaches, that are suggestive for a group of points to be considered as a fixation.

The I-DT is a well-known position-based approach. This algorithm separates gaze data using a predefined maximum dispersion threshold $D$ together with a minimum duration. It uses a fixed-area window to construct fixations by sequentially adding points beyond a minimum duration, until the dispersion threshold is exceeded (Salvucci and Goldberg 2000). The I-DT can yield fairly accurate results, is rather straightforward to implement, and has favorable performance time. However, a significant drawback arises from the interaction with the threshold $D$ and the dispersion metric it uses:

$$D(x, y) = D(x) + D(y) = [max(x) - min(x)] + [max(y) - min(y)]. \tag{1}$$




(a) The I-DT algorithm may misclassify gaze points under static dispersion threshold $D$. Whether to include the sixth point in the fixation, while technically within outer threshold $D$, is questionable.

(b) The I-VT algorithm may misclassify gaze points under constant velocity threshold $V$. The fifth and sixth points, while technically having velocities below threshold $V$, may not belong to the fixation.

Figure 2: Depicting some limitations of standard methods for fixation identification. For both the I-DT and I-VT algorithms, the center points (centroids) appear as lighter triangles, shifted to the upper right, as opposed to those of the denser fixation centroids, which are depicted with darker triangles and are more representative of the center of fixation of interest.

Figure (2a) illustrates some of the challenges with I-DT in assuming a simple, constant dispersion threshold $D$. As long as the $D(x, y)$ measure does not exceed $D$, points are

considered to belong to the same fixation. Figure (2a) raises significant doubts as to whether the sixth point belongs to the same fixation as the first five gaze points. This in turn can skew metrics such as the fixation duration and centroid, which are often used to assess user reaction to stimuli (Sabatos-DeVito et al. 2016, Thorup et al. 2016).

The I-VT algorithm may be the simplest of all fixation detection approaches, which sequentially categorizes each gaze point based on its point-to-point velocity. If the velocity meets or exceeds a velocity threshold $V$, it is identified as a saccade; below, the point belongs to a fixation (Salvucci and Goldberg 2000). I-VT is an elegant algorithm; as Salvucci and Goldberg (2000) discuss, it is a rather straightforward and robust approach, because the physical and physiological nature of the velocity profiles naturally separate data points into fixations or saccades. In fact, the I-VT algorithm serves as the foundation for the fixation detection algorithms in major commercial eye-tracking devices such as Tobii (see, e.g., Olsen 2012). The I-VT algorithm features a simple implementation, efficient performance, and is fairly accurate.

Even so, the I-VT algorithm also has significant limitations. It essentially considers any consecutive group of points below a specified velocity threshold as a fixation. It then uses this grouping as a basis for summary statistics, such as the $(x, y)$ centroid (by collapsing into a single point the individual $x$ and $y$ points according to their average values). Hence, the simplicity of the I-VT algorithm may result in misclassification, that is, points being classified as within the same fixation – when in reality they are distinct – because they do not strictly exceed the velocity threshold. As can be seen in Figure (2b), the inclusion of gaze points that are technically below the velocity threshold, but would not otherwise be included in a fixation, can skew important metrics such as the fixation duration and centroid. When considering the first four fixation points in Figure (2b), the centroid appears lower, and to the left, of where it appears when all six points are included in a fixation.

Although a few studies exist on enhancing the I-VT algorithm (see, e.g., Smeets and Hooge 2003, Nyström and Holmqvist 2010), the aforementioned drawback remains detrimental for applications that demand precision. To resolve inherent discrepancies present in commonly used methods for fixation identification, we propose the concept of *inner-density*, which refers to both the duration and concentration of the gaze points that form a fixation. In the next section we explain how we use inner-density to identify the densest fixations in a stream of temporal gaze data.

## 2.2 Fixation Identification Based on Inner-Density

In this study we propose detecting fixations through inner-density, in lieu of solely distance or velocity. Fixation inner-density represents the spatial concentration of gaze points, and is an attractive metric to optimize because denser fixations reveal more focused attention (Shojaeizadeh et al. 2016). The density of a fixation increases when either a larger number of points are contained in the fixation, or when the constituent points are contained in a more compact region, or both. Mathematically, density can be expressed in multiple ways, such as a ratio or a weighted sum.

It is important to note that fixation inner-density is distinct from the seemingly similar, but separate, notion of *spatial* density, which addresses the concept of the proximity of *multiple* clusters of gaze points (i.e., fixations). Spatial density involves the post-processing step of merging individual fixations into a larger fixation (Goldberg and Kotval 1999, Poole and Ball 2005), for example as done on a fixation density map (Engelke et al. 2013), or the use of Voronoi diagrams to represent the uniformity of fixation density (Over et al. 2006). An excellent, in-depth source of gaze processing information is Holmqvist et al. (2011).

To date, the authors are unaware of any studies that explicitly use inner-density for gaze data, with a possible, though indirect, exception being a recent study on extending the popular DBSCAN clustering algorithm to oculomotor event detection (Li et al. 2016).

## 2.3 Related Work in the Literature

We now review the literature for works that use techniques from mathematical optimization or related exact approaches to conduct, or enhance, eye-tracking event detection. Within the eye-tracking literature, the authors are unaware of any such works. We highlight only a single study that attempts to determine, empirically, an optimal parameter setting. On the other hand, the task of identifying fixations in gaze data can be viewed as a type of *clustering*. Clustering is the analytical process of collecting elements into a group, or *cluster*, such that elements that are gathered together have a greater similarity as opposed to those in other clusters. There are a number of studies that attempt to cluster data using some type of exact (i.e. either optimization-based, or approximation scheme) methodology.

### 2.3.1 Estimating Fixation Detection Parameters

The only work that is somewhat related is the excellent study conducted by Blignaut (2009) to estimate "optimal" dispersion thresholds for dispersion algorithms. Through empirical investigations, the optimal fixation radius threshold was found to be from 0.7° to 1.3° of visual

angle. We use their threshold recommendation in our I-DT implementation (see Section 4.5).

### 2.3.2 Using Optimization for Clustering

We review works from the literature that conduct clustering using some type of exact methodology, either optimization-based or approximation schemes. Such methods that provide a guarantee on the quality of the solution are increasingly viable given the state-of-the-art in optimization technology. As we review work from the literature that use such exact methodologies, it is important to note that any successful method for clustering gaze data must address its *temporal* nature, that is, recovered clusters should contain only consecutive gaze points.

Selim and Ismail (1984) formulate the $k$-means clustering problem as a non-convex mathematical program, and investigate properties related to convergence, local and global optimality. Sağlam et al. (2006) propose a mixed-integer nonlinear formulation for clustering that they subsequently linearize to minimize the maximum cluster diameter among all of the clusters. Bradley et al. (1997) discuss the task of finding $k$ cluster centers using mathematical programming techniques so that the sum of distances of each point to the nearest cluster is minimized. While their approach does not guarantee global optimality, they demonstrate that it has favorable performance to classical methods such as $k$-means. Rao (1971) proposes an integer nonlinear program having a ratio objective to minimize the sum of average squared distances within a given cluster. None of these studies, however, consider temporal data. Seref et al. (2013) build upon the work of Bradley et al. (1997) for time series data, introducing both exact formulations and fast approximation algorithms that compare favorably to existing methods. Concerning approximation algorithms, Hochbaum and Maass (1985) discuss approximation schemes for a related problem of using circles to cover points in a plane. Approximation schemes are also used by Charikar et al. (2004) to address a similar problem of finding good clusters when the data is dynamic.

# 3. Technical Developments

In this section we address the core challenge of *fixation identification* in gaze data. We begin with a formal problem description, provide a combinatorial analysis of the number of ways to form meaningful fixations, touch upon related literature, and discuss challenges in solving related to scaling. We then highlight three unique insights to facilitate efficient solution of the problem, and proceed to introduce two mathematical programming formulations that identify fixations by optimizing for inner-density, together with an iterative algorithm.

## 3.1 Fixation Identification: Formal Problem Description

Fixation identification is the process of translating a longitudinal sequence of raw eye-movement data points into constituent fixation events and, thereby, the saccadic events between them (Salvucci and Goldberg 2000). We are unaware of any formal characterization of the *fixation identification* problem, though a related problem of *sequence segmentation* is discussed in Terzi (2006), from which we adapt some notation.

Formally, we consider a raw time-series sequence $\mathcal{S}$ of $\mathcal{T}$ $d$-dimensional gaze points, so that $\mathcal{S} = \{t_1, \ldots, t_\mathcal{T}\}$. Let $S_\mathcal{T}$ denote all such sequences of length $\mathcal{T}$. We seek to form $\mathcal{F}$ fixations from these $\mathcal{T}$ gaze points. While $\mathcal{F}$ is in general difficult to know with certainty, a suitable value (or range of values) for $\mathcal{F}$ can often be informed by the problem context, and subsequently validated according to the resulting performance. For the sake of the formal problem description, we assume $\mathcal{F}$ is known a priori.

An important consideration is to determine which points belong to fixations; some should not be included as they are saccade points, or possibly some other noise. Points that do form fixations must be consecutive in time, and together should be of a minimum length to have meaning with respect to cognitive processing. At a fixed sampling frequency, this is equivalent to stating that every fixation must contain a minimum number of points $\mathcal{N}$. Hence, the $\mathcal{F}$ formed fixations constitute *segments* of the gaze sequence $\mathcal{S}$ that are mutually exclusive, and of a sufficient minimum length. Of particular interest to us are *dense* fixations, which we will further qualify.

An $\mathcal{F}$-segmentation $F$ of $\mathcal{S}$ can be uniquely represented by $\mathcal{F}$ pairs of fixation "segment" breakpoints. That is, $F = \{(f_1, f_2), \ldots, (f_{2\mathcal{F}-1}, f_{2\mathcal{F}})\}$, with $f_i \in \mathcal{S}$. These pairs of breakpoints denote the fixation points in $F$ through the respective intervals $[f_{2j-1}, f_{2j}]$, $j = 1, \ldots, \mathcal{F}$. Hence fixation $j$ contains $f_{2j} - f_{2j-1} + 1$ gaze points, which must meet or exceed $\mathcal{N}$ for information processing to occur, so that $f_{2j-1} + \mathcal{N} - 1 \leq f_{2j},\ j = 1, \ldots, \mathcal{F}$.

Let $\mathscr{S}_\mathcal{T}$ denote all possible segmentations of gaze sequences of length $\mathcal{T}$, and let $\mathscr{S}_{\mathcal{T},\mathcal{F},\mathcal{N}}$ denote all possible segmentations of sequences of length $\mathcal{T}$ into $\mathcal{F}$ fixation "segments" of length $\mathcal{N}$ or greater. Of particular interest is to minimize error criterion $E : S_\mathcal{T} \times \mathscr{S}_\mathcal{T} \mapsto \mathbb{R}$ that assesses the quality of the formed fixations. Specifically, $E$ should characterize two density-related aspects: fixation duration (a relatively *large number of* gaze points) and compactness (gaze points in *close proximity*).

For sequence $\mathcal{S}$ and error function $E$, we define the optimal $\mathcal{F}$-segmentation $F$ of $\mathcal{S}$ as:

$$F_{opt}(\mathcal{S}, \mathcal{F}) = \underset{F \in \mathscr{S}_{\mathcal{T},\mathcal{F},\mathcal{N}}}{\arg\min}\ E(\mathcal{S}, F), \tag{2}$$

that is, $F_{opt}$ is a grouping of $\mathcal{S}$ into $\mathcal{F}$ fixations that minimizes the function $E(\mathcal{S}, F)$.

**Problem 1** ***Fixation Identification.*** *Given a raw longitudinal gaze sequence $\mathcal{S}$ containing $\mathcal{T}$ total time points, integer values $\mathcal{F}$ and $\mathcal{N}$ respectively denoting the number of fixations and minimum number of points, together with error function $E$, identify $F_{opt}(\mathcal{S}, \mathcal{F})$.*

As it turns out, this problem has a very large number of possible segmentations.

### 3.1.1 Combinatorial Analysis

We illustrate this by counting the number of ways to identify $\mathcal{F}$ fixations in a sequence of $\mathcal{T}$ points. Each fixation must contain at least $\mathcal{N}$ points, which should be consecutive in time. Moreover, not all points must be included in fixations. This suggests that the $\mathcal{T}$ points can be assigned either *i*) to fixations ($\mathcal{F}$), *ii*) to intervals between fixations ($\mathcal{F}-1$), or *iii*) preceding (1) or *iv*) following all fixations (1), so that in general there are $2\mathcal{F}+1$ distinct bins. A small example $\mathcal{S}$ with $\mathcal{T}=8$, $\mathcal{F}=2$ and $\mathcal{N}=3$ is illustrated in Figure 3, where gaze points are denoted as "$\circ$". For this example with $\mathcal{F}=2$, we depict $2\mathcal{F}+1=5$ bins: $a_1, a_2, a_3, a_4, a_5$, as well as three possible ways of grouping the gaze points (there are others). While each depicts $\mathcal{T}=8$ points, note that they differ with respect to which points are included and not included in fixations, and moreover the first fixation of the third grouping differs in size.

| $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ |
|---|---|---|---|---|
| $\circ$ | $\circ\circ\circ$ | | $\circ\circ\circ$ | $\circ$ |
| | $\circ\circ\circ$ | $\circ$ | $\circ\circ\circ$ | $\circ$ |
| | $\circ\circ\circ\circ$ | | $\circ\circ\circ$ | $\circ$ |

Figure 3: Visualizing Three Groupings of Gaze Points.

Multisets are useful to count the number of $\mathcal{F}$-segmentations of $\mathcal{S}$, as they generalize the concept of a set by allowing for multiple instances of elements. The multiplicity of an element is the number of instances of the element in a specific multiset. An infinite number of multisets exist which contain only elements $a_1$ and $a_2$, varying only by multiplicity. A general multiset of $n$ elements can be denoted by

$$M = \{\infty \cdot a_1, \infty \cdot a_2, \ldots, \infty \cdot a_n\}, \tag{3}$$

where $n \in \mathbb{Z}_+$, and $a_1, a_2, \ldots, a_n$ are distinct objects. Then a specific multiset of the form (3) is an $(m_1 + m_2 + \cdots + m_n)$-element multi-subset of $M$, with $m_1, m_2, \ldots, m_n \in \mathbb{Z}_+$ being the respective multiplicities. The number of $m$-element multi-subsets of $M$ is given by $H_m^n$ (see, e.g., Chen and Koh 1992):

$$H_m^n = \binom{m+n-1}{m}. \tag{4}$$

The number of possible ways to form $\mathcal{F}$ fixations of length at least $\mathcal{N}$ in $\mathcal{T}$ gaze points can be viewed as a certain multiset with $2\mathcal{F}+1$ distinct objects $M_p = \{m_1 \cdot a_1, m_2 \cdot a_2, \ldots, m_{2\mathcal{F}+1} \cdot a_{2\mathcal{F}+1}\}$. The quantity $(m_1 + m_2 + \cdots + m_{2\mathcal{F}+1})$ should amount to the total number of points that must be placed. While there are $\mathcal{T}$ total gaze points, because each of the $\mathcal{F}$ fixations must contain at least $\mathcal{N}$ points, we have that $m_1 + m_2 + \cdots + m_{2\mathcal{F}+1} = \mathcal{T} - \mathcal{N}\mathcal{F}$. Hence there are $H^{2\mathcal{F}+1}_{\mathcal{T}-\mathcal{N}\mathcal{F}}$ total ways of assigning $\mathcal{T}$ points into $\mathcal{F}$ fixations of size at least $\mathcal{N}$, where:

$$H^{2\mathcal{F}+1}_{\mathcal{T}-\mathcal{N}\mathcal{F}} = \binom{\mathcal{T} - \mathcal{N}\mathcal{F} + 2\mathcal{F}}{\mathcal{T} - \mathcal{N}\mathcal{F}} = \binom{\mathcal{T} - (N-2)\mathcal{F}}{\mathcal{T} - \mathcal{N}\mathcal{F}}. \tag{5}$$

The quantity in (5) grows very quickly for even small $\mathcal{T}$ and, in particular, $\mathcal{F}$; for example, for only $\mathcal{F} = 10$ and $\mathcal{T} = 50$, the number of combinations is on the order of $10^{11}$.

### 3.1.2 Problems Related to Fixation Identification

A related problem, *sequence segmentation*, is also concerned with optimal segmentation of time series sequences of data (Terzi (2006), Terzi and Tsaparas (2006); also see Bingham 2010). They too consider minimizing an error criterion, for example distance from the center of the sequence. However a key distinction is that in the *sequence segmentation* problem, *all points* must be used to form relevant segments (clusters). On the contrary, the *fixation identification* problem forms fixations with *only the most salient time points* – that is, there are data points in the gaze sequence that should not be included in any fixation. A dynamic programming algorithm is presented in Terzi (2006) to solve the *sequence segmentation* problem in $\mathcal{O}(\mathcal{T}^3\mathcal{F})$ time, and it is further reduced to $\mathcal{O}(\mathcal{T}^2\mathcal{F})$ time through a series of clever algorithmic enhancements.

While a dynamic program similar to that of Terzi (2006) also exists for the *fixation identification* problem, it has $\mathcal{O}(\mathcal{T}^3\mathcal{F})$ complexity due to the need to process the assignment of points to fixations as well as to intervals between fixations, and unfortunately it becomes prohibitive to solve for even modest sizes of the fixation identification problem (indeed, Terzi (2006) similarly notes "*...cubic complexity makes the dynamic programming algorithm prohibitive to use in practice.*"). This suggests alternative solution approaches are necessary that further exploit the structure of the *fixation identification* problem.

## 3.2 Three Mathematical Insights

We next highlight insights that enable us to develop an algorithmic approach to identify the densest fixations in $\mathcal{S}$.

### 3.2.1 Decomposition Principle: Saccades Separate Fixations

A gaze sequence $\mathcal{S}$ contains a large number of $(x, y)$ points over time. Common lengths of gaze data sequences are in the tens to hundreds of seconds. For frequencies of 30 Hz to 300 Hz, $\mathcal{S}$ can contain anywhere from several hundred, to hundreds of thousands of gaze points, and may contain hundreds if not thousands of fixations. For such realistic data instances, the fixation identification problem is prohibitive for even a moderate number of fixations, as proving the optimality of clusters on large datasets is known to be computationally demanding (Trapp et al. 2010, Trapp and Prokopyev 2010, Seref et al. 2013).

An alternative perspective leverages the specific structure of the sequence $\mathcal{S}$. Fixations must occur over temporally consecutive gaze points. Hence, any point that is identified as saccadic (e.g., by the I-VT filter) is a separator of fixations. Moreover, any small number of consecutive points may be removed if they are below a reasonable lower threshold for information processing to occur (similar to the I-DT filter). By removing these two types of gaze points, the gaze sequence $\mathcal{S}$ becomes a collection of disjoint sets, or chunks, of gaze points where fixations may occur – that is, there are no saccadic points, and each chunk contains at least a minimum number of gaze points to be considered a fixation. Such a process separates $\mathcal{S}$ into $\mathcal{K}$ chunks of potential fixation points $\mathcal{C}^k,\ k = 1, \ldots, \mathcal{K}$. In particular, $\mathcal{C}^i \cap \mathcal{C}^j = \emptyset,\ 1 \leq i < j \leq \mathcal{K}$, and $\cup_{k=1,\ldots,\mathcal{K}}\, \mathcal{C}^k \subseteq \mathcal{S}$. Each of these chunks can subsequently be explored, independently, for (dense) fixations, as it is very likely that each chunk contains a minimal number of fixations.

### 3.2.2 Fixations Contain Consecutive Points in Time

There are fundamental differences between clustering temporal versus non-temporal data. In particular, fixations must adhere to temporal restrictions, which represents an extra condition for typical (atemporal) clustering tasks. Once a fixation begins, the included points must be consecutive in time, until the fixation ends. Stated another way, a fixation may conclude only once in a given sequence of gaze points. If this were not the case, fixations that occur in the same proximity, but separated over distinct periods of time, may be considered as a single fixation. Moreover, saccadic points that collect over time in the same region could also be incorrectly classified as a fixation (see, e.g., Li et al. 2016). To facilitate the ensuing discussion, define $\mathcal{TF}$ binary variables $z$, with $z_{tf} = 1$ if gaze point $t$ is included in fixation $f$, and 0 otherwise.

**Proposition 1** *The constraint set*

$$\sum_{j=t+1}^{\mathcal{T}} z_{jf} \leq (\mathcal{T}-t)(1-z_{tf}+z_{t+1,f}),\ t=1,\ldots,\mathcal{T}-1;\ f\ =\ 1,\ldots,\mathcal{F}$$

*ensures that every fixation $f$ has only consecutive gaze points and terminates at most once.*

A proof of Proposition 1 can be found in Appendix 7.1. For a fixation $f$ starting at time point $p$ and concluding at $q$, the constraint set in Proposition 1 ensures in a linear fashion that $\underset{t:1\leq t<p}{z_{t,f}} = 0$, $\underset{t:p\leq t\leq q}{z_{t,f}} = 1$, and $\underset{t:q<t\leq\mathcal{T}}{z_{t,f}} = 0$. Moreover, this is accomplished with $\mathcal{T}\mathcal{F}-\mathcal{F}$ additional constraints, and no new variables.

### 3.2.3 Controlling Inner-Density of Fixations

Given that fixation identification is somewhat subjective in nature, all automated classification methods require some interpretation. Fixations properties can fluctuate as the task and stimulus vary. To account for this, we incorporate a nonnegative parameter $\alpha$ that acts to balance the tradeoff between the inclusion of additional gaze points and the spatial concentration of gaze points within fixations. This is done by incorporating the following term in the objective function:

$$\sum_{f=1}^{\mathcal{F}}\sum_{t=1}^{\mathcal{T}} \alpha(1-z_{tf}), \tag{6}$$

where larger $\alpha$ values provide greater incentive (that is, greater penalty) to include additional fixation points, at the expense of spatial proximity. Fixation inner-density can thereby be controlled by adjusting the level of $\alpha$. As $\alpha$ increases, there is additional incentive to cluster points, with $\alpha\to\infty$ tantamount to clustering all points (as in Terzi (2006)).

## 3.3 Mathematical Modeling

We next present two optimization-based formulations that make use of these three key insights to identify fixations in gaze data chunks by optimizing for density. The first formulation bears some resemblance to a clustering approach proposed by Rao (1971), which has the advantage of finding 2D fixations with no strong regard for their shape. The formulation is nonlinear and rather than using general mixed-integer nonlinear programming solvers such as BARON and SCIP, we pursue the strategy of linearization to efficiently solve the formulation. The second is an original, linear formulation that we develop, and has a related goal of bounding fixations with a square box of minimal diameter. We note that the following mathematical programming formulations are valid for any values of $\mathcal{T}$ and $\mathcal{F}$, notably

including smaller values that arise from the output of the decomposition principle described in Section 3.2.1, i.e. a single chunk $\mathcal{C}^k$.

### 3.3.1 MINLP Formulation: Minimize Average Intra-Fixation Sum of Distances

The main idea of this formulation is to ensure that fixations are constructed by minimizing the average intra-fixation sum of distances. Whereas every point must have a cluster assignment in Rao (1971), in our formulation we enable gaze points to be selected for a fixation only when it improves the objective of optimizing the density-based metric – it is not necessary to include every data point in a given chunk. To offset the tendency to select fixations of minimum duration, we incorporate the idea in (6) to balance the tradeoff between highly compact clusters and non-inclusion. Our formulation uses values $d_{ij}$ as the Euclidean distances between two data points $i$ and $j$, $i < j$, and $\mathcal{N}$ is the minimum number of gaze points that could reasonably constitute a fixation.

$$\text{minimize} \quad \sum_{f=1}^{\mathcal{F}} \left[ \frac{\sum_{i=1}^{\mathcal{T}-1} \sum_{j=i+1}^{\mathcal{T}} d_{ij} z_{if} z_{jf}}{\sum_{t=1}^{\mathcal{T}} z_{tf}} + \alpha \sum_{t=1}^{\mathcal{T}} (1 - z_{tf}) \right] \tag{7a}$$

$$\text{subject to} \quad \sum_{f=1}^{\mathcal{F}} z_{tf} \leq 1, \ t = 1, \ldots, \mathcal{T}, \tag{7b}$$

$$\sum_{t=1}^{\mathcal{T}} z_{tf} \geq \mathcal{N}, \ f \ = \ 1, \ldots, \mathcal{F}, \tag{7c}$$

$$\sum_{j=t+1}^{\mathcal{T}} z_{jf} \leq (\mathcal{T} - t)(1 - z_{tf} + z_{t+1,f}), \ t = 1, \ldots, \mathcal{T} - 1; \ f \ = \ 1, \ldots, \mathcal{F}, \tag{7d}$$

$$z_{tf} \in \{0, 1\}, \ t = 1, \ldots, \mathcal{T}, \ f = 1, \ldots, \mathcal{F}. \tag{7e}$$

Constraint set (7b) ensures that gaze points are assigned to at most one fixation. Constraint set (7c) ensures a fixation contains at least $\mathcal{N}$ points, and as per Proposition 1, constraint set (7d) ensures a fixation concludes at most once. Objective function (7a) contains two terms, one resembling the objective of Rao (1971), and a second that incentivizes inclusion of gaze points into fixations. Rather than $d_{ij}^2$ as in Rao (1971), we use a simpler objective term of $d_{ij}$ (this effect can be offset by adjusting the level of $\alpha$). This formulation has $\mathcal{T}\mathcal{F}$ binary variables and $\mathcal{T}\mathcal{F} + \mathcal{T}$ linear constraints. The specific instance with $\alpha$ very large and $\mathcal{N} = 1$ yields a model that can solve the sequence segmentation problem of Terzi (2006).

The first term of the objective function is nonlinear and fractional. In addition to containing the ratio of variable terms, it has a bilinear product component $z_{if} z_{jf}$ in the numerator. This bilinearity can be linearized by introducing variables $y_{ijf} \in \mathbb{R}_+$ equal to the product of

$z_{if}z_{jf}$, enforced implicitly via the following three constraint sets:

$$y_{ijf} \leq z_{if}, \quad i = 1, \ldots, \mathcal{T}-1;\ j = i+1, \ldots, \mathcal{T};\ f = 1, \ldots, \mathcal{F}, \tag{8a}$$

$$y_{ijf} \leq z_{jf}, \quad i = 1, \ldots, \mathcal{T}-1;\ j = i+1, \ldots, \mathcal{T};\ f = 1, \ldots, \mathcal{F}, \tag{8b}$$

$$y_{ijf} \geq z_{if} + z_{jf} - 1, \quad i = 1, \ldots, \mathcal{T}-1;\ j = i+1, \ldots, \mathcal{T};\ f = 1, \ldots, \mathcal{F}. \tag{8c}$$

The remaining nonlinear fractional term of the objective can be linearized through an approach similar to Wu (1997) and Trapp et al. (2010). Define $u_f = \frac{1}{\sum_{t=1}^{\mathcal{T}} z_{tf}},\ f = 1, \ldots, \mathcal{F}$. Continuous variable $u_f$ has a lower bound of $1/\mathcal{T}$ and, from (7c), an upper bound of $1/\mathcal{N}$. This gives a new objective function of:

$$\sum_{f=1}^{\mathcal{F}} \sum_{i=1}^{\mathcal{T}-1} \sum_{j=i+1}^{\mathcal{T}} d_{ij} y_{ijf} u_f, \tag{9}$$

which remains nonlinear. As $y_{ijf}$ takes a binary value and $u_f$ is a bounded continuous variable, this product can be further linearized in a manner similar to (8a)–(8c). Define continuous variable $v_{ijf}$ to be the product of $y_{ijf}u_f$. We can enforce this relationship implicitly through the following four constraint sets:

$$v_{ijf} \leq \frac{1}{\mathcal{N}} y_{ijf}, \quad i = 1, \ldots, \mathcal{T}-1;\ j = i+1, \ldots, \mathcal{T};\ f = 1, \ldots, \mathcal{F}, \tag{10a}$$

$$v_{ijf} \geq \frac{1}{\mathcal{T}} y_{ijf}, \quad i = 1, \ldots, \mathcal{T}-1;\ j = i+1, \ldots, \mathcal{T};\ f = 1, \ldots, \mathcal{F}, \tag{10b}$$

$$v_{ijf} \leq u_f - \frac{1}{\mathcal{T}}(1 - y_{ijf}), \quad i = 1, \ldots, \mathcal{T}-1;\ j = i+1, \ldots, \mathcal{T};\ f = 1, \ldots, \mathcal{F}, \tag{10c}$$

$$v_{ijf} \geq u_f - \frac{1}{\mathcal{N}}(1 - y_{ijf}), \quad i = 1, \ldots, \mathcal{T}-1;\ j = i+1, \ldots, \mathcal{T};\ f = 1, \ldots, \mathcal{F}. \tag{10d}$$

Lastly, it is important to ensure that $u_f$ is indeed the reciprocal of $\sum_{t=1}^{\mathcal{T}} z_{tf}$. Suppose $\sum_{t=1}^{\mathcal{T}} z_{tf} = \mathcal{P}_f$. To ensure $u_f = 1/\mathcal{P}_f$, we can use a procedure similar to that of Trapp et al. (2010), but leverage existing auxiliary variables so as to avoid creating additional variables. In particular, the $y_{ijf}$ and $v_{ijf}$ variables are defined for $i < j$, and it is not difficult to show that $\sum_{i=1}^{\mathcal{T}-1} \sum_{j=i+1}^{\mathcal{T}} y_{ijf} = \frac{\mathcal{P}_f \cdot (\mathcal{P}_f - 1)}{2}$. To enforce the $u_f = 1/\mathcal{P}_f$ relationship, we multiply both sides by the aforementioned expression involving the $y_{ijf}$ variables, to obtain the equivalent $\sum_{i=1}^{\mathcal{T}-1} \sum_{j=i+1}^{\mathcal{T}} v_{ijf} = \frac{\mathcal{P}_f \cdot (\mathcal{P}_f - 1)}{2} \frac{1}{\mathcal{P}_f} = \frac{\mathcal{P}_f - 1}{2}$. Rewriting this expression yields:

$$2 \sum_{i=1}^{\mathcal{T}-1} \sum_{j=i+1}^{\mathcal{T}} v_{ijf} - \sum_{t=1}^{\mathcal{T}} z_{tf} = -1,\ f = 1, \ldots, \mathcal{F}. \tag{11}$$

The final, linearized reformulation is:

$$\text{minimize} \quad \sum_{f=1}^{\mathcal{F}} \left[ \sum_{i=1}^{\mathcal{T}-1} \sum_{j=i+1}^{\mathcal{T}} d_{ij} v_{ijf} + \alpha \sum_{t=1}^{\mathcal{T}} (1 - z_{tf}) \right], \tag{12a}$$

$$\text{subject to} \quad \text{(7b), (7c), (7d), (8a), (8b), (8c), (10a), (10b), (10c), (10d), (11),} \tag{12b}$$

$$\frac{1}{\mathcal{T}} \leq u_f \leq \frac{1}{\mathcal{N}}, \ f = 1, \ldots, \mathcal{F}, \tag{12c}$$

$$0 \leq v_{ijf} \leq \frac{1}{\mathcal{N}}, \ i = 1, \ldots, \mathcal{T} - 1; \ j = i + 1, \ldots, \mathcal{T}; \ f = 1, \ldots, \mathcal{F}, \tag{12d}$$

$$z_{tf} \in \{0, 1\}, \ t = 1, \ldots, \mathcal{T}, \ f = 1, \ldots, \mathcal{F}, \tag{12e}$$

$$y_{ijf} \in \{0, 1\}, \ i = 1, \ldots, \mathcal{T} - 1; \ j = i + 1, \ldots, \mathcal{T}; \ f = 1, \ldots, \mathcal{F}. \tag{12f}$$

Formulation (12a)–(12f) has $\mathcal{T}^2\mathcal{F}$ binary variables, $\mathcal{T}^2\mathcal{F} - \mathcal{T}\mathcal{F} + \mathcal{F}$ continuous variables, and $7\mathcal{T}^2\mathcal{F} - 6\mathcal{T}\mathcal{F} + \mathcal{T} + \mathcal{F}$ linear constraints (not including simple variable bounds). While substantially larger than formulation (7a)–(7e), it has the advantage of being linear and therefore amenable to powerful commercial solvers such as Gurobi (Gurobi Optimization 2016). An analytical discussion of the sensitivity with respect to $\alpha$ appears in Appendix 7.2.

### 3.3.2 MIP Formulation: Minimize Square Area of Fixations

We now present our second formulation for finding dense fixations. It attempts to balance enveloping the largest number of points with a 2D square of minimal area, as measured by the side length $r$. As in the first formulation, the model is parametrized by the expression described in (6).

$$\text{minimize} \quad \sum_{f=1}^{\mathcal{F}} \left[ r_f + \alpha \sum_{t=1}^{\mathcal{T}} (1 - z_{tf}) \right], \tag{13a}$$

$$\text{subject to} \quad \text{(7b), (7c), (7d),} \tag{13b}$$

$$x_f - r_f - \mathcal{M}_x(1 - z_{tf}) \leq x^t \leq x_f + r_f + \mathcal{M}_x(1 - z_{tf}), \ t = 1, \ldots, \mathcal{T}, \tag{13c}$$

$$y_f - r_f - \mathcal{M}_y(1 - z_{tf}) \leq y^t \leq y_f + r_f + \mathcal{M}_y(1 - z_{tf}), \ t = 1, \ldots, \mathcal{T}, \tag{13d}$$

$$l_x \leq x_f \leq u_x; \ l_y \leq y_f \leq u_y, \ f = 1, \ldots, \mathcal{F}, \tag{13e}$$

$$r_f, x_f, y_f \in \mathbb{R}, \ f = 1, \ldots, \mathcal{F}; \ z_{tf} \in \{0, 1\}, \ t = 1, \ldots, \mathcal{T}, \ f = 1, \ldots, \mathcal{F}. \tag{13f}$$

The model has binary variables $z_{tf}$ for assigning time point $t$ to fixation $f$, continuous variables $x_f$ and $y_f$ that indicate the center of fixation $f$, and continuous variables $r_f$ that indicate the (half-)length of the side of the square bounding box (also known as the apothem

length). Bounds for $x_f$ and $y_f$ are constructed using $l_x = \min_t x^t$, $u_x = \max_t x^t$, $l_y = \min_t y^t$, and $u_y = \max_t y^t$, and further we set the values of $\mathcal{M}_x = \max\{|x^t - l_x|, |u_x - x^t|\}$ and $\mathcal{M}_y = \max = \{|y^t - l_y|, |u_y - y^t|\}$. Constraints (13c)–(13d) are box constraints to ensure that, if time point $t$ is assigned to fixation $f$ (i.e., $z_{tf} = 1$), then it lies geometrically within the appropriate square with side length $r_f$. Again, constraints (13b) represent the fundamental constraints that simply ensure, respectively, no time point is assigned to more than one fixation, every fixation contains a minimum number of points, and every fixation is composed of consecutive time points. Variable definitions and bounds are given in (13e)–(13f), while objective (13a) minimizes the total square fixation area, while the $\alpha$ term accounts for the tradeoff on the number of points included. Similar to formulation (12a)–(12f), an analytical discussion of the sensitivity with respect to $\alpha$ appears in Appendix 7.2.

### 3.4 Algorithm to Identify Densest Fixations

We provide an algorithmic approach to identify the densest fixations from a sequence $\mathcal{S}$ of gaze points using one of optimization formulations (12a)–(12f) or (13a)–(13f).

**Algorithm 1** Identify Densest Fixations

**Input:** Sequence $\mathcal{S}$ separated into distinct chunks of consecutive $(x, y, t)$ data $\mathcal{C}^k,\ k = 1, \ldots, \mathcal{K}$; parameter $\alpha$.

Set $\mathcal{L} \leftarrow \emptyset$.
**for** $k = 1, \ldots, \mathcal{K}$ **do**
  Set $\mathcal{T} \leftarrow \mathcal{T}^k$.
  **for** $\mathcal{F} = \mathcal{F}^k_{min}, \ldots, \mathcal{F}^k_{max}$ **do**
    With $\alpha$, formulate and solve mixed-integer program (12a)–(12f) or (13a)–(13f).
    **if** optimal solution found **then**
      Add solution to $\mathcal{L}$.
**return** $\mathcal{L}$.

Algorithm 1 processes all $(x, y, t)$ gaze-data chunks $\mathcal{C}^k,\ k = 1, \ldots, \mathcal{K}$ from sequence $\mathcal{S}$ into constituent fixations by optimizing for density using formulation (12a)–(12f) or (13a)–(13f). Initially $\mathcal{L}$ is empty, and by sequentially iterating over each chunk $\mathcal{C}^k,\ k = 1, \ldots, \mathcal{K}$, it sets $\mathcal{T}$ to the total number of gaze points $\mathcal{T}^k$ in chunk $\mathcal{C}^k$, and then formulates an optimization problem for every level of $\mathcal{F}$. For each chunk $\mathcal{C}^k$, fixations of maximum density (with respect to $\alpha$) are recorded and stored in $\mathcal{L}$.

## 4. Computational Experiments

We now proceed to discuss the computational performance of using Algorithm 1 to sequentially call formulations (12a)–(12f) and (13a)–(13f), on real gaze datasets from two tasks that

differ with respect to cognitive effort: online shopping (Shah et al. 2016), and solving math problems (Shojaeizadeh et al. 2016). The shopping task requires participants to purchase three items in a simulated grocery store environment, while the math task requires participants to answer a set of Graduate Record Examination Math Section questions. The task of reading and processing Math GRE questions by nature requires a higher level of information processing than the shopping task, hence it is more cognitively complex.

## 4.1 Datasets and Equipment

We considered two datasets, one containing eye movement data from the shopping task and one from the math task. Each dataset contains $\mathcal{R} = 10$ eye-tracking recordings (indexed by $\ell$). Participants were recruited from the student population in a Northeastern university of the United States. The first (shopping task) dataset was recorded by a Tobii Pro X2-30 eye tracker (Tobii 2018), with a frequency of 30 Hz. Each recording is between seven and twelve minutes in duration. The second (math task) dataset was recorded by a Tobii Pro TX300 (Tobii 2018). Each recording is approximately five minutes in duration. This dataset was originally recorded at 300 Hz. The two datasets are available for download at the following site: `http://uxdm.wpi.edu/data/Data_IJOC_2018_Public.zip`.

To compare the fixation patterns between shopping and math tasks, we also downsampled this dataset for each recording by retaining the first gaze point, and every tenth point thereafter, thereby generating a new reading dataset at 30 Hz. All experiments were run on an Intel core i7-4700MQ computer with 2.40 GHz and 8.0 GB RAM running 64-bit Windows 8. We used the Gurobi Optimizer (Gurobi Optimization 2016) with Python 2.7 interface for the optimization modeling, algorithmic design, and solution process, and note that we explicitly pursue global optima for each optimization problem by using default values for the Gurobi MIPGap (1e-4) and MIPGapAbs (1e-10) parameters. MATLAB was used for designing the I-DT filter (MathWorks 2016), while Tobii Studio was used for the I-VT filter (Olsen 2012). A time limit of 12 hours (wall-clock) was present for all computational experiments.

## 4.2 Data Preprocessing

For each recording $\ell$, gaze data is preprocessed, as discussed in Section 3.2.1, by separating the data sequence $\mathcal{S}_\ell$ into chunks $C_\ell^k,\ k = 1, \ldots, \mathcal{K}_\ell$, via saccadic events. We used the Tobii Studio I-VT filter (Salvucci and Goldberg 2000, Olsen 2012) to do so, together with a constant velocity threshold of $V = 30^\circ/s$, which is suitable for a variety of types of data under different sampling frequencies and noises (Olsen 2012). Because the I-VT processing can result in one or more consecutive, non-saccadic gaze points with total duration below a theo-

retical minimum fixation duration (which we take to be $100ms$; see, e.g., Salojärvi et al. 2005, Blignaut 2009, Komogortsev et al. 2010), we also preemptively removed these from consideration. Note that the further processing of chunks using formulations (12a)–(12f) and (13a)–(13f) serves to *refine* the results of the I-VT filter by optimizing for fixation inner-density.

| Stimuli | Frequency (Hertz) | Avg # of All Points in Sequence | Avg # of Data Chunks | Avg # of Valid Data Chunks | Avg # of Points in All Data Chunks | Avg # of Points in Valid Data Chunks |
|---|---|---|---|---|---|---|
| Shopping Data | 30 | 18,207 | 3,017 | 1,178 | 10,153 | 7,737 |
| GRE Math Reading Data | 30 | 9,058 | 752 | 575 | 8,092 | 6,822 |
| GRE Math Reading Data | 300 | 90,580 | 3,612 | 721 | 80,956 | 66,677 |

Table 1: Summary results on separated data with I-VT filter, averaged over ten recordings per dataset.

The minimum number of gaze points for a fixation is dependent on the frequency $h$ (in Hertz) of the eye-tracking device. From the literature, fixation durations are typically estimated in the range of $60 - 400ms$; in general a minimum duration $d_m = 100ms$ is a reasonable lower-bound for information processing to occur (Salojärvi et al. 2005, Blignaut 2009, Komogortsev et al. 2010). A straightforward choice of $\mathcal{N}$ is then $\mathcal{N} = \left\lceil \frac{h \cdot d_m}{1{,}000ms} \right\rceil$. Using this we set the minimum number of gaze points to be $\mathcal{N} = 3$ and $\mathcal{N} = 30$ for the 30 Hz and 300 Hz datasets, respectively. Table 1 details summary results on the processed sequences prior to, and after, removing these small sets of points; we term as *valid* those chunks (and points) that remain after removal. In general, lower values of $\mathcal{N}$ result in smaller, more numerous data chunks for a given data sequence. After preprocessing each $\mathcal{S}_\ell,\ \ell = 1, \ldots, \mathcal{R}$, into chunks $C_\ell^k,\ k = 1, \ldots, \mathcal{K}_\ell$, we then run Algorithm 1 using one of formulations (12a)–(12f) or (13a)–(13f). We set $\mathcal{F}_{min}^k = \mathcal{F}_{max}^k = 1$ for all of our experiments, as after randomly selecting one of the ten 300 Hz recordings, manual inspection strongly indicated that data chunks predominantly contain a single fixation.

## 4.3 Evaluation Metrics

We use several metrics to evaluate the performance of our methods for each dataset and level of $\alpha$. The average *fixation duration* $\delta^{avg}$ of a sequence $\mathcal{S}$ measures, in seconds, the time spent in fixations, averaged over all fixations. The *cover rate* $\gamma$ of a data sequence $\mathcal{S}$ measures the ratio of the number of gaze points included in fixations, to the total number of gaze points (fixation and non-fixation) in a given data instance; Blignaut (2009) also reports this measurement ("the percentage of points-of-regard that are included in fixations"). Cumulative computational runtimes are also recorded, in seconds of wall-clock time, for both the Gurobi Optimizer and Algorithm 1. Each of the metrics we consider are averaged over all ten recordings for each respective dataset.

Figure 4 is a small illustrative example depicting the duration and cover rate. It depicts a small data chunk (outer bounding region) containing eight gaze points obtained from 30 Hz data. The inner fixation, namely gaze points 2 through 7, has a duration of $\delta = \frac{6}{30} = 0.2$ seconds. Supposing that the length of this recording was 8 total points, the cover rate is $\gamma = \frac{6}{8} = 0.75$, because six of the eight points were included in the fixation.

We also consider three distinct representations of density. The paper of Rao (1971) advocates for *minimizing* the average intra-fixation sum of distances, a measure that is inversely proportional to density (so, effectively, the optimization *maximizes density*). Hence, to keep with this convention we present our results from this perspective – the three expressions we use to characterize density are such that *smaller* magnitudes represent *greater* density.

The first of these metrics ($\rho_1$) is the average pairwise distances between points within a fixation. Suppose that $\mathcal{P}$ is the number of points contained in the fixation, $\mathcal{P} > 2$, and $d_{pq}$ is the Euclidean distance between fixation points $p$ and $q$. Then $\rho_1$ is expressed as:

$$\rho_1 = \frac{\sum_{p=1}^{\mathcal{P}-1} \sum_{q=p+1}^{\mathcal{P}} d_{pq}}{\binom{\mathcal{P}}{2}}. \qquad (\rho_1)$$

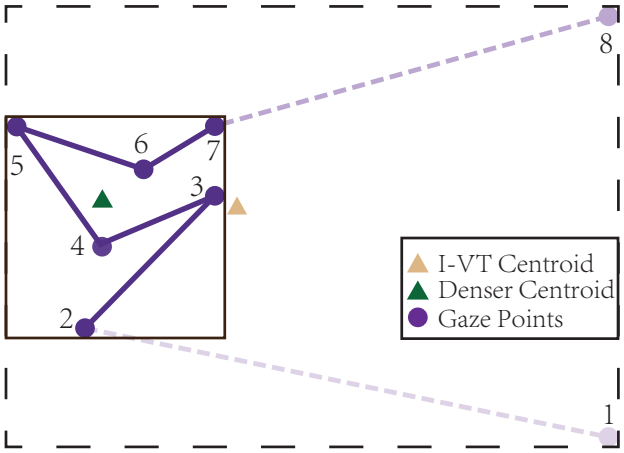



Figure 4: Duration $\delta$ and cover rate $\gamma$ for a single chunk. Our results refine those of I-VT, including only six of the eight points, yielding a denser fixation. In addition to duration and cover rate differences, the centroid shifting is also apparent.

The second metric $\rho_2$ is similar to $\rho_1$. It has the same numerator of summing the pairwise distances of all included fixation points, though the denominator is simply $\mathcal{P}$, which has the effect of increasing the density for fixations with greater number of points:

$$\rho_2 = \frac{\sum_{p=1}^{\mathcal{P}-1} \sum_{q=p+1}^{\mathcal{P}} d_{pq}}{\mathcal{P}}. \qquad (\rho_2)$$

We consider $\rho_2$ because of its clear relationship to objective function (12a) when $\alpha = 0$. It is meaningful to see how this metric varies under differing values of $\alpha$.

The third metric ($\rho_3$) is the minimal area square bounding box surrounding the fixation divided by the number of fixation points it contains:

$$\rho_3 = \frac{(2\hat{r})^2}{\mathcal{P}}. \qquad (\rho_3)$$

| | 30 Hz Shopping Data | | | | | | | | 30 Hz GRE Math Reading Data | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Duration | Density Measures | | | Cover Rate | Center Shift | Avg Runtime (s) | | Duration | Density Measures | | | Cover Rate | Center Shift | Avg Runtime (s) | |
| $\alpha$ | $\delta^{avg}$ (s) | $\rho_1^{avg}$ | $\rho_2^{avg}$ | $\rho_3^{avg}$ | $\gamma^{avg}$ | $\lambda^{avg}$ | Gurobi | Overall | $\delta^{avg}$ (s) | $\rho_1^{avg}$ | $\rho_2^{avg}$ | $\rho_3^{avg}$ | $\gamma^{avg}$ | $\lambda^{avg}$ | Gurobi | Overall |
| 0 | 0.1000 | 21.3300 | 21.3300 | 528.3056 | 0.1877 | 6.9322 | 39.2 | 55.6 | 0.1000 | 5.2815 | 5.2815 | 94.6009 | 0.2637 | 2.7312 | 131.0 | 151.3 |
| 3 | 0.1005 | 21.3200 | 21.3669 | 528.2375 | 0.1888 | 6.9244 | 129.6 | 146.4 | 0.1901 | 5.8623 | 10.9157 | 95.6769 | 0.5052 | 2.2796 | – | – |
| 6 | 0.1075 | 21.3647 | 22.3935 | 527.9422 | 0.2040 | 6.8027 | 1,796.2 | 1,812.9 | 0.2496 | 6.6952 | 18.2435 | 98.9420 | 0.6674 | 1.4921 | – | – |
| 9 | 0.1287 | 21.8642 | 27.2219 | 530.6561 | 0.2493 | 6.3368 | 1,911.2 | 1,928.1 | 0.2658 | 7.1121 | 21.7909 | 101.8035 | 0.7096 | 1.0694 | – | – |
| 12 | 0.1500 | 22.6585 | 33.8933 | 537.9728 | 0.2946 | 5.6161 | 356.6 | 373.4 | 0.2722 | 7.3488 | 23.7653 | 103.7606 | 0.7258 | 0.8422 | 784.6 | 804.7 |
| 15 | 0.1655 | 23.4458 | 40.1121 | 547.7675 | 0.3268 | 4.8396 | 216.8 | 233.8 | 0.2747 | 7.4884 | 24.7490 | 105.3714 | 0.7319 | 0.7191 | 507.9 | 528.0 |
| 18 | 0.1752 | 24.1153 | 44.9136 | 557.6114 | 0.3466 | 4.1185 | 178.3 | 195.2 | 0.2759 | 7.5881 | 25.3391 | 106.7105 | 0.7345 | 0.6306 | 375.2 | 395.3 |
| 21 | 0.1821 | 24.7263 | 48.9410 | 569.2495 | 0.3606 | 3.4843 | 183.0 | 199.9 | 0.2765 | 7.6401 | 25.7045 | 107.6167 | 0.7359 | 0.5857 | 219.6 | 239.6 |
| 24 | 0.1871 | 25.2430 | 52.2651 | 580.8938 | 0.3704 | 2.9871 | 164.8 | 181.6 | 0.2768 | 7.7087 | 25.9153 | 109.1708 | 0.7365 | 0.5390 | 24.8 | 44.8 |
| 27 | 0.1906 | 25.6653 | 54.9576 | 591.0722 | 0.3774 | 2.5722 | 136.9 | 153.8 | 0.2770 | 7.7516 | 26.0471 | 110.1686 | 0.7369 | 0.5068 | 16.5 | 36.6 |
| 30 | 0.1934 | 26.0326 | 57.3728 | 601.0251 | 0.3831 | 2.1891 | 100.5 | 117.6 | 0.2772 | 7.7877 | 26.1815 | 111.2469 | 0.7372 | 0.4865 | 12.7 | 32.9 |

Table 2: Results of Algorithm 1 & formulation (12a)–(12f) on 30 Hz shopping and GRE Math reading datasets.

| | 30 Hz Shopping Data | | | | | | | | 30 Hz GRE Math Reading Data | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Duration | Density Measures | | | Cover Rate | Center Shift | Avg Runtime (s) | | Duration | Density Measures | | | Cover Rate | Center Shift | Avg Runtime (s) | |
| $\alpha$ | $\delta^{avg}$ (s) | $\rho_1^{avg}$ | $\rho_2^{avg}$ | $\rho_3^{avg}$ | $\gamma^{avg}$ | $\lambda^{avg}$ | Gurobi | Overall | $\delta^{avg}$ (s) | $\rho_1^{avg}$ | $\rho_2^{avg}$ | $\rho_3^{avg}$ | $\gamma^{avg}$ | $\lambda^{avg}$ | Gurobi | Overall |
| 0 | 0.1006 | 21.5012 | 21.6441 | 522.6540 | 0.1888 | 6.9010 | 8.2 | 44.2 | 0.1015 | 5.3435 | 5.4356 | 93.7965 | 0.2679 | 2.7093 | 7.7 | 40.4 |
| 1 | 0.1209 | 21.5645 | 26.7576 | 513.3601 | 0.2323 | 6.4350 | 10.3 | 46.6 | 0.2477 | 6.4484 | 19.5562 | 93.3753 | 0.6603 | 1.6307 | 6.4 | 39.6 |
| 2 | 0.1531 | 22.5559 | 37.6929 | 519.0029 | 0.3012 | 5.4906 | 9.7 | 46.3 | 0.2669 | 6.9389 | 23.0267 | 89.6347 | 0.7120 | 1.1170 | 4.1 | 37.3 |
| 3 | 0.1711 | 23.5078 | 45.9917 | 529.5563 | 0.3389 | 4.5600 | 8.7 | 45.3 | 0.2719 | 7.1749 | 24.1726 | 91.7787 | 0.7249 | 0.8920 | 3.5 | 36.6 |
| 4 | 0.1812 | 24.2505 | 51.3489 | 541.7839 | 0.3595 | 3.8215 | 7.7 | 44.3 | 0.2740 | 7.3090 | 24.7753 | 93.4555 | 0.7299 | 0.7780 | 3.2 | 36.5 |
| 5 | 0.1873 | 24.8422 | 54.9765 | 553.5813 | 0.3714 | 3.2194 | 6.9 | 43.6 | 0.2752 | 7.4159 | 25.2025 | 95.1061 | 0.7327 | 0.7007 | 3.1 | 36.3 |
| 6 | 0.1908 | 25.2697 | 57.2888 | 562.2031 | 0.3785 | 2.7894 | 6.2 | 42.9 | 0.2759 | 7.4883 | 25.4862 | 96.3742 | 0.7344 | 0.6406 | 3.0 | 36.3 |
| 7 | 0.1932 | 25.6052 | 59.1002 | 571.0291 | 0.3834 | 2.4927 | 5.8 | 42.5 | 0.2763 | 7.5522 | 25.7068 | 97.5203 | 0.7354 | 0.5861 | 3.0 | 36.5 |
| 8 | 0.1951 | 25.9119 | 60.5458 | 580.0977 | 0.3871 | 2.1942 | 5.4 | 42.0 | 0.2765 | 7.5892 | 25.8133 | 98.1695 | 0.7359 | 0.5581 | 2.9 | 36.2 |
| 9 | 0.1965 | 26.1606 | 61.7261 | 588.6679 | 0.3897 | 1.9573 | 5.1 | 41.8 | 0.2769 | 7.6568 | 26.0204 | 99.6355 | 0.7366 | 0.5136 | 2.8 | 36.2 |
| 10 | 0.1977 | 26.3796 | 62.7787 | 597.2041 | 0.3919 | 1.7403 | 4.9 | 41.7 | 0.2770 | 7.6795 | 26.1019 | 100.1463 | 0.7368 | 0.4983 | 2.8 | 36.2 |

Table 3: Results of Algorithm 1 & formulation (13a)–(13f) on 30 Hz shopping and GRE Math reading datasets.

The minimal square side length $2\hat{r}$ is derived from the optimal $\hat{r}$ value in optimization formulation (13a)–(13f). A final metric, the *center shift* $\lambda^{avg}$, is reported in more detail in Section 4.5.2, in particular with respect to the performance of our approaches versus the standard I-VT filter.

## 4.4 Computational Results and Discussion On Proposed Methods

We now discuss the results of our computational experiments for our proposed methods. Table 2 highlights computational results from running formulation (12a)–(12f) on 30 Hz Shopping data (left) and 30 Hz GRE Math reading data (right). Table 3 documents the same information as Table 2, but using formulation (13a)–(13f). Table 4 details the performance of formulation (13a)–(13f) on the larger 300 Hz dataset (formulation (12a)–(12f) was not competitive at this higher frequency). Each table has a similar format, with the rows indexed by trade-off parameter $\alpha$, and the columns indicating various properties discussed in Section 4.3, which are obtained post-optimization by averaging over all chunks in each of the ten data recordings.

| $\alpha$ | 300 Hz GRE Math Reading Data | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Duration | Density Measures | | | Cover Rate | Center Shift | Avg Runtime (s) | |
| | $\delta^{avg}$ (s) | $\rho_1^{avg}$ | $\rho_2^{avg}$ | $\rho_3^{avg}$ | $\gamma^{avg}$ | $\lambda^{avg}$ | Gurobi | Overall |
| 0 | 0.1062 | 5.8589 | 90.1959 | 31.9361 | 0.2598 | 1.8150 | 574.3 | 659.5 |
| 0.1 | 0.2607 | 6.5335 | 241.3585 | 28.8872 | 0.6528 | 0.9478 | 364.5 | 454.1 |
| 0.2 | 0.2762 | 6.7828 | 268.4264 | 28.5850 | 0.6911 | 0.6739 | 264.7 | 354.7 |
| 0.3 | 0.2803 | 6.8764 | 277.5209 | 28.2034 | 0.7004 | 0.5727 | 207.2 | 299.7 |
| 0.4 | 0.2827 | 6.9654 | 283.6307 | 27.5299 | 0.7053 | 0.5046 | 154.7 | 246.6 |
| 0.5 | 0.2840 | 7.0202 | 287.1474 | 27.7181 | 0.7083 | 0.4589 | 119.0 | 212.0 |
| 0.6 | 0.2848 | 7.0571 | 289.3265 | 27.8777 | 0.7100 | 0.4300 | 87.0 | 178.1 |
| 0.7 | 0.2853 | 7.0816 | 290.6830 | 28.0161 | 0.7112 | 0.4095 | 67.1 | 159.0 |
| 0.8 | 0.2857 | 7.1100 | 292.1223 | 28.1589 | 0.7121 | 0.3880 | 53.9 | 145.1 |
| 0.9 | 0.2860 | 7.1251 | 292.7735 | 28.2548 | 0.7126 | 0.3777 | 43.4 | 136.5 |
| 1.0 | 0.2863 | 7.1483 | 294.0966 | 28.3347 | 0.7134 | 0.3612 | 37.7 | 128.8 |

Table 4: Results of Algorithm 1 & formulation (13a)–(13f) on 300 Hz GRE Math reading dataset.

The parameter $\alpha$ represents the trade-off in emphasis between the spatial compactness versus the number of gaze points contained in a given fixation. At one extreme, a level of $\alpha = 0$ gives no incentive for inclusion, so very compact fixations tend to form with minimal gaze points, that is, near the level of $\mathcal{N}$. At the other extreme, larger $\alpha$ penalties incentivize many gaze points to be included in the fixation, likely at the expense of spatial proximity. Tension exists in between these two extremes for gaze points that, while within a given data chunk $\mathcal{C}^k$, are not near the center of a fixation (see, e.g., the sixth gaze point in Figure 2a). Due to the intrinsic and distinct interpretations of density in (12a) versus (13a), differing levels of $\alpha$ are required to induce similar outcomes. For this reason we varied the range of $\alpha$ values in Tables 2, 3, and 4. Due to the higher frequency of the 300 Hz dataset, greater sensitivity with $\alpha$ was necessary (in the form of smaller values) to influence the results of Table 4.

### 4.4.1 Runtime Discussions

For each sequence $\mathcal{S}_\ell$, the runtime consists of solving an optimization problem for each valid chunk $\mathcal{C}_\ell^k,\ k = 1, \ldots, \mathcal{K}_\ell$. As can be seen in Table 1, on average this implies solving upwards of several hundreds, and sometimes thousands, of small yet still NP-hard optimization problems. Moreover, for every computational test, there is a roughly "constant" time for processing the same dataset. This can be seen in the difference in runtimes between the "Gurobi" and "Overall" columns, with "Overall" being fairly static. Thus, the differences in runtime are largely due to the contribution of Gurobi, which experiences varying levels of computational difficulty as $\alpha$ fluctuates. Moreover, Tables 2 and Table 3 exhibit the general trend that when $\alpha$ increases, the Gurobi runtime initially increases, and then decreases. This is apparent for both the 30 Hz shopping and GRE Math reading datasets, and for both

optimization formulations. This behavior is likely induced by $\alpha$: when $\alpha$ is rather small yet nonzero, there is relatively greater difficulty in balancing the trade-off term in the objective of including a point or adding the penalty.

Looking across Tables 2 and 3, in general formulation (12a)–(12f) exhibits a slower runtime performance than (13a)–(13f). When comparing the algorithmic performances of formulation (12a)–(12f) on shopping and GRE Math reading stimuli as reported in Table 2, we observe that the latter dataset exhibited much longer runtimes for several initial levels of $\alpha$. Generally speaking, many of the GRE Math reading data chunks were much larger than those from the shopping data. These larger data chunks, as well as the numerous new variables and constraints required to linearize formulation (12a)–(12f), are likely the reason that it returned no fixations for several levels of $\alpha$ where the proximity-duration trade-off was most difficult to balance.

Formulation (13a)–(13f) experienced no such performance degradation on the 30 Hz datasets detailed in Table 3. Even so, when comparing the runtimes for the 30 Hz and 300 Hz GRE Math reading data in Tables 3 and 4, formulation (13a)–(13f) exhibits slower performance on the 300 Hz instances. It can be seen from Table 1 that the 300 Hz instances have larger average chunk sizes. Hence, the longer processing times are likely due to Gurobi formulating and solving (13a)–(13f) on larger data chunks. These runtime results from Table 4, while larger than those from Table 3, remain quite promising for future fixation detection on similar datasets, and for those of longer duration and at higher frequencies.

#### 4.4.2 Fixation Duration and Cover Rate Discussions

Fixation duration $\delta$ is a commonly-used metric in eye-tracking research representing the temporal length of a fixation. For each dataset and formulation, we report in Tables 2, 3, and 4 the fixation duration averaged over all chunks and recordings, $\delta^{avg}$. When $\alpha = 0$, there is no incentive to include gaze points beyond the minimum necessary. Hence, the value of $\delta^{avg}$ approaches the minimum defined length of a fixation represented by $\mathcal{N}$. As $\alpha$ increases, the value of $\delta^{avg}$ also increases, indicating that on average, fixations are containing more gaze points. Moreover, independent of dataset and formulation, $\delta^{avg}$ experiences the greatest increase for relatively low values of $\alpha$.

The cover rate $\gamma$ is a measurement that describes the ratio of points included in fixations to the total points in a data recording. For each dataset and formulation, we report the cover rate averaged over all recordings, $\gamma^{avg}$. As $\alpha$ increases, $\gamma^{avg}$ exhibits an increasing trend in Tables 2, 3, and 4. Independent of the formulation, the largest $\gamma$ increases occur at slightly different values of $\alpha$. For the GRE Math reading data, the largest jump in $\gamma$

occurs immediately after $\alpha$ transitions from 0 to the first nonzero value. For the shopping data, however, the greatest $\gamma$ increase occurs somewhat subsequent to the initial nonzero $\alpha$ transition. After these larger jumps, $\gamma$ increases at a decreasing rate.

### 4.4.3 Density Metric Discussions

The three density metrics discussed in Section 4.3 are averaged over all chunks in each of the ten data recordings, and reported in Tables 2, 3, and 4. Recall that, in keeping with Rao (1971), density is largest for small $\rho_1$, $\rho_2$, and $\rho_3$ values. Some general trends across all experiments is that $\rho_1^{avg}$ never exceeds $\rho_2^{avg}$. This is a relatively straightforward observation because, while $\rho_1^{avg}$ and $\rho_2^{avg}$ have identical numerators, $\rho_1^{avg}$ always has at least as large of a denominator (and often larger). The $\rho_3^{avg}$ metric evaluates the ratio of the minimal bounding box area to the number of points in the fixation, hence is a slightly different metric and often differs in magnitude from $\rho_1^{avg}$ and $\rho_2^{avg}$.

For all datasets and formulations, the general trend is for $\rho_1^{avg}$, $\rho_2^{avg}$, and $\rho_3^{avg}$ to increase as $\alpha$ increases, implying that, on average, fixations decrease in density. For all three metrics, both the numerator and denominator will increase as $\alpha$ increases, hence there are some slight fluctuations as $\alpha$ varies, and among the three metrics, $\rho_3^{avg}$ exhibits the greatest variation for early values of $\alpha$. Another observation is that the difference between $\rho_1^{avg}$ and $\rho_2^{avg}$ increases as the value of $\alpha$ increases. This increase is largely attributed to the difference in denominators of $\rho_1^{avg}$ and $\rho_2^{avg}$. For the 300 Hz dataset, as can be seen in Table 4, there is a much larger difference between $\rho_1^{avg}$ and $\rho_2^{avg}$. This is again due to the linear versus quadratic nature of the denominators; with the 300 Hz dataset, the value of the minimum duration threshold $\mathcal{N}$ is much larger, implying that each fixation should contain many more points.

Another important observation is that, independent of formulation, the fixations in GRE Math reading data both exhibit greater density than those for the shopping data, as well as feature longer durations. As it is known that longer fixations are representative of higher levels of information processing (Djamasbi 2014), the results in our study give further support that the math task was cognitively more demanding than the shopping task. Moreover, our results also provide evidence that fixations for more cognitively complex tasks are denser than less demanding tasks. This in turn is a valuable insight for studies that use eye-tracking to capture information processing behavior at the physiological level.

## 4.5 Analysis of I-VT and I-DT Methods

In this section we evaluate the performance of the I-VT and the I-DT methods for the three datasets. Specifics related to our I-VT filter implementation were discussed in Sec-

tion 4.2. For the I-DT filter, we use a commonly used minimum duration threshold of $d_m = 100ms$ (Salvucci and Goldberg 2000, Salojärvi et al. 2005, Blignaut 2009, Komogortsev et al. 2010), measure with the same dispersion metric expressed in (1), and set the dispersion threshold $D$ at 1° of visual angle as recommended in Blignaut (2009).

### 4.5.1 Comparison of I-VT and I-DT Filters

Table 5 contains five evaluation metrics for each of the I-VT and I-DT filters. It reveals that, though the I-VT and I-DT filters have different approaches to identify fixations, they have relatively similar average fixation duration for the shopping data (0.2028 and 0.1939 seconds, respectively). The difference between average fixation duration for the math task calculated by these two methods, however, is quite large. Independent of the method used, average fixation duration for the math task is longer than the average fixation duration for the shopping task. This suggests that the math task required more intense attention than the shopping task.

| | I-VT Fixation Filter | | | | | I-DT Fixation Filter | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Dataset | $\delta^{avg}$ (s) | $\rho_1^{avg}$ | $\rho_2^{avg}$ | $\rho_3^{avg}$ | $\gamma^{avg}$ | $\delta^{avg}$ (s) | $\rho_1^{avg}$ | $\rho_2^{avg}$ | $\rho_3^{avg}$ | $\gamma^{avg}$ |
| 30 Hz Shopping Data | 0.2028 | 28.5099 | 70.0920 | 752.7819 | 0.4012 | 0.1939 | 25.9076 | 57.0907 | 388.4482 | 0.6875 |
| 30 Hz GRE Math Reading Data | 0.2778 | 8.6233 | 28.4239 | 273.2999 | 0.7385 | 0.6240 | 16.9943 | 152.8289 | 154.4743 | 0.9770 |
| 300 Hz GRE Math Reading Data | 0.2893 | 7.7921 | 318.7482 | 83.1764 | 0.7195 | 0.4699 | 12.4661 | 934.2180 | 34.5853 | 0.9092 |

Table 5: Performance of I-VT and I-DT filters on five metrics for chunks $\mathcal{C}^k,\ k = 1, \ldots, \mathcal{K}$ from our study. I-VT performance bears resemblance to results from larger values of $\alpha$ reported in Tables 2, 3, and 4.

No clear trend between the I-VT and I-DT filters was discernible among the $\rho_1^{avg}$, $\rho_2^{avg}$, and $\rho_3^{avg}$ metrics. We believe this to simply be due to the substantial design differences between the I-VT and I-DT filters. We also note that the $\gamma^{avg}$ values for the I-DT filter are larger than those of the I-VT filter for every dataset. Similarly, we attribute this to the I-DT filter tending to identify fixations of longer duration than the I-VT filter (in Andersson et al. 2016, a similar phenomenon was reported).

### 4.5.2 Comparing Our Methods with Existing Methods

Our final discussion compares the computational results of our proposed methods with existing methods, namely the I-VT and I-DT filters. It is important to note that many of our comparison metrics are based on fixation properties, which in turn depend upon their separation from saccadic events. Comparisons between the I-VT and I-DT methods are somewhat incongruent, as their differing implementations lead to fundamental differences in the way they identify fixations, including the total number of fixations and the fixation

durations. Even so, of the two, our methods most closely compare with the I-VT filter – as we use the velocity threshold strategy of the I-VT filter to divide the gaze sequence $\mathcal{S}$ into chunks. We explain this data preprocessing step in Section 4.2.

Most of the I-VT filter performance in Table 5 can be viewed through the lens of the parameter $\alpha$. In particular, the I-VT results resemble those of Tables 2, 3, and 4 for increasingly large values of $\alpha$. That is, because the straightforward I-VT implementation has no way of further reducing the chunks it identifies as fixations, it can be seen as the extreme of our methods, for very high $\alpha$ levels. Table 5 reveals that the I-VT metrics represent the limiting values for each of the five metrics, over all three datasets and both formulations. This gives credence to the idea that fixations identified by the I-VT filter often contain additional gaze points that should be viewed as outliers. Our methods are able to further filter these fringe points by optimizing for inner-density, thereby refining the classification results of the I-VT filter. The ability to use user-defined $\alpha$ values to distill larger chunks of data into more refined fixations is a key feature of our proposed fixation identification approaches, as these refined fixations represent the core of focused attention.

Concerning the I-DT filter, comparisons to our methods are at best indirect. Even so, we make comparisons as they may have some limited utility. For the 30 Hz shopping data, the I-DT filter performance yielded fixations with somewhat similar or slightly larger durations, having more total gaze points covered than the results from our formulations (12a)–(12f) and (13a)–(13f). Again, these effects may be attributable simply to the specifics of the I-DT filter design (in a related study, Andersson et al. 2016, similar observations are made). Moreover, on the aggregate our methods are generally able to identify fixations with greater density than those of the I-DT filter.

Last, we remark on refining the center location of a fixation. Having already observed that fixation duration is strongly influenced by the level of $\alpha$, which controls for inner-density, we now demonstrate that our approaches can fine-tune the locational precision of the I-VT method. We introduce the *center shift* $\lambda^{avg}$, which measures the straight-line (Euclidean) distance, in pixels, between the I-VT fixation centroid and the densest fixation centroid, averaged over all fixations. These values are reported in Tables 2, 3, and 4. It can be clearly seen that lower $\alpha$ values yield larger $\lambda^{avg}$ values than do higher $\alpha$ values. This is because smaller $\alpha$ values increase the inner-density of the resulting fixations, and in so doing, the fixation centroids become more centralized due to the exclusion of some peripheral points existing in data chunks.

In summary, our proposed fixation identification approach both builds on strengths of both the I-VT and I-DT filters, and avoids shortcomings. Velocity-based methods serve as a suitable method to group a gaze data sequence into fixation chunks by removing saccadic

points (as per the I-VT filter). Moreover, excluding consecutive gaze points for which the duration is below a realistic threshold is also a useful way to remove gaze points unrelated to fixations (similar to the I-DT filter). By optimizing for inner-density on each resulting data chunk, we essentially use a dispersion-based approach to identify fixations. A key difference is that, rather than a static threshold used in I-DT, our dispersion threshold is dynamic – this is directly expressed by the variable $r$, characterizing bounding square side length, that is minimized in formulation (13a)–(13f). By doing so, we minimize the inclusion of fringe points in fixations and thus improve the accuracy of fixation duration and location. Hence, our methods are a refinement of both approaches.

Our computational findings have important implications for eye-tracking research. First, they show that considering fixations at a more refined scale can provide important insights into cognitive processing levels, as our computational experiments reveal that tasks with greater cognitive complexity featured longer-lasting fixations with heightened density. Hence, the results provide a rationale and theoretical direction for studying behavior via a new metric in user experience and human-computer interaction studies. Additionally, our results demonstrate that inner-density is a valuable concept; when combined with optimization-based approaches, it is a useful and novel way to identify fixations. In particular, the inner-density parameter $\alpha$ provides a previously unavailable level of control for studying focused fixation, which we believe will prove fruitful in many fields of study that use fixation duration and location to identify behavior, including marketing, user experience, human-computer interaction, and medical diagnosis.

# 5. Conclusions

This paper addresses the task of identifying eye-movement fixation events in temporal $(x, y)$ gaze data obtained from eye-tracking devices. Fixations carry information about cognitive processing and thus their properties, including fixation duration and location, are often used to understand behavior (Poole and Ball 2005). Fixation duration refers to the temporal length of a fixation, calculated as the number of gaze points within a fixation divided by the recording frequency, whereas fixation location refers to the center (centroid) of a gaze point cluster identified as a fixation event. Both duration and location properties are sensitive to how gaze points are grouped into fixations because they are influenced by the number, and the spatial proximity, of gaze points in an individual fixation.

Common methods of identifying fixations, such as the I-VT and I-DT filters, can lead to issues with precision regarding duration and location of fixations. This can have unintended ramifications when exact, rather than approximate, oculomotor behavior location or duration

are key outcomes. Indeed, in some eye-movement applications such as psychopathological diagnoses (such as Autism spectrum disorder – see, e.g., Sabatos-DeVito et al. 2016, Thorup et al. 2016), it can be argued that such accuracy is essential.

Figure 5 contrasts the performance of the raw I-VT filter with the performance of formulation (13a)–(13f) and $\alpha = 0.1$ on the same data sequence depicted in Figure 1. The callouts denote saccadic points by stars, fixation points by circles, and points that are eliminated by our approach by triangles. The smallest 2D boundaries for both approaches are also drawn. Some I-VT fixations (e.g. Fixation 1) contain nearly 35% more points as compared to ours (66 vs. 49 points). This refinement can have a large affect on key gaze metrics such as fixation duration and center.

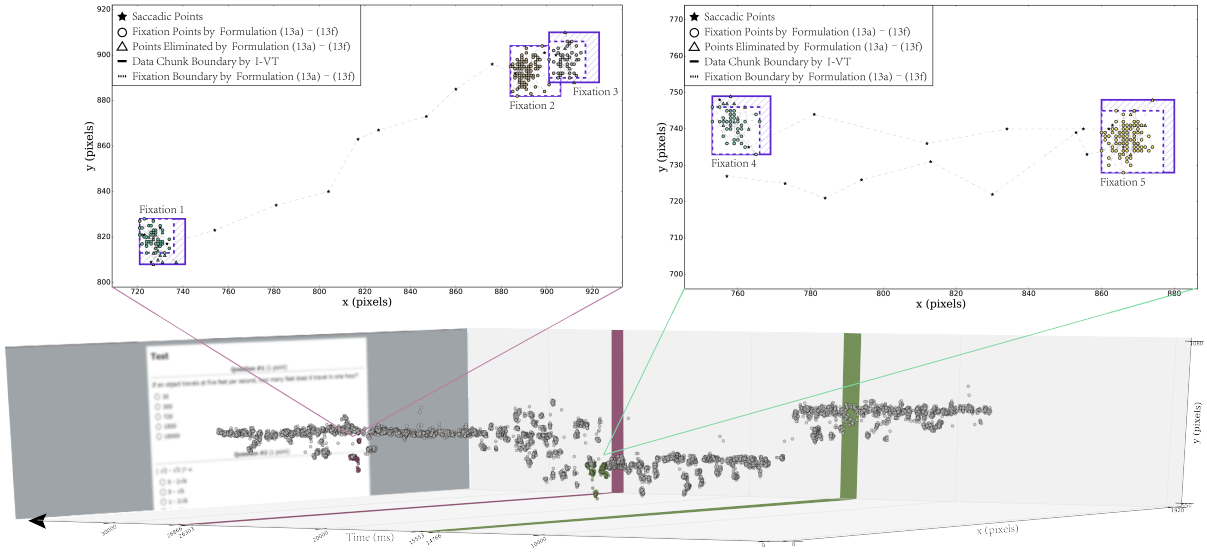



Figure 5: Comparing fixations identified with standard I-VT, versus formulation (13a)–(13f), $\alpha = 0.1$, in the gaze stream depicted in Figure 1. Some I-VT fixations contain nearly 35% more points than our approach.

Fixations indicate user effort to stabilize gaze when viewing an object, and the density of gaze points within a fixation carry information about user focus on that object; denser fixations represent more focused attention (Shojaeizadeh et al. 2016). This inner-density property of individual fixations, however, has gone largely unstudied in eye-tracking research. In this work we approach the problem of identifying fixations from the perspective of inner-density. Inner-density intrinsically values both duration of gaze, as well as the compactness of fixations, and so can reasonably represent focused processing. Additionally, because dense fixations place a high value on proximity, they are less likely to include outliers, and hence their duration and location measures are likely to be more accurate and representative. We provide several alternative mathematical programming formulations together with an algorithm to identify the densest fixations in a sequence of gaze data. To our best knowledge,

there are no explicit density-based approaches to identify fixations in gaze data, nor are there any that optimize for density.

Our computational experiments on two actual shopping and GRE Math reading datasets yielded encouraging results, in particular formulation (13a)–(13f) is quite robust to the larger 300 Hz GRE Math reading dataset over a variety of parametrized $\alpha$ values. The reasonable runtimes suggest further scalability for formulation (13a)–(13f). Moreover, both formulations are able to identify fixations with greater density than the standard I-VT filter, revealing that finer detail is available than what the I-VT can otherwise provide.

We note some limitations. Our formulations (12a)–(12f) and (13a)–(13f) can accommodate multiple dense fixations within a given data chunk, and especially for the latter, sequences of reasonably large size. Through limited manual inspections on individual data chunks separated by the I-VT filter, we overwhelmingly observed that a single actual dense fixation existed within a given data chunk. However, multiple fixations may occasionally exist within a single chunk. Here, the size of the formulations and the complexity in solving them also grow with the size of $\mathcal{F}_{max}$, so we would expect the runtime performance of our approaches to deteriorate as $\mathcal{F}_{max}$ increases. On the other hand, when seeking multiple fixations, symmetry exists in formulations (12a)–(12f) and (13a)–(13f), and hence symmetry-breaking constraints may help improving computational progress toward global optimality. A future avenue of work may involve ways to automatically detect the number of fixations within a given data chunk, and whether there exist statistical measures that can account, or even optimize, for these. We could also conduct side-by-side runtime performance comparisons of our methods with I-DT and I-VT filters, which require more information than is presently available in the data obtained from Tobii eye-tracking software (Olsen 2012). Future studies involving customized I-VT and I-DT implementations using eye-tracking SDKs are needed to enable such comparisons.

Formulation (12a)–(12f) does not perform well computationally, for at least some $\alpha$ values, when the number of gaze points exceeds approximately 100. We attribute this to the nonlinearity of the initial formulation, and subsequent increase in model size required by the linearization. On the other hand, formulation (13a)–(13f) appears to demonstrate greater scalability. However, a potential limitation is that it uses a square bounding box, when there is no specific reason to expect that fixations will be bounded by a quadrilateral of equal sides. Future work includes exploring alternative bounding regions such as circles, rectangles, and ellipses which, while likely more representative, require more expressive modeling through nonlinear representations of area.

Viewed at the two extremes of $\alpha = 0$ and very large $\alpha$, our formulations can be seen as either finding very compact fixations of short duration, or alternatively an essential equivalent

to the I-VT filter. In between these extremes, it will prove interesting to explore appropriate levels of $\alpha$ for varying tasks (e.g. recognition or selection) and stimuli (e.g. dynamic or interactive). This may be the focus of subsequent investigations. Another idea is to allow for the possibility of slightly relaxing the fairly restrictive temporal adherence conditions outlined in Proposition 1. This could replace that constraint set with an alternative that could allow for small deviations from absolute time consistency. Moreover, while the solution approaches presented in this paper specifically address 2D gaze data, the increasing trend of exploring and interpreting 3D gaze data (see, e.g., Blascheck et al. 2014) presents additional opportunities. Formulations (12a)–(12f) and (13a)–(13f) can both readily accommodate 3D gaze data with minor modifications (namely, adapting the $d_{ij}$ measure in the former to represent 3D distances, and incorporating a continuous variable for the dimension beyond $x$ and $y$ together with associated box constraints).

Finally, gaze data likely exhibits different tendencies when stimuli and task vary. We observed this when comparing the shopping versus GRE Math reading data – eye movement metrics such as average fixation duration and average cover rate had discernible changes across stimuli. Further, eye movement metrics are likely sensitive to subject variation. We believe our algorithmic approaches are extendible to recognize fixations under a variety of different scenarios. Moreover, because we focus only on refining fixations, our approaches pose no issue, and may even enhance, the identification of less studied eye movement events such as smooth pursuit and glissades.

# 6. Acknowledgments

The authors would like to thank the WPI User Experience and Decision Making (UXDM) Laboratory, and in particular Mina Shojaeizadeh for her assistance with recording the eye-tracking data used in this paper.

# 7. Appendices

## 7.1 Proof of Proposition 1

**Proof.** For any fixation $f$, the variables corresponding to every consecutive time pair ($z_{tf}$, $z_{t+1,f}$) can take one of four alternatives: *i*) $(0,0)$ is outside of $f$; *ii*) $(0,1)$ starts $f$; *iii*) $(1,1)$ is inside of $f$; and *iv*) $(1,0)$ terminates $f$. The constraint set causes no restriction for the first three alternatives, as every constraint is trivially satisfied with a right-hand side of $(\mathcal{T}-t)$ or $2(\mathcal{T}-t)$. The right-hand side constrains only for alternative iv), with $z_{tf}=1$ and $z_{t+1,f}=0$.

Here it becomes zero, immediately ensuring that $z_{t,f} = 0$ for all $t+1, \ldots, \mathcal{T}$. Constraints corresponding to $t = 1, \ldots, \mathcal{T}-1$ ensure this to be in effect over all pairs of time points, thereby disallowing fixation $f$ to terminate more than once. This completes the proof. ■

## 7.2 Sensitivity Analysis for Objective Function Density Terms

We consider the sensitivity of objective functions of formulations (12a)–(12f) and (13a)–(13f), for constant $\bar{\alpha}$, in balancing the tradeoff between including additional points in the fixation versus the spatial compactness of the fixation. We analyze a general chunk $\mathcal{C}^k$ with $\mathcal{T}^k$ gaze points. Without loss of generality, suppose $\mathcal{F} = 1$, so the fixation subscript $f$ may be omitted for notational simplicity. Further suppose a feasible solution $\bar{\mathcal{X}} = \{\bar{u}, \bar{v}, \bar{y}, \bar{z}\}$ exists for formulation (12a)–(12f), and consider the equivalent form of the objective found in (7a). The objective function value for feasible solution $\bar{\mathcal{X}}$ is:

$$\mathcal{Z}_{feas} = \frac{\sum_{i=1}^{\mathcal{T}^k-1} \sum_{j=i+1}^{\mathcal{T}^k} d_{ij} \bar{z}_i \bar{z}_j}{\sum_{i=1}^{\mathcal{T}^k} \bar{z}_i} + \bar{\alpha} \sum_{i=1}^{\mathcal{T}^k} (1 - \bar{z}_i). \tag{14}$$

Now suppose there exists a gaze point $\ell \in \mathcal{C}^k$ that is not presently contained in the fixation and immediately precedes the first gaze point in the fixation (or succeeds the last). We consider whether to add gaze point $\ell$. The new objective function would then be:

$$\mathcal{Z}_{new} = \frac{\sum_{i=1}^{\mathcal{T}^k-1} \sum_{j=i+1}^{\mathcal{T}^k} d_{ij} \bar{z}_i \bar{z}_j + \sum_{i:\bar{z}_i=1}^{\mathcal{T}^k} d_{i\ell}}{\sum_{i=1}^{\mathcal{T}^k} \bar{z}_i + 1} + \bar{\alpha} \sum_{i=1}^{\mathcal{T}^k} (1 - \bar{z}_i) - \bar{\alpha}. \tag{15}$$

Adding gaze point $\ell$ improves the objective if $\mathcal{Z}_{diff} = \mathcal{Z}_{new} - \mathcal{Z}_{feas} < 0$, where $\mathcal{Z}_{diff}$, the difference between $\mathcal{Z}_{new}$ and $\mathcal{Z}_{feas}$, is (after some algebraic rearrangement):

$$\mathcal{Z}_{diff} = \frac{\left(\sum_{i=1}^{\mathcal{T}^k} \bar{z}_i\right)\left(\sum_{i:\bar{z}_i=1}^{\mathcal{T}^k} d_{i\ell}\right) - \sum_{i=1}^{\mathcal{T}^k-1} \sum_{j=i+1}^{\mathcal{T}^k} d_{ij} \bar{z}_i \bar{z}_j}{\left(\sum_{i=1}^{\mathcal{T}^k} \bar{z}_i\right)\left(\sum_{i=1}^{\mathcal{T}^k} \bar{z}_i + 1\right)} - \bar{\alpha}. \tag{16}$$

Hence, the new gaze point $\ell$ improves the objective, and should be added to the fixation, whenever the penalty $\bar{\alpha} > \frac{\left(\sum_{i=1}^{\mathcal{T}^k} \bar{z}_i\right)\left(\sum_{i:\bar{z}_i=1}^{\mathcal{T}^k} d_{i\ell}\right) - \sum_{i=1}^{\mathcal{T}^k-1} \sum_{j=i+1}^{\mathcal{T}^k} d_{ij} \bar{z}_i \bar{z}_j}{\left(\sum_{i=1}^{\mathcal{T}^k} \bar{z}_i\right)\left(\sum_{i=1}^{\mathcal{T}^k} \bar{z}_i + 1\right)}$.

A similar derivation exists for formulation (13a)–(13f). Suppose there is a feasible solution $\bar{\mathcal{X}} = \{\bar{r}, \bar{x}, \bar{y}, \bar{z}\}$. Consider whether to add a new gaze point $\ell \in \mathcal{C}^k$ that is presently outside of the bounding box formed by $r$, $x$, and $y$, yet temporally adjacent to the first or

last point in the fixation. The radius will increase from $\bar{r}$ to $\bar{r} + \delta_r$, and one additional point will be included in the fixation, which will contribute $-\bar{\alpha}$ to the objective. Hence the gaze point $\ell$ should be included in the fixation if $\bar{\alpha} > \delta_r$.

Through this analysis we can see that varying $\alpha$ affects the density of resulting fixations in both formulations (12a)–(12f) and (13a)–(13f). Depending on such factors as the task and stimuli, different $\alpha$ levels may be necessary to induce desired levels of fixation density.

# References


Andersson, Richard, Linnea Larsson, Kenneth Holmqvist, Martin Stridh, Marcus Nyström. 2016. One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods* 1–22.

Bertsimas, Dimitris, Angela King. 2015. OR Forum–An algorithmic approach to linear regression. *Operations Research* **64** 2–16.

Bingham, Ella. 2010. Finding segmentations of sequences. *Inductive Databases and Constraint-Based Data Mining*. Springer, 177–197.

Blascheck, T., K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, T. Ertl. 2014. State-of-the-Art of Visualization for Eye Tracking Data. R. Borgo, R. Maciejewski, I. Viola, eds., *EuroVis - STARs*. The Eurographics Association.

Blignaut, Pieter. 2009. Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, Perception, & Psychophysics* **71** 881–895.

Bradley, Paul S, Olvi L Mangasarian, W Nick Street. 1997. Clustering via concave minimization. *Advances in Neural Information Processing Systems* 368–374.

Charikar, Moses, Chandra Chekuri, Tomas Feder, Rajeev Motwani. 2004. Incremental clustering and dynamic information retrieval. *SIAM Journal on Computing* **33** 1417–1440.

Chen, Chuan-Chong, Khee-Meng Koh. 1992. *Principles and Techniques in Combinatorics*. World Scientific.

Cockerham, Glenn C, Eric D Weichel, James C Orcutt, Joseph F Rizzo, Kraig S Bower. 2009. Eye and visual function in traumatic brain injury. *Journal of Rehabilitation Research and Development* **46** 811–818.

Djamasbi, Soussan. 2014. Eye tracking and web experience. *AIS Transactions on Human-Computer Interaction* **6** 37–54.

Djamasbi, Soussan, Marisa Siegel, Tom Tullis. 2010. Generation Y, web design, and eye tracking. *International Journal of Human Computer Studies* **66** 307–323.

Engelke, Ulrich, Hantao Liu, Junle Wang, Patrick Le Callet, Ingrid Heynderickx, Hans-Jurgen

Zepernick, Andreas Maeder. 2013. Comparative study of fixation density maps. *IEEE Transactions on Image Processing* **22** 1121–1133.

Estivill-Castro, Vladimir. 2002. Why so many clustering algorithms: A position paper. *ACM SIGKDD Explorations Newsletter* **4** 65–75.

Goldberg, Joseph H, Xerxes P Kotval. 1999. Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics* **24** 631–645.

Goldberg, Joseph H, Mark J Stimson, Marion Lewenstein, Neil Scott, Anna M Wichansky. 2002. Eye tracking in web search tasks: Design implications. *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*. ACM, 51–58.

Gurobi Optimization, Inc. 2016. Gurobi Optimizer 6.5.0 Reference Manual.

Hochbaum, Dorit S, Wolfgang Maass. 1985. Approximation schemes for covering and packing problems in image processing and VLSI. *Journal of the ACM (JACM)* **32** 130–136.

Holmqvist, Kenneth, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, Joost Van de Weijer. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.

Komogortsev, Oleg V, Denise V Gobert, Sampath Jayarathna, Do Hyong Koh, Sandeep M Gowda. 2010. Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on Biomedical Engineering* **57** 2635–2645.

Li, Beibin, Quan Wang, Erin Barney, Logan Hart, Carla Wall, Katarzyna Chawarska, Irati Saez de Urabain, Timothy J Smith, Frederick Shic. 2016. Modified DBSCAN algorithm on oculomotor fixation identification. *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. ACM, 337–338.

MathWorks, Inc. 2016. MATLAB Users Guide.

Nyström, Marcus, Kenneth Holmqvist. 2010. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods* **42** 188–204.

Olsen, Anneli. 2012. The Tobii I-VT fixation filter. *Copyright© Tobii Technology AB* .

Over, Eelco AB, Ignace TC Hooge, Casper J Erkelens. 2006. A quantitative measure for the uniformity of fixation density: The Voronoi method. *Behavior Research Methods* **38** 251–261.

Poole, Alex, Linden J. Ball. 2005. Eye tracking in human-computer interaction and usability research: Current status and future. C. Ghaoui, ed., *Encyclopedia of Human Computer Interaction*. Idea Group Reference, Hershey, Pennsylvania, 211–219.

Radvay, Xavier, Stéphanie Duhoux, Françoise Koenig-Supiot, François Vital-Durand. 2007. Balance training and visual rehabilitation of age-related macular degeneration patients. *Journal of Vestibular Research* **17** 183–193.

Rao, MR. 1971. Cluster analysis and mathematical programming. *Journal of the American Statistical Association* **66** 622–626.

Sabatos-DeVito, Maura, Sarah E Schipul, John C Bulluck, Aysenil Belger, Grace T Baranek. 2016. Eye tracking reveals impaired attentional disengagement associated with sensory response patterns in children with autism. *Journal of Autism and Developmental Disorders* **46** 1319–1333.

Sağlam, Burcu, F Sibel Salman, Serpil Sayın, Metin Türkay. 2006. A mixed-integer programming approach to the clustering problem with an application in customer segmentation. *European Journal of Operational Research* **173** 866–879.

Salojärvi, Jarkko, Kai Puolamäki, Jaana Simola, Lauri Kovanen, Ilpo Kojo, Samuel Kaski. 2005. Inferring relevance from eye movements: Feature extraction. *NIPS Workshop*. 45–67.

Salvucci, Dario D, Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*. ACM, 71–78.

Selim, Shokri Z, Mohamed A Ismail. 1984. K-means-type algorithms: A generalized convergence theorem and characterization of local optimality. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 81–87.

Seref, Onur, Ya-Ju Fan, Wanpracha Art Chaovalitwongse. 2013. Mathematical programming formulations and algorithms for discrete $k$-median clustering of time-series data. *INFORMS Journal on Computing* **26** 160–172.

Shah, Purvi, Minal Goyal, Daiyang Hu. 2016. Role of expiration dates in grocery shopping behavior: An eye tracking perspective. *Proceedings of the Twenty-Second Americas Conference on Information Systems (AMCIS)*. Association for Information Systems, 1–5.

Shojaeizadeh, Mina, Soussan Djamasbi, Andrew C. Trapp. 2016. Density of gaze points within a fixation and information processing behavior. *Proceedings of the 2016 Human-Computer Interaction International (HCII) Conference*. Springer, 1–8.

Smeets, Jeroen BJ, Ignace TC Hooge. 2003. Nature of variability in saccades. *Journal of Neurophysiology* **90** 12–20.

Terzi, Evimaria. 2006. Problems and algorithms for sequence segmentations. Ph.D. thesis, The University of Helsinki, Finland.

Terzi, Evimaria, Panayiotis Tsaparas. 2006. Efficient algorithms for sequence segmentation. *Proceedings of the 2006 SIAM International Conference on Data Mining*. SIAM, 316–327.

Thorup, Emilia, Pär Nyström, Gustaf Gredebäck, Sven Bölte, Terje Falck-Ytter. 2016. Altered gaze following during live interaction in infants at risk for autism: An eye tracking study. *Molecular Autism* **7** 1–10.

Tobii. 2018. Tobii technology. `http://www.tobii.com`. Accessed: 2018-08-02.

Trapp, Andrew, Oleg A Prokopyev, Stanislav Busygin. 2010. Finding checkerboard patterns via fractional 0–1 programming. *Journal of Combinatorial Optimization* **20** 1–26.

Trapp, Andrew C., Oleg A. Prokopyev. 2010. Solving the order-preserving submatrix problem via integer programming. *INFORMS Journal on Computing* **22** 387–400.

Wedel, Michel, Rik Pieters. 2008. A review of eye-tracking research in marketing. *Review of Marketing Research* **4** 123–147.

Wu, T. 1997. A note on a global approach for general 0–1 fractional programming. *European Journal of Operational Research* **101** 220–223.