Evaluating Special Event Transit Demand: A Robust Principal Component Analysis Approach

Pramesh Kumar and Alireza Khani

***Abstract*—The special events such as games, concerts, state fairs, etc. attract a large amount of population, which requires proper planning of transit services to meet the induced demand. Previous studies have proposed methods for estimating an average daily weekday demand, which have an inherent disadvantage in estimating the demand for a special event. We solve an idealized version of this problem i.e., we decompose a special event affected demand matrix into a regular and an outlier matrix. We start with detecting the special events in large scale transit data using the Mahalanobis distance, an outlier detection method for high dimensional data. Then, a special event demand is evaluated using state-of-the-art dimensionality reduction technique known as robust principal component analysis (RPCA), which is formulated as a convex optimization program. We show the application of the proposed method using Automatic Passenger Count (APC) data from Twin Cities, MN, USA. The methods are general and can be applied to any type of data related to the flow of passengers available with respect to time. Of practical interest, the methods are scalable to large-scale transit systems.**

***Index Terms*—special event, origin-destination (O-D) matrix, transit data, Automatic Passenger Count (APC), Robust Principal Component Analysis (RPCA), Mahalanobis distance, outlier detection**

I. Introduction

The National Highway Institute [1] in 1988 defined a "special event" as an occurrence that "abnormally increases the traffic demand, unlike an accident, construction, or maintenance activities, which typically restrict the roadway capacity". Special events can range from big events such as Olympics, Super Bowl, Concerts, etc. to small events such as a local community festival. These events have now become an important aspect of our lives and culture as the United States is becoming a leisure-oriented society [2]. Florida Department of Transportation Report 2006 categorizes these events as planned and unplanned events [3]. The planned events have fixed schedule, time, location, and duration, e.g., sporting events, concerts, festivals, parades, fireworks, conventions, and so on. On the other hand, unplanned events such as natural disasters, do not have a fixed schedule and duration. The current study focuses on the post-analysis of planned events. They attract a large amount of population, which requires planning from various aspects. Federal Highway Administration (FHA) of the US Department of Transportation recommends that a general feasibility study for a planned special event should at least include aspects such as travel forecast, market area analysis, parking demand analysis, travel demand analysis, roadway capacity analysis, and mitigation of impacts. For the induced demand created by special events, transportation agencies have to make arrangements to provide extra parking space, transit service, and better traffic management. Furthermore, special events, if not planned properly, may have disruptive impacts on our transportation infrastructure as current transportation infrastructure and services are not designed for extreme events. For example, high passenger inflows can cause extreme delays on both highway and transit networks.

During these events, many travelers, in order to avoid congestion on highways and high parking cost, decide to take transit to attend them. In such cases, the goal of a transit agency is to [3]:

1) reduce the delay for people attending and not attending the event.
2) improve mobility by providing convenient service.
3) expose transit system to non-riders
4) attract potential new riders

To achieve above goals and provide efficient service, a transit agency needs to know the induced demand for these events. One possible way is to conduct passenger surveys during these events to estimate this demand [4]. However, the data collected through surveys is limited, and cannot give a full estimate of this demand. On the other hand, transit Automated Data Collection Systems (ADCS) such as Automatic Passenger Count (APC) system or Automatic Fare Collection (AFC) system can provide a rich source of information about passengers' travel pattern on a continuous basis [5]. This ITS data can be used to evaluate an origin-destination (O-D) flow matrix using trip chaining of AFC tags ([6], [7]), or using novel optimization techniques employing APC data [8]. Although these methods promise to give a high-quality O-D matrix, the estimated matrix does not give us any information about whether the demand is regular or special event. An ideal way to pose this problem is as follows. Given a demand matrix $M \in \mathbb{R}^{m\times n}$, can we decompose it into a regular matrix $L \in \mathbb{R}^{m\times n}$ and a special event/outlier matrix $S \in \mathbb{R}^{m\times n}$?, i.e.,

$$M = L + S \tag{1}$$

The problem seems daunting at first sight, but under certain conditions [9], both L and S can be recovered. We study this decomposition problem and make the following contributions through this article:

Pramesh Kumar and Alireza Khani are with the Department of Civil, Environmental, and Geo-Engineering, Twin Cities, MN, 55455 USA (e-mail: kumar372@umn.edu and akhani@umn.edu, Web: http://http://umntransit.weebly.com/)

Color versions of one or more of the figures in this paper are available onlineat http://ieeexplore.ieee.org.

- Describe a procedure to detect special event(s) in a large scale time-series transit passenger flow data using an outlier detection method that leverages Mahalanobis distance [10].
- Use state-of-the-art dimensionality reduction technique known as Robust Principal Component Analysis (RPCA) via Principal Component Pursuit (PCP) to solve the decomposition problem (1) and estimate a special event demand matrix.
- Show application of these methods to evaluate the Minnesota State Fair demand on a transit route using APC data from Twin Cities, MN.

Although this research uses existing statistical techniques, it is a novel application of them in transportation science literature. The methods are general and can be applied to any type of transit or highway network data available with respect to time. The rest of the paper is organized as follows. §II describes previous work related to the demand estimation, followed by the methodology in §III. Then, a case study about the Minnesota State Fair as a special event is presented in §IV. Finally, conclusions and directions for future research are presented in §V.

II. Related work

The literature review is presented in two subsections. §II-A describes the literature related to special events and §II-B describes the literature on recent advances in performing Robust Principal Component Analysis (RPCA).

A. Special event literature

There is limited literature on recurring special event demand estimation. The Department of Transportation (DOTs) and other transportation agencies have published reports on general guidelines to follow when planning a special event [11], [2]. These reports discuss how various agencies should plan, coordinate, and manage different transportation systems for special events. The report is useful to any organization that is involved in the planning of a special event. These organizations include, but not limited to the Department of Transportation (DOTs), law enforcement agencies, media, event planners, consulting firms, and the military.

One of the steps in the planning of a special event is the demand estimation which can be classified into two categories [12]: long-term prediction and short-term prediction of demand. The short term prediction is necessary to avoid congestion and disruption on highway and transit network in real-time. Existing literature has considered short-term prediction methods based on neural networks ([13], [14]), time series analysis ([15], [16], [17], [12]), support vector machines [14], fuzzy logic [18], and Kalman filtering [19]. Although these methods are able to capture demand fluctuation in a shorter time span, they are not suitable to forecast demand for long-term planning. For long term planning, Pereira et al. used neural networks to predict the transit passenger arrival using social media and smart card data [20] and Ni et al. developed a hashtag-based event detection algorithm by combining optimization with hybrid loss function, and linear regression [21]. Kuppam used a traditional four-step model and calibrated choice models using a survey to predict the special event demand [4]. However, surveys are associated with inherent disadvantages such as limited size, general reporting errors, and so on which are not able to capture complete demand. One of the pioneer efforts in this regard is by Wong et al. [22]. They used a bi-level optimization with a multi-class traffic assignment at the lower level to evaluate a special event O-D matrix for Macau Grand Prix. With the advent of Intelligent Transit Data Collection Systems, namely, Automatic Fare Collection (AFC), Automatic Passenger Count (APC), and Automatic Vehicle Location (AVL) system, it is now possible to do a detailed analysis of transit passenger travel behavior [23]. We can perform a post-analysis of the demand and evaluate a high-quality OD matrix. Recent application of AFC and AVL data can estimate a stop-level transit O-D matrix using a method known as trip chaining [6], [7]. Trip chaining links various taps of a passenger, made using a smart card, throughout the day and predicts their boarding and alighting locations. The quality of this matrix depends on the number of passengers using a smart card to travel and trip chaining method used to estimate missing boarding/alighting location [7]. The penetration of smart cards is particularly important because visitors attending the special event may not possess a transit smart card. The Automatic Passenger Count (APC) data, on the other hand, provides a full picture by recording the number of boarding and alighting at each stop in the transit network. However, it requires solving an ill-posed system of linear equations to evaluate an O-D matrix. The methods which use boarding and alighting counts obtained from APC data to estimate an O-D matrix are either statistical methods ([24], [25]), or optimization methods ([8], [26]). To the best of the authors' knowledge, there is no study that uses automated transit data to do post-analysis and evaluate a special event OD matrix. This is because the methods proposed in the literature ([6], [7], [26], [24], [25], [4]), are able to evaluate a reliable O-D flow matrix, but unable to evaluate how much of that flow belongs to a special event. This naturally raises a question that can we decompose the given matrix into a regular and a special event matrix? The application of RPCA can help in such decomposition. It can be used to evaluate a low-dimensional matrix (regular matrix) along with the separation of special event matrix lying inside this high-dimensional time series data.

B. Robust Principal Component Analysis (RPCA)

Principal Component Analysis (PCA) is one of the most extensively used statistical technique for dimensionality reduction. The method is used to convert a set of correlated data into uncorrelated vectors using an orthogonal transformation. In l_2 sense, the problem tries to find a low rank matrix $L \in \mathbb{R}^{m\times n}$ having rank less than $r \in \mathbb{N}$ out of the given data matrix $M \in \mathbb{R}^{m\times n}$ using the following convex optimization program:

$$\begin{aligned} \underset{L}{\text{minimize}} \quad & \|M - L\| \\ \text{subject to} \quad & rank(L) \leq r \end{aligned} \tag{2}$$

where $\|M\|$ represents the spectral norm of the matrix M, which is equal to the largest singular value of M. The optimization program (2) can be efficiently solved using singular value decomposition (svd) of M [27]. However, PCA is extremely sensitive to the outliers present in the data matrix and performs poorly when grossly corrupted entries are present. Even one corrupted entry can result in a matrix $L^{'}$ which is significantly different from the true low-rank matrix L. To avoid this problem, several approaches have been proposed in the literature to robustify PCA, such as influence function techniques [28], multivariate trimming [29], random sampling [30], and so on. However, these techniques are either non-exact or exact with no polynomial-time algorithm to solve them. This makes it difficult to use these techniques for large-scale data. Recent advances on subspace estimation by rank minimization and sparse representation give a good framework for separating a low rank matrix from sparse corruptions. For Robust PCA via a low-rank and sparse decomposition, various formulations are proposed in the literature. This includes RPCA via Principal Component Pursuit [9], RPCA via Outlier Pursuit [31], RPCA via Iteratively Reweighted Least Squares [32], and Bayesian RPCA [33]. We do not describe every approach here but refer the interested reader to the review article by Bouwmans and Zahzah on their use of RPCA in video surveillance [34]. For this research, we use RPCA via Principal Component Pursuit [9] because of its suitability to the current problem.

III. Methodology

This section presents a methodology for detecting the special events and estimating the demand for it. Before describing the methods, we first show the systematic procedure to prepare the data on which the proposed methods can suitably be applied. The notations used in this article are summarized in Appendix A.

A. Preparation of time series data

For the application of methods presented in this article, we require passenger flow count along a transit route with respect to time. Such information can be obtained from transit automated data, (e.g., using APC or AFC data). Using the methods reviewed in §II, a time-dependent origin-destination (O-D) flow matrix can be obtained. We describe the steps to aggregate the time series data in a matrix form, required for this study. The passenger trips can be aggregated by their origin-destination pair or simply by their origin or destination only with respect to time in a matrix form. Such a matrix will help in extracting the useful statistical measures to detect unusual events and then identifying the duration of a special event. The methodology presented here is applied to the flow matrix of a transit route. However, it is straightforward to extend it for a network-level passenger flow matrix by including the origin-destination pairs or boarding/alighting stops for the whole network.

Let $N = \{1, 2, ..., |N|\}$ be the set of stops/stations along a transit route (For a network-level analysis, include all the stops/stations in the network). For a particular day, time can be discretized into h hour intervals, denoting it by the set $H = \{0, 1, 2, ..., \lfloor \frac{24}{h} \rfloor\}$. Let D be the set of days in our analysis period. We recommend using a large scale time-series data for this purpose. This will help in learning the average pattern of the trips in the given matrix. Also, the methods described in this paper can be scaled to a large amount of data which is one of its advantages. Let $T \to D \times H$ corresponds to a day-time mapping. This set contains time intervals for different days in our analysis period. Depending upon the availability of the data, there are two ways of aggregating transit trips with respect to time:

1) If we have a true time-dependent O-D matrix available, then we can aggregate the trips by their origin-destination pair, which in this case, is the combination of boarding and alighting stops of the transit route denoted by $K \to N \times N$. For a fixed $t \in T$, the total number of trips between different origin-destination pairs can be aggregated as a flow vector denoted as $m(t) \in \mathbb{R}^{|K|}$. By stacking these aggregated trip vectors $m(t)$ column-wise for each time period in T will create a time-dependent flow matrix $M = \{m_i(t) | i \in K, t \in T\}$.
2) As mentioned in the §II, a true time-dependent O-D matrix may not be easy to obtain. To avoid problems in obtaining a time-dependent O-D matrix, we can use the commonly available Automatic Passenger Count (APC) data, which provides the number of boarding and alighting on every stop along a transit route. In this case, two different aggregated flow matrices are prepared i.e., a boarding matrix and an alighting matrix because we do not know the actual flow between origin-destination pairs. For a fixed $t \in T$, the number of boarding and alighting at different stops $n \in N$ can be aggregated to create a boarding and alighting vector as $b(t) \in \mathbb{R}^{|N|}$ and $a(t) \in \mathbb{R}^{|N|}$ respectively. By arranging these vectors along different columns will form a boarding matrix $B = \{b_i(t) | i \in N, t \in T\}$, and an alighting matrix $A = \{a_j(t) | j \in N, t \in T\}$. Here, $b_i(t)$ and $a_i(t)$ represents the total number of boarding and alighting at stop i during time period t.

In any of the above cases, we will finally prepare a flow matrix represented as $M = \{m_j(t)\}$, where $j \in K$ can either represent an O-D pair $K \to N \times N$ or a stop location $K = N$. Table I shows the structure of such a matrix. Before moving further, we make the following assumption:

TABLE I: Flow matrix M (rows represents the O-D pairs or stops and columns represents day-time)

B / T	d_1-t_1	.	.	.	d_k-t_k	.	.	.	$d_{\|D\|}$-$t_{\|H\|}$
$od_1/b_1/a_1$									
$od_2/b_2/a_2$									
.									
.									
.									
$od_{\|K\|}/b_{\|N\|}/a_{\|N\|}$									

Assumption 1: The flow of passengers can be viewed as a distribution conditioned on time, which follows a periodic trend. The trend is observed according to the time of the day

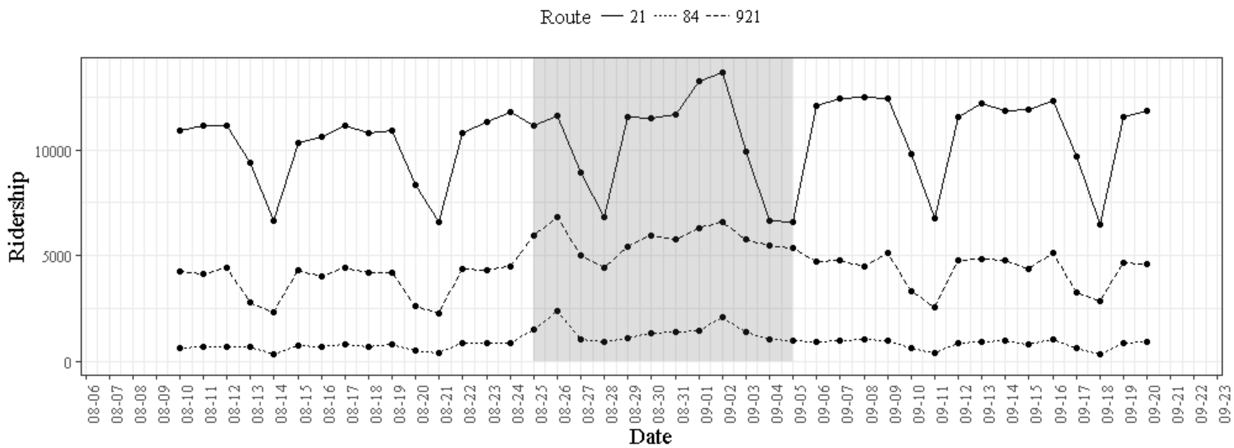


Fig. 1: Ridership of route 84, 21, and 921 (A line) (The Minnesota State Fair time frame is shaded)

and the day of the week. We assume that the high dimensional flow matrix M lies in approximately low dimensional subspace.

The intuition behind this assumption is that there exists a periodic pattern in the travel pattern (Figure 1) of passengers with some noise in it. This travel pattern can be observed during peak and non-peak hours. For example, on weekdays, some particular stops along a transit route show a high number of boarding during morning peak hours and alighting during evening peak hours. If there is a special event, the flow distribution would deviate from this periodic pattern.

B. Detection of a special event

The prepared flow matrix M can be used to detect special events. We assume that the trend (Assumption 1) in the number of boarding and alighting follows a weekly pattern that can be observed (Figure 1) by plotting the number of boarding or alighting with respect to time. To capture this periodic pattern, let us define a reference set $R_t = \{m(h) | h \equiv t \bmod 168, h \neq t\}$ for the given flow matrix with 168 hours in a week [35]. Also, to see how typical flow is at time t, it is recommended not to use $m(t)$ as a part of the definition. The reference set R_t can be used for calculating the expected boarding/alighting vector $\mu_m(t)$ and a covariance matrix $\Sigma_m(t)$ as below:

$$\mu_m(t) = \frac{1}{|R_t|} \sum_{m \in R_t} m \tag{3}$$

$$\Sigma_m(t) = \frac{|R_t|}{|R_t| - 1} \sum_{m \in R_t} \left(\frac{mm^T}{|R_t|} - \mu_m(t)\mu_m(t)^T \right) \tag{4}$$

(4) defines a non-standard but equivalent formula to compute the sample covariance matrix. The purpose of this is to improve the computational efficiency when computing the covariance of a high-dimensional matrix [35].

The mean $\mu_m(t)$ and $\Sigma_m(t)$ calculated above tells us the expected ridership and the daily variation in it along with the correlation among different dimensions at a particular time $t \in T$. If the flow at time t deviates more than a certain threshold from the mean vector $\mu_m(t)$, then that time duration can be flagged as an outlier or a special event. This deviation from the mean vector can be calculated as a standard z-score in one dimension. The generalization of this notion for higher dimensions (i.e., how many standard deviations a point is far from the mean of the distribution) is known as Mahalanobis distance [10]. For our case, the Mahalanobis distance for the flow vector $m(t)$ can be calculated as below:

$$\mathcal{M}(t) = \sqrt{\left(m(t) - \mu_m(t)\right)^T \Sigma_m(t)^{-1} \left(m(t) - \mu_m(t)\right)} \quad \forall t \in T \tag{5}$$

$\mathcal{M}(t)$ fluctuates periodically depending on the day of the week, and time of the day. It is a natural way of detecting outliers in a multivariate normally distributed data, but it has been shown to work well even when the data is not normally distributed [**?**]. Setting $\mathcal{M}(t)$ equals to a constant c defines a multi-dimensional ellipse with centroid at μ. The boundary of this ellipse is a probability density contour defined by the probability distribution of c^2, which follows χ_p^2 distribution with p degrees of freedom (in our case, $p = |T|$). This Mahalanobis measure can be used to detect a special event by flagging a time period $t \in T$ as an outlier event if $\mathcal{M}(t)$ is higher than a certain threshold value. The distribution of c^2 gives us a probabilistic bound on calculating this threshold value [**?**]. The probability of $M(t)^2 \leq \chi_p^2(\alpha)$ is $1 - \alpha$, where α is the significance level.

$$Prob\left[\left(m(t) - \mu_m(t)\right)^T \Sigma_m(t)^{-1} \left(m(t) - \mu_m(t)\right) \leq \chi_p^2(\alpha)\right] = 1 - \alpha \tag{6}$$

A similar bound based on the generalization of Chebyshev's inequality was developed by [**?**], however, it is a weaker bound than given in (6). For $p = |T|$ and $\alpha = 0.01$, one can use $\sqrt{\chi_p^2(\alpha)}$ as a threshold value to detect outliers. Geometrically, the value of $\sqrt{\chi_p^2(\alpha)}$ gives us a boundary, out of which the points can be considered as outliers with high probability. Using this technique, we can determine the duration of a special event in high dimensional data [36]. We show the application of this method in §IV.

C. Evaluating the special event flow matrix

In this section, we describe the main focus of this research, which is evaluating the demand for a special event. First, we formulate the problem as an optimization problem and then present the solution algorithm for it.

1) Mathematical formulation: As we assume that the flow matrix lies in some low dimensional subspace (Assumption 1), we aim to recover that low dimensional matrix, which can be obtained using Principal Component Analysis (PCA), a standard problem in the literature ([37], [38]). Since, PCA is not able to perform efficiently in case of gross corruptions, in idealized settings, one would like to decompose the matrix M into a low rank component (regular demand) and an outlier component of grossly corrupted values (special event demand) in order to apply PCA. This can be written as follows: Given a flow matrix $M \in \mathbb{R}^{m\times n}$, we would like to recover low rank matrix $L \in \mathbb{R}^{m\times n}$ and sparse matrix $S \in \mathbb{R}^{m\times n}$ such that

$$M = L + S \tag{7}$$

We make an advantage of this decomposition and evaluate a special event matrix S, which is a sparse flow matrix having outlier entries uniformly distributed throughout the matrix. S is sparse as the special events do not happen regularly as often in the high dimensional data. Note that we do not have any prior information about the column space of L or support of S, where it is sparse. The decomposition (7) can be achieved by solving the optimization program (8), which tries to minimize the rank of the matrix L (to recover a low-rank matrix for PCA) along with the number of non-zero entries in S (to recover a sparse matrix), known as Robust Principal Component Analysis (RPCA).

$$\begin{aligned} \underset{L,S}{\text{minimize}} \quad & rank(L) + \lambda\|S\|_0 \\ \text{subject to} \quad & M = L + S \end{aligned} \tag{8}$$

where, $\|S\|_0 = \lim_{p\to 0}\sum_{i,j}|S_{ij}|^p$ represents l_0 norm of matrix S which is the number of non-zero entries in S. The optimization program (8) is a non-convex and an NP-hard problem which is not easy to solve for a high dimensional matrix to achieve a global optimum. Recently, a tractable convex optimization program to solve (8) is proposed by [39] and [9] known as Principal Component Pursuit (PCP). PCP is inspired by the recent advancement in the field of compressed sensing [40], [41] which tries to obtain the sparsest solution planted in an underdetermined system of equations. The program can be written as follows:

$$\begin{aligned} \underset{L,S}{\text{minimize}} \quad & \|L\|_* + \lambda\|S\|_1 \\ \text{subject to} \quad & M = L + S \end{aligned} \tag{9}$$

where, $\|L\|_* = \sum_i \sigma_i(L)$ represents the nuclear norm of the matrix L which is sum of singular values of L and $\|S\|_1 = \sum_{ij}|S_{ij}|$ represents the l_1 norm of S which is sum of absolute values of elements of S. In the program (9), l_1 norm is used as the tightest convex relaxation of l_0 norm by minimizing the sum of non-zero entries instead of the number of non-zero entries of a matrix. This convex relaxation has been greatly used in recovering a sparse matrix from an underdetermined system of equations [40]. For example, [42] used this framework to evaluate an O-D matrix on a highway network, [8] used it to evaluate a transit route OD matrix using APC data, and [43] used it to optimally locate sensors on a highway network for O-D estimation. Similarly, the nuclear norm is used as the tightest convex relaxation of rank function. The intuition behind this relaxation is that a matrix L with rank r has exactly r non-zero singular values, which means that the rank is simply the number of non-vanishing singular values. So, minimizing the sum of singular values of a matrix which is its nuclear norm can be understood as the minimization of rank of a matrix [44]. The use of l_1 norm is justified when S satisfies a Restricted Isometry Property (RIP) [40]. This condition is satisfied by most of the random matrices and its successful application in estimating O-D matrix can be found in ([43], [8]). Similar RIP condition for nuclear norm can be found in [45]. The parameter λ is a critical parameter, higher value of which detects fewer outliers in S. More details about the choice of λ is given in §IV-C. In this way, we now have a tractable convex optimization program (9) which is far easier to solve than (8). It is shown in [9] that under a few assumptions, we can exactly and efficiently recover L and S even though we do not have any information about the low rank structure of L and location of outliers in matrix S. These assumptions are discussed below:

Assumption 2: The matrix L should satisfy incoherence conditions (10) which state that the singular vectors of L should be reasonably spread out and the entries in S are located uniformly at random.

Assumption 2 tries to avoid the extreme cases such as matrix $M = e_1 e_1^T$ (e_1 is the standard basis), which has 1 at the top left corner and zeros elsewhere. In this case, it is not possible to find L and S unless we know all the entries. Such situations can be avoided by imposing incoherence conditions proposed by [44]. Let us denote the singular value decomposition of L as $L = U\Sigma V^T = \sum_{i=1}^{r}\sigma_i u_i v_i^T$, where $r = rank(L)$, σ_i is the i^{th} positive singular values, and U and V are the left and right singular matrices with first r columns. Then according to the incoherence conditions specifies that,

$$\begin{aligned} \max_i \|U^T e_i\|^2 &\le \frac{\mu r}{m} \\ \max_i \|V^T e_i\|^2 &\le \frac{\mu r}{n} \\ \|UV^T\|_\infty &\le \sqrt{\frac{\mu r}{mn}} \end{aligned} \tag{10}$$

$\|X\|_\infty$ is the l_∞ norm which is defined as $\max_{i,j}|X_{ij}|$. The conditions in (10) state that the orthogonal projection onto U or V should be less than the rank multiplied by the parameter μ and divided by the dimension of the matrix. If (10) is satisfied, then the separation (8) makes sense because the singular vectors of L would be spread out or not sparse. In our numerical experiment (§IV-C), the value of μ was found to be equal to 130.12.

The optimization program (9) assumes that L is exactly low rank and S is exactly sparse. However, the flow matrix obtained from data such as APC is often corrupted by daily noise and it only has an approximate low rank structure. The noise can be attributed to the failure of APC systems correctly recording the boarding and alighting data or change in the regular travel pattern of passengers during the special events. For example, some transit riders might avoid their regular trips by working from home. In that case, our flow matrix M can be considered as the sum of three components:

$$M = L + S + N \tag{11}$$

where, $N \in \mathbb{R}^{m \times n}$ represents the noise matrix. Assuming that entries in N follows i.i.d. Gaussian distribution and $\|N\|_F \leq \delta$ for some value of $\delta > 0$, [46] proposed the program (12) to exactly recover L and S. The optimization program (12) is known as Stable Principal Component Pursuit (SPCP) in the literature.

$$\begin{aligned} \underset{L,S}{\text{minimize}} \quad & \|L\|_* + \lambda \|S\|_1 \\ \text{subject to} \quad & \|M - L - S\|_F \leq \delta \end{aligned} \tag{12}$$

The parameter δ shows the accuracy of matrix M and can be adjusted to represent the actual noise in it. Note that there is no restriction on the sign of the entries in these matrices. Therefore, if regular riders decide to work from home, then N would take negative entries representing the reduction in the number of regular trips. Another possibility is that sometimes there are missing entries in the data. This can happen when the automated data collection system fails to record the values. Even in such cases, we can recover L and S using (12). For those cases, let us assume a set $\Omega = \{(i,j)$ where M_{ij} is observed$\}$ and $\mathcal{P}_\Omega(X)$ be the projection of X onto the set of observed entries Ω i.e.,

$$\mathcal{P}_\Omega(X) = \begin{cases} X_{ij}, & (i,j) \in \Omega \\ 0, & (i,j) \notin \Omega \end{cases}$$

and then the optimization program (12) can be modified as below:

$$\begin{aligned} \underset{L,S}{\text{minimize}} \quad & \|L\|_* + \lambda \|S\|_1 \\ \text{subject to} \quad & \|\mathcal{P}_\Omega(M - L - S)\|_F \leq \delta \end{aligned} \tag{13}$$

The program (13) will decompose the matrix M along with the prediction of missing entries. This formulation is an extension of the matrix completion problem proposed by [44], which is a popular technique to do collaborative filtering.

2) Solution Algorithm: We can treat (9), (12), and (13) as a general convex optimization problem and solve it using an interior point method after formulating it as a semidefinite program. The semidefinite reformulation can be found in [47]. However, the interior point methods perform poorly with high dimensional matrices as they rely on the Hessian of the objective function, resulting in prohibitive computational time even for moderately large size problems (e.g., one with the dimension of the order of 100). In such cases, first order methods are often preferred for large-scale optimization. [48] and [49] proposed various first order optimization algorithms for this problem. We use Accelerated Proximal Gradient (APG) method because of its suitability to the problem structure and faster convergence rate. Instead of solving (12), we can equivalently solve the following dual problem:

$$\underset{L,S}{\text{minimize}} \quad \mu(\|L\|_* + \lambda\|S\|_1) + \frac{1}{2}\|M - L - S\|_F^2 \tag{14}$$

(14) is equivalent to (12) for a given value of $\mu(\delta)$ [46]. The proximal gradient method naturally applies to such composite functions as this is the sum of a smooth (l_2 norm) and non-smooth functions (l_1 and nuclear norm). Let us denote X as the ordered pair (L, S) and define $f(X) = \frac{1}{2}\|M - L - S\|_F^2$ and $g(X) = \|L\|_* + \lambda\|S\|_1$. Then, we can write (14) as:

$$\underset{X}{\text{minimize}} \quad F(X) = f(X) + \mu g(X) \tag{15}$$

where, $f(X)$ is smooth and convex with gradient being Lipshitz continuous having Lipschitz constant $L_f = 2$ and $g(X)$ is convex but non-smooth. In proximal gradient method, we approximate the smooth function $f(X)$ by it's second order Taylor series expansion $Q(X_0, Y)$ given the value of X_0 (see 17. Clearly, $Q(X_0, Y)$ which is an upper bound to $F(X)$.

$$Q(X_0, Y) = \mu g(X_0) + \langle \nabla f(Y), X_0 - Y \rangle + \|X_0 - Y\|^2 \tag{16}$$

$$= \mu g(X_0) + \|X_0 - (Y - \frac{1}{2}\nabla f(Y))\|_2^2 \tag{17}$$

Definition 1: (Proximal Mapping). For a closed function $g(X)$ and a parameter $t \in \mathbb{R}$, the proximal mapping $\text{prox}_h(X)$ is defined as follows:

$$\text{prox}_h(X) = \underset{Z}{\text{argmin}} \ \frac{1}{2t}\|X - Z\|_2^2 + h(Z) \tag{18}$$

In proximal gradient descent method, we choose initial iterate $X^{(0)}$, and then repeat

$$X_k = \text{prox}_{t_k}(X_{k-1} - t_k \nabla f(X_{k-1})), \ k = 1, 2, ... \tag{19}$$

We can see that the next iterate using (19) is obtained by minimizing (17) with $t_k = \frac{1}{L_f}$. This method works well in practice if it is easy to evaluate the proximal mapping. In our case, it is found that the proximal mapping for $h(X)$ which is the sum of l_1 norm and nuclear norm can be evaluated in a closed form. This closed form expression is known as soft-thresholding operator which is being frequently used in l_1 norm minimization arising in compressed sensing problems ([50], [51]). Similar iterative thresholding operator can also be used for nuclear norm minimization [52].

By defining gradient step update $G = Y - \frac{1}{L_f}\nabla f(Y)$ having order pair $G^L = Y^L - \frac{1}{2}(Y^L + Y^S - M)$ and $G^S = Y^S - \frac{1}{2}(Y^L + Y^S - M)$, we can repeatedly get the next iterate X_{k+1} using (19).

Definition 2: (Soft-thresholding operator). The minimizer in each iteration which is the soft-thresholding operator $\mathcal{S}_\epsilon[x]$ can be defined for $x \in \mathbb{R}, \epsilon > 0$ as below:

$$\mathcal{S}_\epsilon[x] = \begin{cases} x - \epsilon, & \text{if } x > \epsilon \\ x + \epsilon, & \text{if } x < -\epsilon \\ 0, & \text{otherwise} \end{cases}$$

This makes it easy to compute the minimizer by just soft-thresholding the singular values of L and soft-thresholding the individual values in S. We have $G_k = (G_k^L, G_k^S)$ and let the singular value decomposition (svd) of $G_k^L = U\Sigma V^T$. Then,

$$L_{k+1} = U\mathcal{S}_{\frac{\mu_k}{2}}(\Sigma)V^T \quad S_{k+1} = \mathcal{S}_{\frac{\lambda\mu_k}{2}}(G_k^S) \tag{20}$$

The natural choice of $Y_k = X_k$, for which the convergence rate is no worse than $\mathcal{O}(\frac{1}{k})$ [50]. We can accelerate the convergence by setting $Y_k = X_k + \frac{t_{k-1}-1}{t_k}(L_k - L_{k-1})$, having a step size satisfying $t_{k+1}^2 - t_{k+1} \leq t_k^2$, which results in improvement of convergence rate up to $\mathcal{O}(\frac{1}{k^2})$. Thus, for $\epsilon > 0$, when $k > k_0 + \frac{2\|X_{k_0} - X^{opt}\|_F}{\sqrt{\epsilon}}$, we can guarantee that $F(X_k) < F(X^{opt}) + \epsilon$, where k_0 is the first iteration and X^{opt} is the optimal value of X. The overall method given in [48] in summarized in Algorithm 1.

Algorithm 1 PCP using Accelerated Proximal Gradient method

Input Flow matrix $M \in \mathbb{R}^{|K|\times|T|}$, λ
Initialization $L_{-1}, L_0 \leftarrow 0^{|K|\times|T|}$; $S_{-1}, S_0 \leftarrow 0^{|K|\times|T|}$; $\tau \leftarrow 10^{-5}$, $\eta \leftarrow 0.9$ and $\mu \leftarrow 0.99\|M\|_F$, $t_{-1} = t_0 \leftarrow 1$; $\bar{\mu} \leftarrow \tau\mu$
while not converged **do**:
$Y_k^L \leftarrow L_k + \frac{t_{k-1}-1}{t_k}(L_k - L_{k-1})$, $Y_k^S \leftarrow S_k + \frac{t_{k-1}-1}{t_k}(S_k - S_{k-1})$
$G_k^L \leftarrow Y_k^L - \frac{1}{2}(Y_k^L + Y_k^S - M)$, $G_k^S \leftarrow Y_k^S - \frac{1}{2}(Y_k^L + Y_k^S - M)$
$(U, \Sigma, V) \leftarrow \text{svd}(G_k^L)$,
$L_{k+1} \leftarrow U\mathcal{S}_{\frac{\mu_k}{2}}(\Sigma)V^T$ and $S_{k+1} \leftarrow \mathcal{S}_{\frac{\lambda\mu_k}{2}}(G_k^S)$
$t_{k+1} \leftarrow \frac{1+\sqrt{4t_k^2+1}}{2}$, $\mu_{k+1} \leftarrow \max(\eta\mu_k, \bar{\mu})$, $k \leftarrow k+1$
end while

IV. Application for Twin Cities transit data

In this section, we show the application of the proposed methodology using APC data from Twin Cities, MN. This data was obtained from Metro Transit, which is the primary transit agency in Minneapolis/St. Paul region offering a connected network of buses, light rail and commuter rail services. The Automatic Passenger Count (APC) data used for this research contains transit trip information, such as date and time of the operation, routeID, stopID, departure and arrival time, number of boarding and alighting on each stop, and the geographical coordinates of the stops. To get insights into the results obtained after applying our methodology, we select a known event beforehand. However, the methods would work in the presence of both known/unknown events.

A. Minnesota State Fair

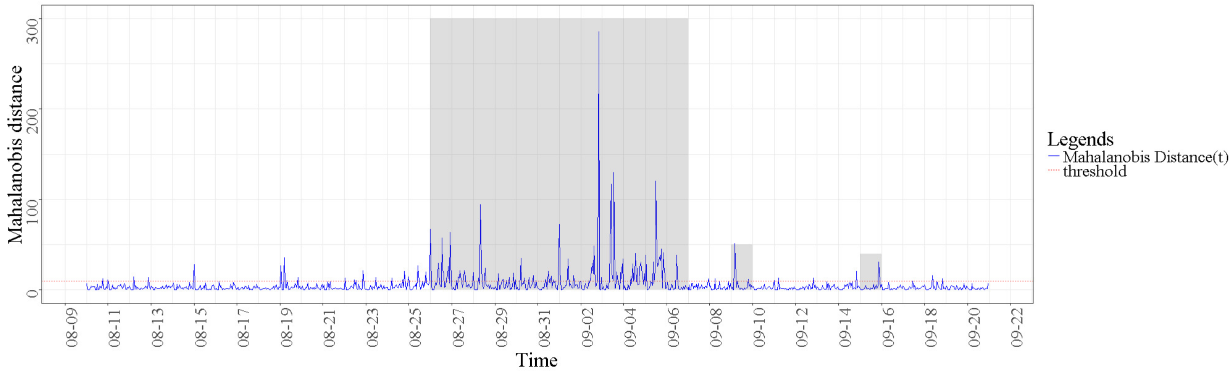
We present a case study of the Minnesota state fair as a special event. Minnesota state fair is the largest state fair in the United States by average daily attendance [53]. In 2016, it was held from 08/25/2016 to 09/05/2016 having 1,943,719 attendees from all over the country [53]. The fair is organized in the State Fair Grounds located in Falcon Heights, halfway between the capital of Minnesota, City of St. Paul and its largest city, Minneapolis. To avoid driving on congested highways during the state fair, many people decide to take transit to attend the state fair. Several new state fair buses are arranged to serve the induced demand. There are some regular buses such as route 84, route 21, and route 921 (A Line BRT), which also serve the State Fair Grounds. Figure 1 shows the ridership of these three routes from 08/10/2016 to 09/20/2016. The duration of the state fair is shown by the shaded region in the figure. Although we can observe a rise in the ridership of all three routes during that period, we do not know how much of that ridership belongs to the state fair. Due to heavy demand, the buses run overcrowded during that period due to which passengers have to stand inside the bus. The quantification of special event demand will help in designing adequate frequency of transit service during that period.

For this research, we analyze the effect of Minnesota state fair on the demand of route 921 (A line). This line is a bus rapid transit (BRT) service in the Twin Cities region which runs on the Snelling Ave corridor. It has 20 stations, with Snelling & Como Av Station being the closest station to the State Fair Grounds. We use APC data from 08/10/2016 to 09/20/2016 for this analysis. The matrix M is prepared using the aggregation procedure described in §III-A. The dimension of the final matrix was $\mathbb{R}^{20\times 336}$ having 20 transit stops and 336 time intervals for different days, which is 8 time intervals per day.

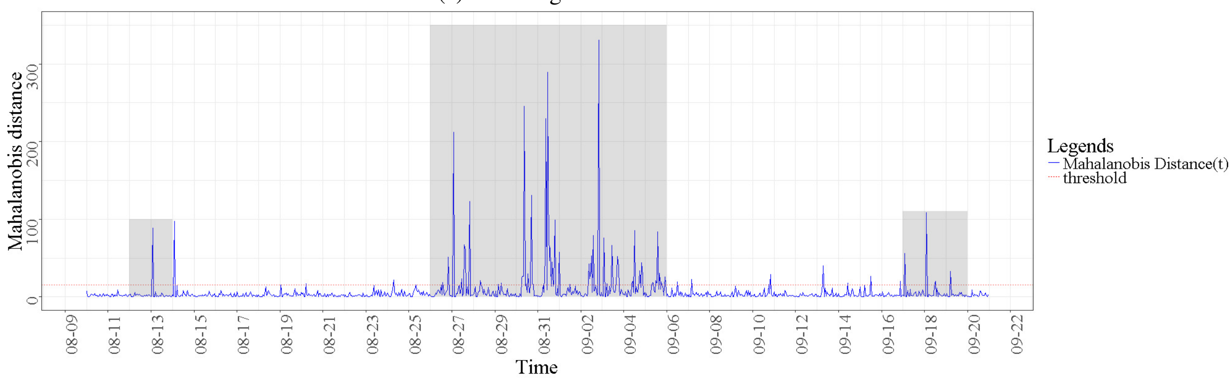
B. Analysis of the special event using Mahalanobis Distance

We prepared four different matrices for this analysis, each for the number of boarding and alighting in the northbound and southbound direction respectively. After that, corresponding mean and covariance matrices are calculated using (3), and (4) respectively. Finally, the Mahalanobis distance $\mathcal{M}(t)\ \forall t \in T$ was calculated using equation (5). To see whether Mahalanobis distance can detect the special event, the results are presented in Figure 2. We plotted $\mathcal{M}(t)$ against t to observe the outliers in the time range. Figure 2(a) and (b) show $\mathcal{M}(t)$ for boarding and alighting matrix in southbound direction.

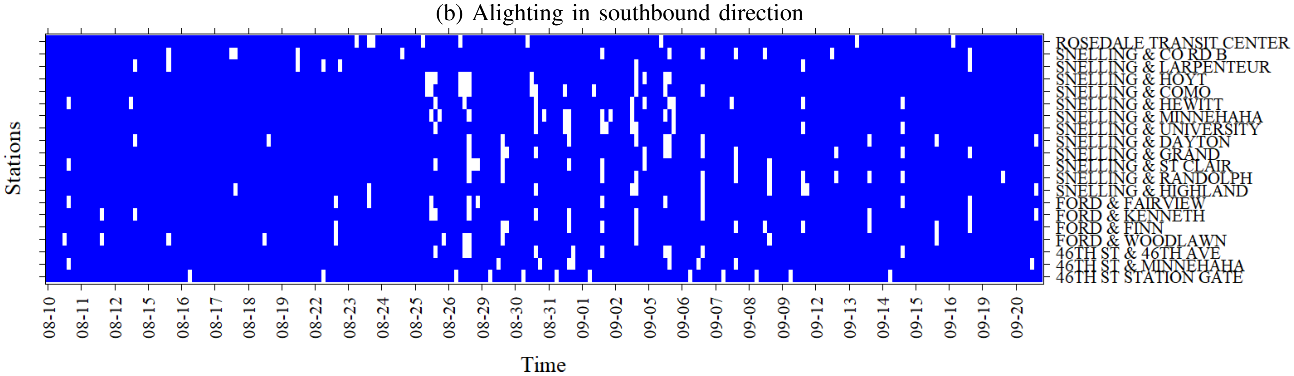
The Mahalanobis distance is intuitively the number of the standard deviation a given vector is away from the mean



(a) Boarding in southbound direction



(b) Alighting in southbound direction



(c) Detecting outliers in southbound alighting using 95th percentile outlier detection method

Fig. 2: Outlier event detection. Figure (a) and (b) shows $\mathcal{M}(t)$ versus time and (c) shows heatmap of outliers using 95th percentile outlier detection method (the white color indicates an outlier and blue color indicates a non-outlier) (For interpretation of colors, please refer to the web version of this article.)

vector. If this value is high, then we expect to see an unusual peak during that time period. The time range can be flagged as an outlier if $\mathcal{M}(t)$ rises above a given threshold. The threshold value can be decided by observing a regular pattern in the peaks of the plot or using the bound given in (6). In our case, the threshold value is equal to $\sqrt{\chi_p^2(0.01)} = 16.52$, which is marked by a red line in Figure 2(a)-(b). By making use of this threshold value, the outliers time ranges are shown by the shaded portions in these figures.

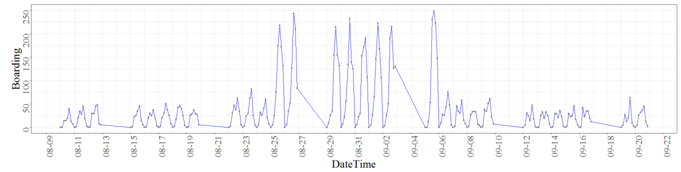
In 2016, the Minnesota state fair was held from 08/25/2016 to 09/05/2016. In Figure 2(a)-(b), we can observe that the $\mathcal{M}(t)$ started to rise on 08/26/2016, showing unusual peaks during the state fair period and then got back to normal trend on 09/06/2016. Although the state fair ended on 09/05/2016, the peaks can still be observed for the next day which is the labor day holiday. The highest peak in both figures was observed on 09/03 which was a weekend during the state fair. We can also see a few other peaks outside the state fair. For example, Figure 2(a) shows a high number of boarding on 09/09 and 09/16 in the southbound direction because of some other event such as a game, concert, etc. This would help a transit agency to look into unknown events.

To show the benefit of using Mahalanobis distance to detect outlier events, we compare its results with the *percentile outlier detection method*. This is a generic method which flags a time interval as an outlier event if the number of boarding/alighting at a stop during that time interval exceeds 95th percentile value. The results of the 95th percentile outlier detection method are shown in Figure 2(c). In this heatmap, the outliers and non-outliers are indicated in white and blue color respectively. Unlike Mahalanobis distance which computes a single measure for each time interval, the 95th percentile method shows outliers in two dimensions. We can observe that the results computed using this generic method are quite sensitive to the noise in data, and it detects outliers that are scattered all over the time horizon without giving a clear indication of the state fair duration. The problem with this method is that it fails to capture the correlation among different dimensions to create a trend in the boarding/alighting pattern.

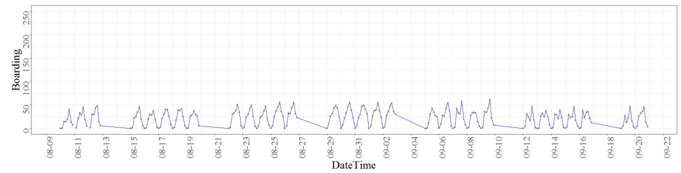
C. Evaluating outlier flow matrix for Minnesota state fair

In this section, we discuss the implementation and results of our outlier flow matrix estimation using RPCA discussed in §III-C. The results are computed for both boarding and alighting in each direction but we only present the result for boarding in southbound direction to conserve space. To obtain the regular matrix L and the special event matrix S, Algorithm 1 is implemented in Python 2, which is shared as a public source code [54]. The algorithm requires two inputs, matrix M and λ. [9] suggested that the value of $\lambda = \frac{1}{\sqrt{\max(m,n)}}$, (where $M \in \mathbb{R}^{m \times n}$) to exactly recover L and S theoretically, but it may require further tuning of this parameter to get the best results. In our case, $\lambda = \frac{1}{\sqrt{336}} = 0.05$ did not work well. There are other values of λ suggested in the literature. For example, [55] suggested $\lambda = \frac{1}{\sqrt{\log n}}$. However, none of the value of λ suggested in the literature worked best for the current study. So, we performed repeated adjustment of λ in order to get the best results by observing the rank of the matrix L after every adjustment which can be done by plotting the flow from low rank matrix L as shown in Figure 3(b). For an appropriate value of λ, we should see a regular pattern in the flow. We used $\lambda = 0.09$ to solve the program for both matrices.

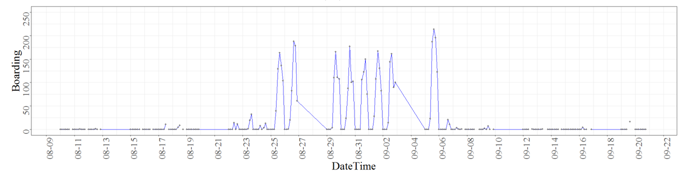
To present the flow in the original and the recovered flow matrices, we prepared heatmaps for boarding in the southbound direction which is shown in Figure 4. The colors show the intensity of flow from various A line stations (on the vertical axis) during different time intervals (on the horizontal axis). The state fair period is enclosed in a rectangle on the horizontal axis. In Figure 4(a), we can observe a high number of boarding on the commencing station which is Rosedale Transit Center and other stations such as Snelling & Como Av and Snelling & University Av station. Snelling & University Av station shows a high number of boarding because it is a transfer station to the Metro Green line, which connects Downtown Minneapolis and Downtown St. Paul via the University of Minnesota campus. We also see a high number of boarding on Snelling & Como Av during the state fair because this is the closest station to State Fair Grounds. RPCA seems to perform an excellent job in recovering the regular matrix L along with outlier matrix S, heatmaps of which are shown in Figure 4(b) and 4(c) respectively. The stations before Snelling & Grand Av show regular boarding as shown by the color intensity in Figure 4(b). The extra demand during the state fair (Figure 4(c)) was generated from Rosedale Transit Center, Snelling & County Rd and Snelling & Hoyt station to go to the state fair. We can also see high number of boarding on Snelling & Como station to alight at all the remaining stations in the southbound direction. To see how RPCA recovered L and S matrices, the number of boarding in southbound direction for Snelling & Como Av station is plotted against the time in Figure 3 for M, L and S. We can see that the extra number of boarding created during state fair at Snelling & Como Station (Figure 3(a)) is successfully recovered from matrix M as S component (Figure 3(c)), leaving behind the regular component (Figure 3(b)).



(a) M



(b) L



(c) S

Fig. 3: M, L, and S for Snelling Av and Como Av Station

A similar analysis was done for the alighting matrix in the southbound direction. We found that Snelling & University Av, Snelling & Grand, Snelling & Randolph, and the concluding station, 46th Street Station are the most popular alighting stations for regular passengers. During the state fair, passengers who boarded at Snelling & Como Station seemed to alight at Snelling & University Av, Snelling & Dayton, Snelling & Grand, Snelling & St. Claire, Snelling & Randolph, and 46th Street Station.

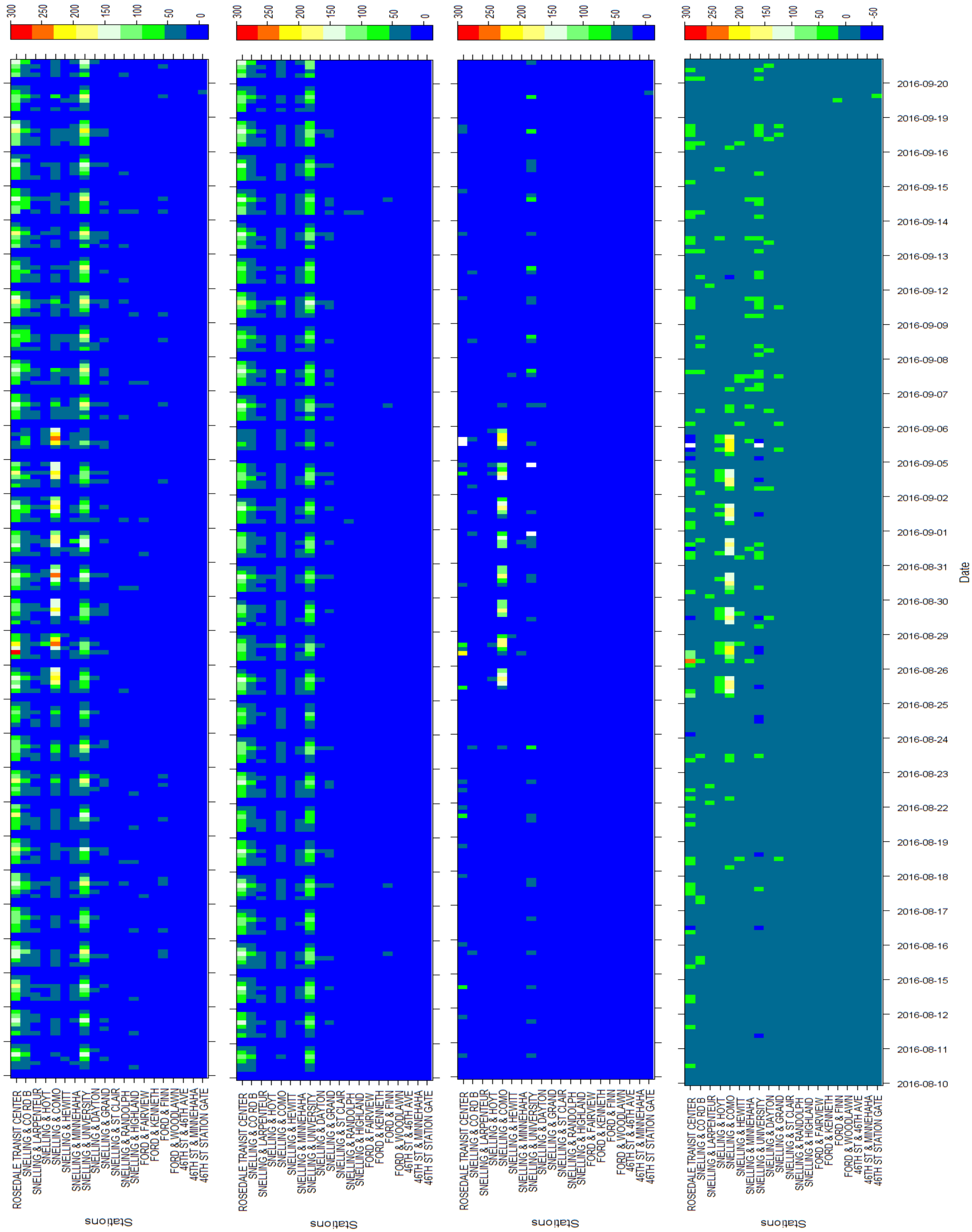
To show the benefit of using RPCA in evaluating the special event demand matrix, we compare its results with the *averaging method*. We assume that the regular demand matrix L_{avg} is the historical average of the weekly demand pattern. To be fair in comparison, we excluded the Minnesota State Fair time duration while computing the average demand. Then, the outlier demand S_{avg} is evaluated by subtracting L_{avg} from M. The results are shown in Figure 4(d). The outlier demand evaluated using the averaging method shows extra demand both during the state fair as well as outside the state fair time duration. We also observe negative values for some time intervals, that is, the reduction in the number of trips during the state fair, which seems unlikely as we expect more demand during that duration. Overall, the averaging method suffers from limitations such as assumption on the structure of the low-rank matrix, which RPCA avoids in its calculation.

To analyze which stretch of the A line is most affected by the state fair, we created a passenger load map for southbound direction. The load is calculated by subtracting the cumulative sum of alighting from the cumulative sum of boarding. In southbound direction (Figure 6), we can observe a heavy passenger load between Snelling & Como Av station and Snelling & St. Clair station. The load is highest between Snelling & Como Av and Snelling & University Av because Snelling & University Av station is a transfer point from Metro Green line to A line. These observations can help Metro Transit to increase the frequency of the bus only along a particular stretch instead of the full route. For example, considering the capacity of the bus is 40, for a total demand of 832 passengers in 3 hour period between Snelling & Como Av and Snelling & University Av, the required headway is $\frac{60*3*40}{832} \approx 8$ min in comparison to current headway of 10 minutes. Increasing the frequency only along a small stretch will save the operational cost to handle the extra demand. This is shown in Figure 6, where we can observe that increasing the frequency only along a stretch (i.e. Kenneth to Como) would help us avoid the reduction in the unused capacity of the bus. The shaded area in the figure shows the unused capacity of A Line route. This is higher if we increase the frequency along the complete route in comparison to a particular stretch where more buses are needed. For example, in our case, the unused capacity in the first figure is 24 (min) * 8 (buses/hrs) * 3 (hrs) * (40 seats) - 103 (seat-hrs) = 281 seat-hrs in comparison to [12 (min) * 6 (buses/hrs) * 3 (hrs) * (40 seats)] + [12 (min) * 8 (buses/hrs) * 3 (hrs) * (40 seats)] - 103 (seat-hrs) = 233 seat-hrs in the second figure. Such analysis would help transit planners to identify the stretch where more buses are needed and evaluating the appropriate frequency for that.

V. Conclusions and recommendations for future work

The special events such as games, sports, state fairs, etc. can affect the regular transit ridership for which the service is designed. This induced demand must be managed properly, otherwise, it can have a disruptive impact on the transit service. Previous approaches are not applicable in evaluating demand during such events. This is for the first time that we approach the problem directly by decomposing the given demand matrix into a regular and a special event matrix. We propose to use Mahalanobis distance to see how atypical flow is with respect to time to detect the duration of any special event. The method is easy to implement and gives us an idea of how severe an event is. After this, RPCA via PCP is used to evaluate the special event demand. Due to the unavailability of a full origin-destination matrix, we used the boarding and alighting counts obtained from APC data to evaluate and analyze the demand during Minnesota State Fair. We observed that the Mahalanobis distance did an excellent job in identifying the outlier time range of the Minnesota state fair. We also observed that the outlier demand generated during the state fair can be successfully recovered by applying RPCA. The extra demand (outlier flow) generated during the state fair is evaluated in terms of the number of boarding and alighting at each stop. Furthermore, we found that the evaluated regular matrix could capture the systematic pattern of boarding/alighting of the passengers, whereas the outlier matrix could capture the extra demand generated during the special event. The extra demand can be used to evaluate an adequate frequency of bus route on a particular stretch of the transit route for a future event.

One of the limitations of this method is that it cannot differentiate the demand for several special events in the region. There is a need for investing this issue further and propose methods to evaluate the demand for multiple special events. Due to the unavailability of complete AFC or survey data, we could not validate the results. Future studies are encouraged to validate the results of the proposed methods. This research can be extended in multiple directions. The idea of detecting outlier event using Mahalanobis distance can be used to measure the resilience of other transportation systems. For example, it can be applied to time-series traffic speed data to measure the resiliency of a highway network. Similarly, RPCA can be applied to evaluate automobile demand during special events. Furthermore, the presented analysis can be extended for a citywide transit network using a network-wide flow matrix. This will help in evaluating the extra demand for other routes during a special event. Not only the special events, the impact of land-use changes from time to time (e.g., the opening of a new supermarket, transit route, and so on) or declining ridership due to weather, which actively affects the origin-destination flow, can also be evaluated.



(a) matrix M (b) matrix L (c) matrix S (d) S_{avg}

Fig. 4: Boarding in southbound direction. Figure (a), (b), and (c) shows actual, regular, and outlier demand respectively calculated using RPCA method, and (d) shows outlier demand evaluated using averaging method (For interpretation of colors, please refer to the web version of this article)

.

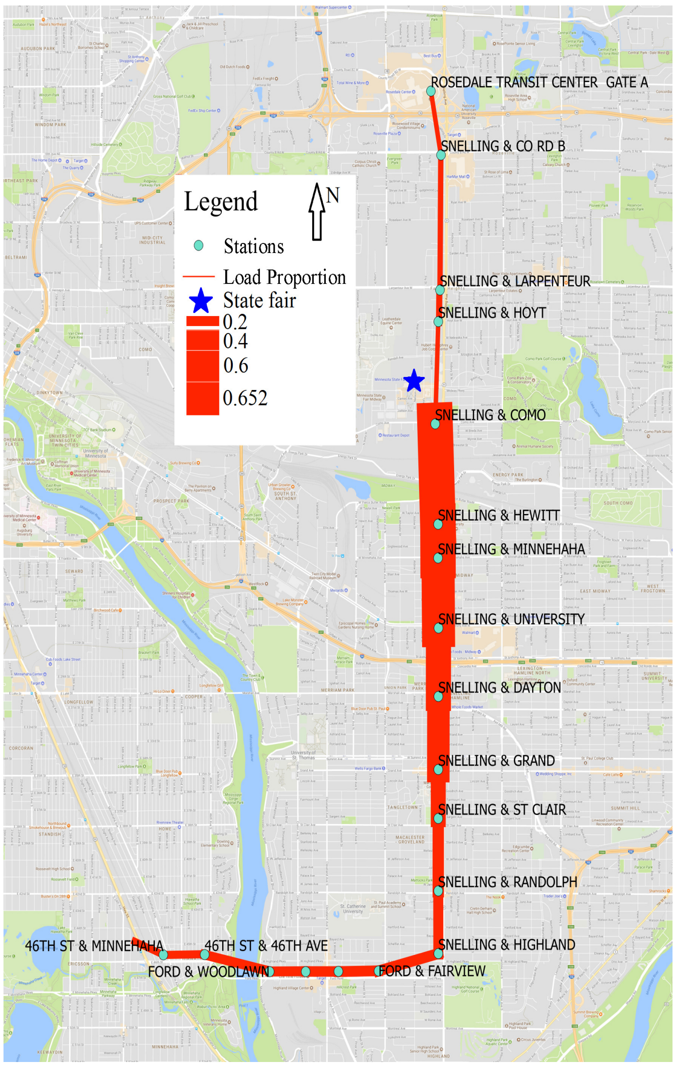


Fig. 5: Passenger load in southbound direction

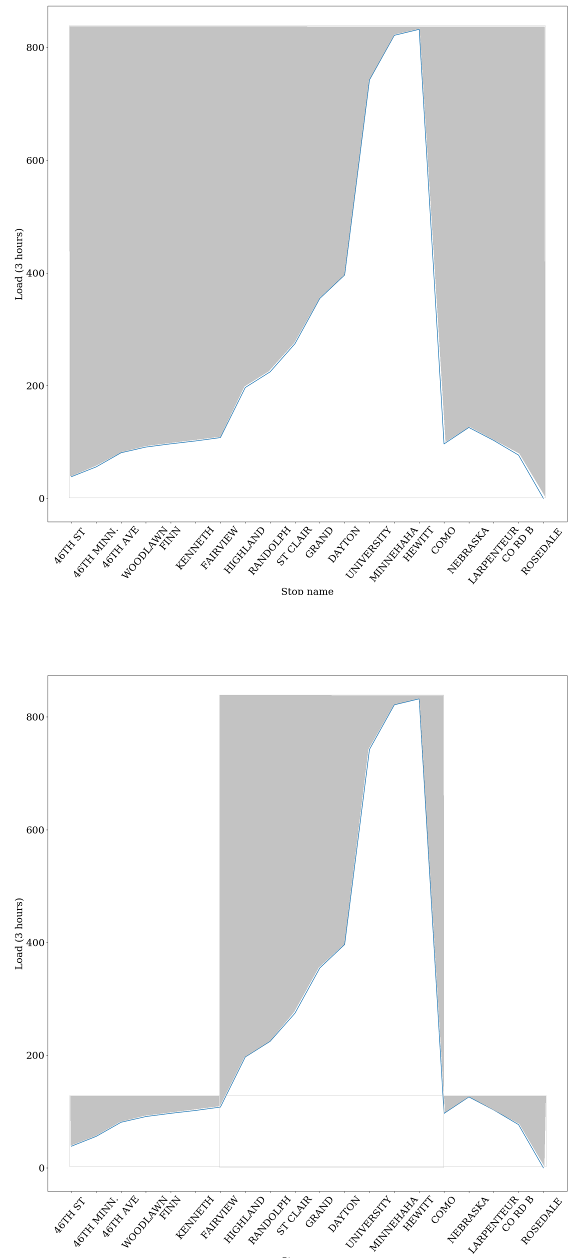


Fig. 6: Unused capacity (shaded area) by increasing the frequency

ACKNOWLEDGMENT

This research is conducted at the University of Minnesota Transit Lab, currently supported by the following, but not limited to, projects:

- National Science Foundation, award CMMI-1637548
- National Science Foundation, award CMMI-1831140
- Minnesota Department of Transportation, Contract No. 1003325 Work Order No. 15
- Minnesota Department of Transportation, Contract No. 1003325 Work Order No. 44
- Transitways Research Impact Program (TIRP), Contract No. A100460 Work Order No. UM2917

The authors are grateful to Metropolitan Council for sharing the data. We are also grateful to the anonymous referees for their constructive input to improve the quality of this article. Any limitation of this study remains the responsibility of the authors.

APPENDIX A
NOTATIONS USED IN THIS ARTICLE

Variable	Definition
N	Set of stops/stations along a transit route
D	Set of days in our analysis period
H	Set of time intervals in a day
T	Day-time mapping
K	Set of origin-destination pairs
B	Boarding matrix
A	Alighting matrix
R	Reference set used for computing mean and covariance matrix
$\mu_m(t)$	Mean flow vector for time interval t
$\Sigma_m(t)$	Covariance matrix for flows during time t
$\mathcal{M}(t)$	Mahalanobis distance for time interval t
M	Flow matrix
L	Regular (low-rank) flow matrix
S	Outlier (sparse) flow matrix
$\|M\|$	Spectral norm of a matrix M
$rank(M)$	Rank of a matrix M
M^T	Transpose of a matrix M
$\mathcal{C}(M)$	Column space of matrix M
$supp(M)$	Support set of matrix M
$\|S\|_0$	l_0 norm of a matrix M, $\|S\|_0 = \lim_{p\to 0}\sum_{i,j}\|M_{ij}\|^p$
$\|L\|_*$	nuclear norm of matrix M, $\|M\|_* = \sum_i \sigma_i(M)$
$\sigma_i(M)$	i^{th} singular value of matrix M, $\sigma_i(M) = \sqrt{\lambda_i(M^TM)}$
$\lambda_i(M)$	i^{th} eigen value of matrix M
$\|M\|_1$	l_1 norm of matrix M, $\|M\|_1 = \sum_{ij}\|M_{ij}\|$
$\|M\|_\infty$	l_∞ norm of matrix M, $\|M\|_\infty = \max_{i,j}\|M_{ij}\|$
N	Noise matrix
L_f	Lipshitz constant
$\langle,\rangle$	Inner product

REFERENCES

[1] J. L. Carson and R. G. Bylsma, "Transportation Planning and Management for Special Events - A Synthesis of Highway Practice," *Transportation Research Board*, 2003.

[2] J. Skolnik, R. Chami, and M. D. Walker, "Planned Special Events - Economic Role and Congestion Effects," *Federal Highway Administration*, no. August, 2008. [Online]. Available:

[3] Florida Department of Transportation Research Report, "Special Event Transportation Service Planning and Operations Strategies for Transit," *National special events, transportation planning, Technical Information Service (NTIS)*, no. March, 2006.

[4] A. Kuppam, R. Copperman, J. Lemp, T. Rossi, V. Livshits, L. Vallabhaneni, K. Jeon, and E. Brown, "Special events travel surveys and model development," *Transportation Letters*, vol. 5, no. 2, pp. 67–82, 2014.

[5] W. Wang, J. P. Attanucci, and N. H. M. Wilson, "Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems," *Journal of Public Transportation*, vol. 14, no. 4, pp. 131–150, 2011.

[6] N. Nassir, A. Khani, S. Lee, H. Noh, and M. Hickman, "Transit Stop-Level Origin-Destination Estimation Through Use of Transit Schedule and Automated Data Collection System," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2263, pp. 140–150, 2011. [Online]. Available:

[7] P. Kumar, A. Khani, and Q. He, "A robust method for estimating transit passenger trajectories using automated data," *Transportation Research Part C: Emerging Technologies*, vol. 95, 2018.

[8] P. Kumar, A. Khani, and G. A. Davis, "Transit Route Origin–Destination Matrix Estimation using Compressed Sensing," *Transportation Research Record: Journal of the Transportation Research Board*, p. 036119811984589, 2019.

[9] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust Principal Component Analysis?" *Journal of the ACM*, vol. 58, no. 3, 2011. [Online]. Available:

[10] P. Mahalanobis, "on the Generalized Distance in Statistics," pp. 109–115, 2013. [Online]. Available:

[11] S. Latoski, W. Dunn Jr, B. Wagenblast, J. Randall, and M. Walker, "Managing travel for planned special events," *US Department of Transportation*, vol. 1, no. September, p. 11, 2003. [Online]. Available:

[12] E. Chen, Z. Ye, C. Wang, and M. Xu, "Subway Passenger Flow Prediction for Special Events Using Smart Card Data," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, pp. 1–12, 2019. [Online]. Available:

[13] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015. [Online]. Available:

[14] Y. Li, X. Wang, S. Sun, X. Ma, and G. Lu, "Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 306–328, 2017. [Online]. Available:

[15] R. Xue, D. J. Sun, and S. Chen, "Short-Term Bus Passenger Demand Prediction Based on Time Series Model and Interactive Multiple Model Approach," *Discrete Dynamics in Nature and Society*, vol. 2015, no. i, pp. 1–11, 2015.

[16] M. Van Der Voort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transportation Research Part C: Emerging Technologies*, vol. 4, no. 5, pp. 307–318, 1996.

[17] M. C. Tan, S. C. Wong, J. M. Xu, Z. R. Guan, and P. Zhang, "An aggregation approach to short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 60–69, 2009.

[18] Y. Zhang and Z. Ye, "Short-Term Traffic Flow Forecasting Using Fuzzy Logic System Methods," *Journal of Intelligent Transportation Systems*, vol. 12, no. 3, pp. 102–112, 2008. [Online]. Available:

[19] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 871–882, 2013.

[20] F. C. Pereira, F. Rodrigues, and M. Ben-Akiva, "Using Data From the Web to Predict Public Transport Arrivals Under Special Events Scenarios," *Journal of Intelligent Transportation Systems*, vol. 19, no. 3, pp. 273–288, 2015. [Online]. Available:

[21] M. Ni, Q. He, and J. Gao, "Forecasting the Subway Passenger Flow under Event Occurrences with Social Media," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1623–1632, 2017.

[22] K. I. Wong and S. A. Yu, "Estimation of origin-destination matrices for mass event: A case of Macau Grand Prix," *Journal of King Saud University - Science*, vol. 23, no. 3, pp. 281–292, 2011. [Online]. Available:

[23] A. Webb, P. Kumar, and A. Khani, "Estimation of passenger waiting time using automatically collected transit data," *Public Transport*, pp. 1–13, 2020.

[24] G. A. Davis, "A STATISTICAL THEORY FOR ESTIMATION OF ORIGIN-DESTINATION PARAMETERS FROM TIME-SERIES OF TRAFFIC COUNTS." in *Proceedings of 12th International Symposium on Transportation and Traffic Theory*, 1993.

[25] B. Li, "Markov models for Bayesian analysis about transit route origin-destination matrices," *Transportation Research Part B: Methodological*, vol. 43, no. 3, pp. 301–310, 2009. [Online]. Available:

[26] S. C. Wong and C. O. Tong, “Estimation of time-dependent origin-destination matrices for transit networks,” *Transportation Research Part B: Methodological*, vol. 32, no. 1, pp. 35–48, 1998.

[27] I. Jolliffe, *Principal component analysis*. Springer, 2011.

[28] F. D. Torre and M. J. Black, “A Framework for Robust Subspace Learning,” *International Journal of Computer Vision*, vol. 54, no. 1, pp. 117–142, 2002.

[29] R. Gnanadesikan and J. R. Kettenring, “Robust estimates, residuals, and outlier detection with multiresponse data,” *Biometrics*, pp. 81–124, 1972.

[30] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[31] H. Xu, C. Caramanis, and S. Sanghavi, “Robust PCA via Outlier Pursuit,” *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3047–3064, 2010. [Online]. Available:

[32] C. Guyon, T. Bouwmans, and E.-H. Zahzah, “Moving object detection via robust low rank matrix decomposition with IRLS scheme,” in *International symposium on visual computing*. Springer, 2012, pp. 665–674.

[33] X. Ding, L. He, and L. Carin, “Bayesian robust principal component analysis,” *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3419–3430, 2011.

[34] T. Bouwmans and E. H. Zahzah, “Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance,” *Computer Vision and Image Understanding*, vol. 122, pp. 22–34, 2014.

[35] B. Donovan and D. B. Work, “Empirically quantifying city-scale transportation system resilience to extreme events,” *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 333–346, 2017.

[36] P. Filzmoser, “A multivariate outlier detection method,” *Seventh International Conference on Computer Data Analysis and Modeling*, vol. 1, no. 1989, pp. 18–22, 2004. [Online]. Available:

[37] H. Hotelling, “Analysis of a complex of statistical variables into principal components.” *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933. [Online]. Available:

[38] C. Eckart and G. Young, “The Approximation of One Matrix by Another Low Rank,” *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[39] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, “Sparse and low-rank matrix decompositions,” *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 15, no. PART 1, pp. 1493–1498, 2009.

[40] E. Candès and T. Tao, “Decoding by Linear ProgrammingEmmanuel Candes†,” *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[41] E. Candes and M. Wakin, “An Introduction To Compressive Sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[42] B. M. Sanandaji and P. P. Varaiya, “Compressive Origin-Destination Matrix Estimation,” *Transportation Letters*, vol. 8, no. 3, pp. 148–157, 2014. [Online]. Available:

[43] P. Ye and D. Wen, “Optimal Traffic Sensor Location for Origin-Destination Estimation Using a Compressed Sensing Framework,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1857–1866, 2017.

[44] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[45] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization,” vol. 52, no. 3, pp. 471–501, 2007. [Online]. Available:

[46] Z. Zhou, X. Li, J. Wright, E. Cand??s, and Y. Ma, “Stable principal component pursuit,” *IEEE International Symposium on Information Theory - Proceedings*, pp. 1518–1522, 2010.

[47] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, “Rank-Sparsity Incoherence for Matrix Decomposition,” *SIAM Journal on Optimization*, vol. 21, no. 2, pp. 572–596, 2011. [Online]. Available:

[48] Z. Lin, A. Ganesh, J. Wright, and L. Wu, “Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix,” *Computational Advances in ...*, pp. 1–18, 2009. [Online]. Available:

[49] Z. Lin, M. Chen, and Y. Ma, “The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices,” 2010. [Online]. Available:

[50] A. Beck and M. Teboulle, “A Fast Iterative Shrinkage-Thresholding Algorithm,” *Society for Industrial and Applied Mathematics Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[51] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, “Bregman iterative algorithms for L1-minimization with applications to compressed sensing,” *Siam*, vol. 1, no. 1, pp. 143–168, 2008.

[52] B. Recht and P. A. Parrilo, “Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization,” *SIAM Review*, vol. 52, no. 3, pp. 1–33, 2008. [Online]. Available:

[53] MSF, “Minnesota State Fair,” 2018. [Online]. Available:

[54] P. Kumar and A. Khani, “Outlier-OD-Estimation: Special event transit flow matrix estimation,” mar 2018. [Online]. Available:

[55] H. Zhang, Z. Lin, C. Zhang, and E. Y. Chang, “Exact Recoverability of Robust PCA via Outlier Pursuit with Tight Recovery Bounds,” *AAAI Conference on Artificial Intelligence*, pp. 3143–3149, 2015.

Pramesh Kumar received the Master of Science in Civil Engineering and Industrial Engineering from the University of Minnesota in 2020, and the Bachelor of Technology degree in Civil Engineering from the Indian Institute of Technology Roorkee, India in 2015. He is currently a PhD candidate in the Department of Civil, Environmental, and Geo-Engineering of the University of Minnesota. He is interested in transportation network optimization problems and emerging technologies.

Alireza Khani Alireza Khani received the PhD degree in Civil Engineering and Engineering Mechanics from the University of Arizona in 2013. He is currently an assistant professor in the department of Civil, Environmental, and Geo- Engineering at the University of Minnesota. His research interests include transportation network modeling, planning and operations of transit systems, and transit ITS data.