ROBUST NETWORK TRAFFIC ESTIMATION VIA SPARSITY AND LOW RANK

Morteza Mardani and Georgios B. Giannakis

Dept. of ECE, University of Minnesota, Minneapolis, MN

ABSTRACT

Accurate estimation of origin-to-destination (OD) traffic flows provides valuable input for network management tasks. However, lack of flow-level observations as well as intentional and unintentional anomalies pose major challenges toward achieving this goal. Leveraging the low intrinsic-dimensionality of OD flows and the sparse nature of anomalies, this paper proposes a convex program with nuclear-norm and ℓ_1-norm regularization terms to estimate the nominal and anomalous traffic components, using a *small subset* of (possibly anomalous) flow counts in addition to link counts. Analysis and simulations confirm that the said estimator can *exactly* recover sufficiently low-dimensional nominal traffic and sparse enough anomalies when the routing matrix is column-incoherent, and an adequate amount of flow counts are randomly sampled. The results offer valuable insights about the measurement types and network scenaria giving rise to accurate traffic estimation. Tests with real Internet data corroborate the effectiveness of the novel estimator.

Index Terms— Sparsity, low rank, traffic estimation.

1. INTRODUCTION

Monitoring origin-to-destination (OD) traffic flows over operational Internet Protocol (IP) networks is of paramount interest for quality-of-service provisioning and capacity planning. Direct measurement of all *nominal* OD flows however, is impossible due to the huge number of OD pairs, lack of measurement infrastructure, and potential anomalies arising due to cyber-attacks and network failures [9]. Typically, the available data sources are: D1) link counts comprising the superposition of OD flows per link; these counts can be readily obtained using the single network management protocol (SNMP); and D2) *partial* (possibly anomalous) flow counts recorded via NetFlow [9]. Traffic estimation is an arduous task because the number of unknown OD flows far exceeds the number of observations.

Relation to prior work. Given D1 and/or D2 measurements, ample research has been carried out to tackle the ill-posed traffic estimation problem using different inference techniques that leverage traffic features; see e.g., [7], [17] and references therein. Several studies have demonstrated that nominal traffic exhibits low intrinsic dimensionality, which is mainly due to common temporal patterns across OD flows, and periodic behaviors across time [9]. Moreover, traffic spikes (anomalies) are rare across time and flows, and tend to last for short periods of time relative to the measurement horizon. Capitalizing on these traffic features, data of the type D1 were used recently to unveil network anomalies with remarkable performance guarantees [10]. Without OD flows however, the nominal flow-level traffic cannot be identified using the approach of [10].

This paper's contribution: Building on [10], the present work utilizes data D1 jointly with data D2 to estimate the nominal and anomalous components of OD flow traffic. The fresh look advocated here permeates benefits from rank minimization and compressive sensing to traffic estimation. Leveraging the ability of the nuclear-norm and the ℓ_1-norm to recover low-rank and sparse components, a convex program is formulated to estimate the unknowns. Exact recovery is studied in the absence of noise, using a deterministic approach along the lines of [4], [10]. Introducing the notion of incoherence (angle) between a pair of subspaces, sufficient conditions are developed under which the convex program can exactly recover the nominal and anomalous traffic components.

The results yield valuable insights about network and measurement characteristics giving rise to an accurate traffic estimation. Intuitively, one can expect accurate recovery if:
a) NetFlow measures sufficently many randomly selected OD flows;
b) the OD routing paths form a column-incoherent routing matrix;
c) the nominal traffic is sufficiently low dimensional; and
d) anomalies are sporadic enough.

In addition, the convex optimization approach to robust traffic estimation opens the door for efficient in-network and online processing along the lines of [11] and [12]. Simulations with synthetic and real Internet data corroborate the effectiveness of the novel scheme.

Notation: Operators $(\cdot)'$ and $\oplus$ will denote transposition and subspace direct-sum, respectively; $|\cdot|$, $\|\mathbf{x}\|$ will denote the cardinality of a set and the ℓ_2-norm of a vector, respectively. For matrix $\mathbf{A} \in \mathbb{R}^{n\times p}$ denotes $\|\mathbf{A}\|_F$ the Frobenious norm and $\|\mathbf{A}\|_\infty := \max_{i,j}|a_{ij}|$ the ℓ_∞-norm. The $n \times n$ identity matrix will be represented by $\mathbf{I}_n$, and its i-th column by $\mathbf{e}_i$; likewise $\mathbf{0}_{n\times p} := \mathbf{0}_n\mathbf{0}'_p$. Define also the support set $\text{supp}(\mathbf{A}) := \{(i,j) : a_{ij} \neq 0\}$.

2. PRELIMINARIES AND PROBLEM STATEMENT

Consider a backbone IP network described by the directed graph $G(\mathcal{N},\mathcal{L})$, where $\mathcal{L}$ and $\mathcal{N}$ denote the set of links and nodes (routers) of cardinality $|\mathcal{L}| = L$ and $|\mathcal{N}| = N$, respectively. A set of end-to-end flows $\mathcal{F}$ with $|\mathcal{F}| = F$ traverse different OD pairs. In backbone networks, the number of OD flows is much larger than the number of physical links ($F \gg L$). Per traffic-flow, single-path routing is considered from each origin to its intended destination. Accordingly, for a particular flow multiple links connecting the corresponding OD node pair along a single path are chosen to carry the traffic. Let $r_{l,f}$ denote the flow $f \in \mathcal{F}$ to link $l \in \mathcal{L}$ routing assignment taking the value one whenever flow f traverses link l, and zero otherwise. The routing matrix $\mathbf{R} := [r_{l,f}] \in \{0,1\}^{L\times F}$ is assumed fixed and given. Likewise, let $x_{f,t}$ denote the unknown traffic rate of flow f at time t. The traffic carried over link l is then the superposition of flows routed through link l, that is, $\sum_{f\in\mathcal{F}} r_{l,f}x_{f,t}$.

It is not uncommon for some of the flow rates to experience unusual sudden changes, which are termed *traffic volume anomalies* and are typically due to the network failures, or cyber attacks [9]. With $a_{f,t}$ denoting the unknown traffic volume anomaly of flow f at time t, the measured traffic carried by link l at time t is then given

Work was supported by an AFOSR MURI grant no. FA 9550-10-1-0567.

by

$$y_{l,t} = \sum_{f \in \mathcal{F}} r_{l,f}(x_{f,t} + a_{f,t}) + v_{l,t},\ t = 1, ..., T \quad (1)$$

where $v_{l,t}$ accounts for the measurement errors. In IP networks, link loads can be readily measured via SNMP supported by most routers [9]. Introducing the matrices $\mathbf{Y} := [y_{l,t}], \mathbf{V} := [v_{l,t}] \in \mathbb{R}^{L \times T}$, $\mathbf{X} := [x_{f,t}]$, and $\mathbf{A} := [a_{f,t}], \in \mathbb{R}^{F \times T}$, link counts in (1) can be expressed in a compact matrix form as

$$\mathbf{Y} = \mathbf{R}(\mathbf{X} + \mathbf{A}) + \mathbf{V}. \quad (2)$$

Matrices $\mathbf{X}$ and $\mathbf{A}$ contain, respectively, the nominal and anomalous traffic flows over the time horizon of interest (T). Inferring $\{\mathbf{X}, \mathbf{A}\}$ from the compressed measurements $\mathbf{Y}$ and knowledge of $\mathbf{R}$ is a difficult task, especially because $L \ll F$.

Typically, additional data sources are utilized to enhance estimation accuracy. A useful such source is the direct flow-level measurements expressed per flow f as

$$z_{f,t} = x_{f,t} + a_{f,t} + w_{f,t},\ t = 1, ..., T \quad (3)$$

where $w_{f,t}$ accounts for measurement errors. The flow traffic (3) is sampled via NetFlow [9] at each origin node. However, due to the high cost of NetFlow one can only partially measure (3) [9]. To account for missing flow-level data, collect the available pairs (f, t) in the set $\Omega \in [1, 2, ..., F] \times [1, 2, ..., T]$. Introduce also the matrices $\mathbf{Z}_\Omega := [z_{f,t}], \mathbf{W}_\Omega := [w_{f,t}] \in \mathbb{R}^{F \times T}$, where $z_{f,t} = w_{f,t} = 0$ for $(f, t) \notin \Omega$, and associate the sampling operator $\mathcal{P}_\Omega$ with the set Ω, which assigns entries of its matrix argument not in Ω equal to zero, and keeps the rest unchanged. The flow counts in (3) can then be compactly written as

$$\mathbf{Z}_\Omega = \mathcal{P}_\Omega(\mathbf{X} + \mathbf{A}) + \mathbf{W}_\Omega. \quad (4)$$

Common temporal patterns among traffic flows in addition to their periodic behavior, render most rows (correspondingly columns) of $\mathbf{X}$ linearly dependent, and thus $\mathbf{X}$ typically exhibits low rank [9]. Anomalies on the other hand are expected to occur sporadically over time, and only last for short periods relative to the (possibly long) measurement interval $[1, T]$. In addition, only a small fraction of flows are supposed to be anomalous at a any given time instant. This renders the matrix $\mathbf{A}$ sparse across both rows and columns.

Given the link counts $\mathbf{Y}$ obeying (2) with the partial flow-counts $\mathbf{Z}_\Omega$ adhering to (4), and with $\{\mathbf{R}, \Omega\}$ known, this paper aims at accurately estimating the unknown *low-rank* nominal and *sparse* anomalous traffic components, namely $\mathbf{X}$ and $\mathbf{A}$.

3. TRAFFIC AND ANOMALY MATRIX RECOVERY

Consider first the noiseless scenario, i.e., $\mathbf{V} = \mathbf{0}_{L \times T}$ and $\mathbf{W}_\Omega = \mathbf{0}_{F \times T}$. A natural estimator accounting for the low rank of $\mathbf{X}$ and the sparsity of $\mathbf{A}$ will be sought to minimize the rank of $\mathbf{X}$, and the number of nonzero entries of $\mathbf{A}$ measured by its ℓ_0-(pseudo) norm. Unfortunately, both rank and ℓ_0-norm minimization problems are in general NP-hard [14, 15]. The nuclear-norm $\|\mathbf{X}\|_* := \sum_k \sigma_k(\mathbf{X})$, where $\sigma_k(\mathbf{X})$ denotes the k-th singular value of $\mathbf{X}$, and the ℓ_1-norm $\|\mathbf{A}\|_1 := \sum_{f,t} |a_{f,t}|$ are typically adopted as *convex* surrogates [15, 1]. Accordingly, one solves

$$\text{(P1)} \qquad \min_{\{\mathbf{X},\mathbf{A}\}} \|\mathbf{X}\|_* + \lambda \|\mathbf{A}\|_1$$
$$\text{s.t.} \quad \mathbf{Y} = \mathbf{R}(\mathbf{X} + \mathbf{A}), \quad \mathbf{Z}_\Omega = \mathcal{P}_\Omega(\mathbf{X} + \mathbf{A})$$

where $\lambda \geq 0$ is the sparsity-controlling parameter.

It is worth mentioning that (P1) subsumes several important special cases, which yield accurate recovery of $\{\mathbf{X}, \mathbf{A}\}$ [10, 4]. In the absence of flow counts for instance, exact recovery of the *sparse* matrix $\mathbf{A}$ from link loads is reported in [10]. The key to this is the sparsity present, which enables recovery from compressed linear-measurements. However, the (possibly huge) nullspace of $\mathbf{R}$ challenges identifiability of the nominal traffic matrix $\mathbf{X}$, as validated also via extensive simulated tests. As another instance, suppose only flow counts are available, in which case (P1) boils down to the robust principal component analysis with missing data, for which one can exactly recover the *low-rank* component [4]. Instrumental role in this case is played by the dependencies among entries of the low-rank component, reflected in the observations. Indeed, the anomaly matrix is not recoverable since observed entries do not convey any information about the unobserved anomalies.

These considerations regarding recovery in the aforementioned special cases, makes one hopeful to retrieve $\mathbf{X}$ and $\mathbf{A}$ via (P1). Note that [6] has recently studied the recovery of compressed low-rank-plus-sparse matrices, where the compression is an *orthogonal projection* onto a low-dimensional subspace. However, it does not necessarily subsume the present model since in general the latter may not entail an orthogonal projection of a low-rank-plus-sparse matrix onto a subspace. In the sequel, the recovery performance of (P1) is analyzed.

4. LOCAL IDENTIFIABILITY

Let $\mathbf{X}_0$ and $\mathbf{A}_0$ be the *true* low-rank and sparse matrices of interest with $r := \text{rank}(\mathbf{X}_0)$ and $s := \|\mathbf{A}_0\|_0$. The first issue to address is identifiability, which asserts that there is a *unique* pair $\{\mathbf{X}_0, \mathbf{A}_0\}$ fulfilling the data constraints: d1) $\mathbf{Y} = \mathbf{R}(\mathbf{X}_0 + \mathbf{A}_0)$ and d2) $\mathbf{Z}_\Omega = \mathcal{P}_\Omega(\mathbf{X}_0 + \mathbf{A}_0)$. Apparently, if multiple solutions exist, one cannot hope finding $\{\mathbf{X}_0, \mathbf{A}_0\}$. Before examining this issue, introduce the subspaces: s1) $\mathcal{N}_R := \{\mathbf{H} : \mathbf{R}\mathbf{H} = \mathbf{0}_{L \times T}\}$ as the nullspace of the linear operator $\mathbf{R}$, and s2) $\mathcal{N}_\Omega := \{\mathbf{H} \in \mathbb{R}^{F \times T} : \text{supp}(\mathbf{H}) \subseteq \Omega^\perp\}$ as the nullspace of the linear operator $\mathcal{P}_\Omega(.)$ [$\Omega^\perp$ is the complement of Ω]. Now, if there exist perturbations $\{\mathbf{H}_1, \mathbf{H}_2\}$ with $\mathbf{H}_1 + \mathbf{H}_2 \in \mathcal{N}_R \cap \mathcal{N}_\Omega$ so that $\mathbf{X}_0 + \mathbf{H}_1$ and $\mathbf{A}_0 + \mathbf{H}_2$ are still low-rank and sparse, one may pick the pair $\{\mathbf{X}_0 + \mathbf{H}_1, \mathbf{A}_0 + \mathbf{H}_2\}$ as another legitimate solution. This section aims at resolving such identifiability issues.

Let $\mathbf{U}_0 \mathbf{\Sigma}_0 \mathbf{V}_0'$ denote the singular value decomposition (SVD) of $\mathbf{X}_0$, and consider the subspaces: s3) $\Phi_{X_0} := \{\mathbf{Z} \in \mathbb{R}^{F \times T} : \mathbf{Z} = \mathbf{U}_0 \mathbf{W}_1' + \mathbf{W}_2 \mathbf{V}_0',\ \mathbf{W}_1 \in \mathbb{R}^{T \times r},\ \mathbf{W}_2 \in \mathbb{R}^{F \times r}\}$ of matrices in either the column or row space of $\mathbf{X}_0$; s4) $\Omega_{A_0} := \{\mathbf{H} \in \mathbb{R}^{F \times T} : \text{supp}(\mathbf{H}) \subseteq \text{supp}(\mathbf{A}_0)\}$ of matrices whose support is contained in that of $\mathbf{A}_0$. Noteworthy properties of these subspaces are: i) both Φ_{X_0} and $\Omega_{A_0} \subset \mathbb{R}^{F \times T}$, hence it is possible to directly compare elements from them; ii) $\mathbf{X}_0 \in \Phi_{X_0}$ and $\mathbf{A}_0 \in \Omega_{A_0}$; and iii) if $\mathbf{Z} \in \Phi_{X_0}^\perp$ is added to $\mathbf{X}_0$, then $\text{rank}(\mathbf{Z} + \mathbf{X}_0) > r$, and likewise if $\mathbf{Z} \in \Omega^\perp$ is added to $\mathbf{A}_0$, then $|\text{supp}(\mathbf{A}_0 + \mathbf{Z})| > s$.

Suppose temporarily that the subspaces Φ_{X_0} and Ω_{A_0} are also known. This extra piece of information helps identifiability based on data d1) and d2), since the potentially troublesome solutions

$$\Upsilon_1 := \{(\mathbf{X}_0 + \mathbf{H}_1, \mathbf{A}_0 + \mathbf{H}_2) : \mathbf{H}_1 + \mathbf{H}_2 \in \mathcal{N}_R \cap \mathcal{N}_\Omega\} \quad (5)$$

are restricted to a smaller set. If $(\mathbf{X}_0 + \mathbf{H}_1, \mathbf{A}_0 + \mathbf{H}_2) \notin \Upsilon_2$, where

$$\Upsilon_2 := \{(\mathbf{X}_0 + \mathbf{H}_1, \mathbf{A}_0 + \mathbf{H}_2) : \mathbf{H}_1 \in \Phi_{X_0},\ \mathbf{H}_2 \in \Omega_{A_0}\} \quad (6)$$

that candidate solution is not admissible since it is known a priori that $\mathbf{X}_0 \in \Phi_{X_0}$ and $\mathbf{A}_0 \in \Omega_{A_0}$. This notion of exploiting additional knowledge to assure uniqueness is known as *local identifiability* [4].

Global identifiability from d1) and d2) is not guaranteed. However, local identifiability will become essential later on to establish the main result. Under these assumptions, the following Lemma puts forth the necessary and sufficient conditions for local identifiability.

Lemma 1. *Matrices* $\{\mathbf{X}_0, \mathbf{A}_0\}$ *satisfy d1) and d2) uniquely if and only if: c1)* $\Phi_{X_0} \cap \Omega_{A_0} = \{\mathbf{0}\}$*; and, c2)* $\Upsilon_1 \cap \Upsilon_2 = \{\mathbf{0}\}$.[1]

Condition c1) implies that for the solutions in Υ_2 to be admissible, $\mathbf{H}_1 + \mathbf{H}_2$ must belong to the subspace $\Phi_{X_0} \oplus \Omega_{A_0}$. Accordingly, c2) holds if

$$\mathcal{N}_R \cap \mathcal{N}_\Omega \cap (\Phi_{X_0} \oplus \Omega_{A_0}) = \{\mathbf{0}\}. \tag{7}$$

Notice that c1) appears also in the context of low-rank-plus-sparse recovery results in [2, 4]. However, c2) is unique to the setting here. It captures the impact of the overlap between nullspace of $\mathbf{R}$ and the sampling operator $\mathcal{P}_\Omega(.)$. Finding simpler sufficient conditions to assure c1) and c2) is studied next.

4.1. Incoherence Measures

The overlap between any pair of subspaces $\{\Phi_{X_0}, \Omega_{A_0}, \mathcal{N}_R, \mathcal{N}_\Omega\}$ plays a crucial role in identifiability and exact recovery as seen e.g., from Lemma 1. To quantify the overlap of the subspaces e.g., Φ_{X_0} and Ω_{A_0}, consider the *incoherence* parameter

$$\mu(\Phi_{X_0}, \Omega_{A_0}) := \max_{\substack{\mathbf{X} \in \Omega_{A_0} \\ \|\mathbf{X}\|_F = 1}} \|\mathcal{P}_{\Phi_{X_0}}(\mathbf{X})\|_F. \tag{8}$$

Observe that $\mu(\Phi_{X_0}, \Omega_{A_0}) \in [0, 1]$. The lower bound is achieved when Φ_{X_0} and Ω_{A_0} are orthogonal, whereas the upperbound is attained when $\Phi_{X_0} \cap \Omega_{A_0}$ contains a nonzero element. To gain further geometric intuition, $\mu(\Phi_{X_0}, \Omega_{A_0})$ represents the cosine of the angle between subspaces when $\Phi_{X_0} \cap \Omega_{A_0} = \{\mathbf{0}\}$ [5]. Small values of $\mu(\Phi_{X_0}, \Omega_{A_0})$ indicate enough separation between Φ_{X_0} and Ω_{A_0}, and thus less chance of ambiguity when discerning $\mathbf{X}_0$ from $\mathbf{A}_0$.

It will be seen later that satisfying c1) requires $\mu(\Phi_{X_0}, \Omega_{A_0}) < 1$. In addition, to ensure c2) one needs the incoherence parameter $\mu(\mathcal{N}_R \cap \mathcal{N}_\Omega, \Phi_{X_0} \oplus \Omega_{A_0}) < 1$. In fact, $\mu(\mathcal{N}_R \cap \mathcal{N}_\Omega, \Phi_{X_0} \oplus \Omega_{A_0})$ captures the ambiguity inherent to the nullspace of the compression and sampling operators. It depends on all subspaces s1)-s4), and it is desirable to express it in terms of the incoherence of different subspace pairs, namely $\mu(\mathcal{N}_R, \Omega_{A_0})$, $\mu(\mathcal{N}_R, \Phi_{X_0})$, $\mu(\mathcal{N}_\Omega, \Omega_{A_0})$, and $\mu(\mathcal{N}_\Omega, \Phi_{X_0})$. This is formalized in the next claim.

Proposition 1. *Assume that* $\mu(\Omega_{A_0}, \Phi_{X_0}) < 1$. *If either* $\dim(\mathcal{N}_R \cap \mathcal{N}_\Omega) = 0$*; or,* $\dim(\mathcal{N}_R \cap \mathcal{N}_\Omega) \geq 1$ *and*

$$\xi := \left[\frac{\mu(\mathcal{N}_R, \Phi_{X_0})\mu(\mathcal{N}_\Omega, \Phi_{X_0}) + \mu(\mathcal{N}_R, \Omega_{A_0})\mu(\mathcal{N}_\Omega, \Omega_{A_0})}{1 - \mu(\Omega_{A_0}, \Phi_{X_0})}\right]^{1/2} < 1$$

hold, then $\Phi_{X_0} \cap \Omega_{A_0} = \{\mathbf{0}\}$ *and* $\mathcal{N}_R \cap \mathcal{N}_\Omega \cap (\Phi_{X_0} \oplus \Omega_{A_0}) = \{\mathbf{0}\}$.

Apparently, small values of $\mu(\mathcal{N}_R, \Omega_{A_0})$ and $\mu(\mathcal{N}_\Omega, \Phi_{X_0})$ can render ξ small enough. In fact, the incoherence $\mu(\mathcal{N}_R, \Omega_{A_0})$ measures whether $\mathcal{N}_R$ contains sparse elements, and it is tightly related to the incoherence among the sparse column-subsets of $\mathbf{R}$. Specifically, if $\mathbf{R}\mathbf{R}' = \mathbf{I}$, the incoherence reduces to the restricted isometry constant of $\mathbf{R}$ [1]. Moreover, $\mu(\mathcal{N}_\Omega, \Phi_{X_0})$ measures whether the low-rank matrices fall into the nullspace of the sampling operator, which is linked to the incoherence metrics introduced in the context of matrix completion; see e.g., [3]. It is also worth mentioning that a wide class of matrices resulting in small incoherence $\mu(\mathcal{N}_R, \Omega_{A_0})$, $\mu(\mathcal{N}_\Omega, \Phi_{X_0})$ and $\mu(\Omega_{A_0}, \Phi_{X_0})$ are provided in [1], [3], [2], which give rise to a sufficiently small value of ξ.

[1] The proofs in this work will be provided in the journal version [13]

5. EXACT RECOVERY VIA CONVEX OPTIMIZATION

Besides $\mu(\Omega_{A_0}, \Phi_{X_0})$ and ξ, there are other incoherence measures which play an important role in the conditions for exact recovery. These measures are introduced to avoid ambiguity when the (feasible) perturbations $\mathbf{H}_1$ and $\mathbf{H}_2$ do not necessarily belong to the subspaces Φ_{X_0} and Ω_{A_0}, respectively. Before moving on, it is worth noting that these measures resemble the ones for matrix completion and decomposition problems; see e.g., [2, 3]. For instance, consider a feasible solution $\{\mathbf{X}_0 + a_{i,j}\mathbf{e}_i\mathbf{e}_j', \mathbf{A}_0 + a_{i,j}\mathbf{e}_i\mathbf{e}_j'\}$, where $(i, j) \notin \text{supp}(\mathbf{A}_0)$ and thus $a_{i,j}\mathbf{e}_i\mathbf{e}_j' \notin \Omega_{A_0}$. It may happen that $a_{i,j}\mathbf{e}_i\mathbf{e}_j' \in \Phi_{X_0}$ and $\text{rank}(\mathbf{X}_0 + a_{i,j}\mathbf{e}_i\mathbf{e}_j') = \text{rank}(\mathbf{X}_0) - 1$, while $\|\mathbf{A}_0 - a_{i,j}\mathbf{e}_i\mathbf{e}_j'\|_0 = \|\mathbf{A}_0\|_0 + 1$, challenging identifiability when Φ_{X_0} and Ω_{A_0} are unknown. Similar complications arise if $\mathbf{X}_0$ has a sparse row space that can be confused with the row space of $\mathbf{A}_0$. These issues motivate defining

$$\gamma(\mathbf{U}_0) := \max_i \|\mathbf{P}_U \mathbf{e}_i\|, \qquad \gamma(\mathbf{V}_0) := \max_i \|\mathbf{P}_V \mathbf{e}_i\|$$

where $\mathbf{P}_U := \mathbf{U}_0\mathbf{U}_0'[\mathbf{P}_V := \mathbf{V}_0\mathbf{V}_0']$ are the projectors onto the column [row] space of $\mathbf{X}_0$. Notice that $\gamma(\mathbf{U}_0), \gamma(\mathbf{V}_0) \in [0, 1]$. The maximum of $\gamma(\mathbf{U}_0)[\gamma(\mathbf{V}_0)]$ is attained when $\mathbf{e}_i$ is in the column [row] space of $\mathbf{X}_0$ for some i. Small values of $\gamma(\mathbf{U}_0)[\gamma(\mathbf{V}_0)]$ imply that the column[row] spaces of $\mathbf{X}_0$ do not contain sparse vectors, respectively.

Another identifiability instance arises when $\mathbf{X}_0$ is sparse, in which case each column of $\mathbf{X}_0$ is spanned by a few canonical basis vectors. Consider the parameter

$$\gamma(\mathbf{U}_0, \mathbf{V}_0) := \|\mathbf{U}_0\mathbf{V}_0'\|_\infty = \max_{i,j} |\mathbf{e}_i{}'\mathbf{U}_0\mathbf{V}_0\mathbf{e}_j|.$$

A small value of $\gamma(\mathbf{U}_0, \mathbf{V}_0)$ indicates that each column of $\mathbf{X}_0$ is spanned by sufficiently many canonical basis vectors. It is worth noting that $\gamma(\mathbf{U}_0, \mathbf{V}_0)$ can be bounded in terms of $\gamma(\mathbf{U}_0)$ and $\gamma(\mathbf{V}_0)$, but it is kept here for the sake of generality.

From c2) in Lemma 1 it is evident that the dimension of the nullspace $\mathcal{N}_R \cap \mathcal{N}_\Omega$ is critical for identifiability. In essence, the lower $\dim(\mathcal{N}_R \cap \mathcal{N}_\Omega)$, the higher is the chance of exact reconstruction. In order to quantify the size of the nullspace, define

$$\tau(\mathcal{N}_R, \mathcal{N}_\Omega) := \max_{\substack{\mathbf{X} \in \mathcal{N}_R \cap \mathcal{N}_\Omega \\ \|\mathbf{X}\|_F = 1}} \|\mathbf{X}\|_\infty \tag{9}$$

which will appear later in the exact recovery conditions. All elements are now on place to state the main result in the next section.

5.1. Main Result

Theorem 5.1. *Consider given matrices* $\mathbf{Y} \in \mathbb{R}^{L \times T}$*, and* $\mathbf{R} \in \mathbb{R}^{L \times F}$ *obeying* $\mathbf{Y} = \mathbf{R}(\mathbf{X}_0 + \mathbf{A}_0)$ *together with the partial matrix* $\mathbf{Z}_\Omega \in \mathbb{R}^{F \times T}$*, sampled from the given set* Ω*, adhering to* $\mathbf{Z}_\Omega = \mathcal{P}_\Omega(\mathbf{X}_0 + \mathbf{A}_0)$. *Suppose that every column of* $\mathbf{A}_0$ *has at most* k *nonzero elements, and let* $\mathbf{X}_0 := \mathbf{U}_0\mathbf{\Sigma}_0\mathbf{V}_0'$, $r :=$ *rank*$(\mathbf{X}_0)$, $s := \|\mathbf{A}_0\|_0$. *If the following conditions*

I) $\lambda_{\max} > \lambda_{\min}$
II) $\theta := 1 - \mu - 2\xi^2 > 0$
III) $\nu := 1 - \mu^2 - \eta\sqrt{s}\mu - \beta\xi\sqrt{s} > 0$
IV) $\kappa := k + (\sqrt{s} - k)\mu^2 + \alpha\xi\sqrt{s} > 0$

hold, where $\mu := \mu(\Omega_{A_0}, \Phi_{X_0})$, $\eta := \gamma(\mathbf{U}_0) + \gamma(\mathbf{V}_0)$, $\tau := \tau(\mathcal{N}_R, \mathcal{N}_\Omega)$*, and* $\gamma := \gamma(\mathbf{U}_0, \mathbf{V}_0)$,

$$\alpha := \frac{(1 + \mu)[1 - \mu + \xi]}{\theta}, \quad \beta := \frac{(1 + \mu)[\tau(1 - \mu) + \eta\xi]}{\theta},$$

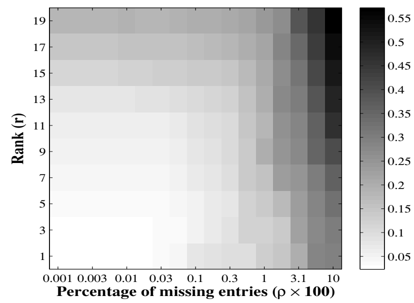

Fig. 1. Relative error for various values of rank (r) and sparsity level ($s = \rho FT$) when $F = T = 210$, $L = F/2$, and $\pi = 0.1$. White represents exact recovery ($e \approx 0$) while black represents $e \approx 1$.

and

$$\lambda_{\max} := \frac{1 - \mu^2 - \sqrt{r}(\mu - \alpha\xi)}{\nu}, \quad \lambda_{\min} := \frac{\gamma + \sqrt{r}(\eta\gamma\mu^2 + \beta)}{\kappa},$$

then there exists $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$ *for which the convex program (P1) exactly recovers* $\{\mathbf{X}_0, \mathbf{A}_0\}$.

Satisfaction of the conditions in Theorem 5.1 hinges upon the incoherence parameters $\{\mu, \gamma, \eta, \xi, \tau\}$ whose sufficiently small values fulfil I)-III). In fact, these parameters are increasing functions of the rank r and the sparsity level s. In particular, $\{\mu, \gamma, \eta\}$, capturing the ambiguity of the additive components $\mathbf{X}_0$ and $\mathbf{A}_0$, are known to be small enough for small values of $\{r, s, k\}$ [3, 4]. Regarding ξ, recall that it is an increasing function of $\mu(\Omega_{\mathbf{A}_0}, \mathcal{N}_R)$ and $\mu(\Phi_{\mathbf{X}_0}, \mathcal{N}_\Omega)$. Similar to $\mu(\Phi_{\mathbf{X}_0}, \Omega_{A_0})$, the parameter $\mu(\Phi_{\mathbf{X}_0}, \mathcal{N}_\Omega)$ takes a small value when NetFlow samples an adequately large subset of OD flows uniformly at random. Moreover, in large-scale networks with distant OD node pairs, and routing paths that are sufficiently 'spread-out', the sparse column-subsets of $\mathbf{R}$ tend to be incoherent, and thus $\mu(\Omega_{\mathbf{A}_0}, \mathcal{N}_R)$ takes a small value. Likewise, for sufficiently many NetFlow samples and column-incoherent routing matrices, τ takes a small value.

Remark 1 (**Satisfiability**). Notice that I)-III) in Theorem 5.1 are expressible in terms of the angle between subspaces s1)-s4). In general, they are NP-hard to verify. Introducing a class of (possibly random) traffic matrices $\{\mathbf{X}_0, \mathbf{A}_0\}$ and realistic network settings giving rise to a desirable routing matrix $\mathbf{R}$ is the subject of ongoing research and will be detailed in an extended report.

6. NUMERICAL TESTS

Performance of the (P1) solver is assessed in this section via computer simulations.

Exact recovery. Data matrix $\mathbf{Y}$ is generated according to $\mathbf{Y} = \mathbf{V}'_R(\mathbf{X}_0 + \mathbf{A}_0)$, where $\mathbf{V}_R \in \mathbb{R}^{F \times L}$ comprises the right singular vectors of the binary $\{0, 1\}$-valued random matrix $\mathbf{R} = \mathbf{U}_R \mathbf{\Sigma}_R \mathbf{V}'_R$ with i.i.d. entries equal to one with probability $1/2$. The low-rank component $\mathbf{X}_0$ is produced from the bilinear factorization $\mathbf{X}_0 = \mathbf{W}\mathbf{Z}'$, where $\mathbf{W}$ and $\mathbf{Z}$ are $L \times r$ and $T \times r$ matrices with i.i.d. entries drawn from Gaussian distributions $\mathcal{N}(0, 1/L)$ and $\mathcal{N}(0, 1/T)$, respectively. Every entry of $\mathbf{A}_0$ is randomly drawn from the set $\{-1, 0, 1\}$ with $\Pr(a_{i,j} = -1) = \Pr(a_{i,j} = 1) = \rho/2$. Likewise, the partial flow-traffic matrix $\mathbf{Z}_\Omega$ is generated according to $\mathbf{Z}_\Omega = \mathbf{\Omega} \odot (\mathbf{X}_0 + \mathbf{A}_0)$, where $\odot$ is the element-wise product. Entries of the binary $\{0, 1\}$-valued sampling matrix $\mathbf{\Omega} \in \mathbb{R}^{F \times T}$ are also i.i.d. taking the value one with probability π. Set also $T = F = 210$. To demonstrate that (P1) is capable of recovering the exact values of $\{\mathbf{X}_0, \mathbf{A}_0\}$, the optimization problem is solved for a wide range of values of (r, s). Let $(\hat{\mathbf{X}}, \hat{\mathbf{A}})$ denote the solution obtained via (P1) for a suitable value of λ. Fig. 1 depicts the relative error in recovering $(\mathbf{X}_0, \mathbf{A}_0)$, namely $e := \|\hat{\mathbf{A}} - \mathbf{A}_0\|_F / \|\mathbf{A}_0\|_F + \|\hat{\mathbf{X}} - \mathbf{X}_0\|_F / \|\mathbf{X}_0\|_F$ for various values of (r, s). It is apparent that (P1) succeeds in recovering $\{\mathbf{X}_0, \mathbf{A}_0\}$ for sufficiently sparse $\mathbf{A}_0$ and low-rank $\mathbf{X}_0$ from the observed data $\{\mathbf{Y}, \mathbf{Z}_\Omega\}$.

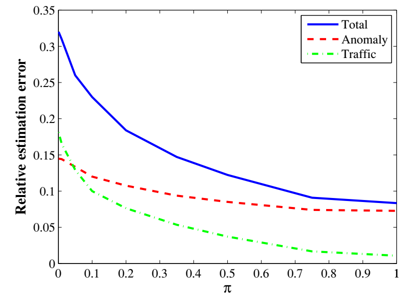

Fig. 2. Relative estimation error versus NetFlow sampling rate.

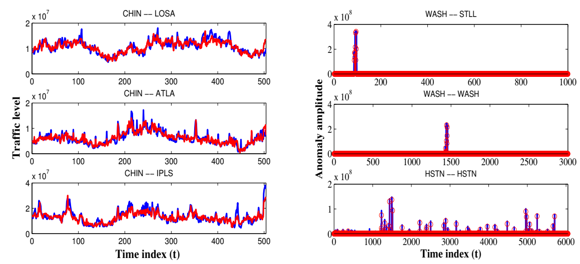

Fig. 3. Performance for representative flows of Internet-2 network when $\pi = 0.1$. (a) Estimated (red) versus true (blue) nominal traffic. (b) Estimated (circle markers) versus true (solid) anomalous traffic.

Real network data tests. Real data including OD-flow traffic-levels are collected from the operation of the Internet-2 network (Internet backbone network across USA) [18]. For further details about the setup refer to [12]. To study the impact of NetFlow sampling rate on the recovery performance, Fig. 2 depicts relative estimation error e for various amounts of NetFlow data. The candidate OD flows are selected independently with probability π. Here, the anomaly estimation error refers to $\|\hat{\mathbf{A}} - \mathbf{A}_0\|_F / \|\mathbf{A}_0\|_F$ (likewise, for the traffic estimation error). Naturally, the higher the sampling rate, the better estimation accuracy is attained. The improvement is seen to be more pronounced for recovering the nominal traffic relative to the anomalous traffic, so as for large π the anomalous-traffic estimation-error bottlenecks the total estimation error. Interestingly, when link loads are utilized alone ($\pi = 0$) to estimate $\{\mathbf{X}_0, \mathbf{A}_0\}$, adding 10% NetFlow samples improves the nominal-traffic estimation-error by 45%, and the anomalous-traffic estimation-error by 18%. This observation in turn corroborates the effectiveness of exploiting partial NetFlow data to improve the traffic estimation accuracy. For an instance of $\pi = 0.1$, the true and estimated nominal and anomalous traffic time-series for representative flows are depicted in Fig. 3.

7. REFERENCES

[1] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Info. Theory*, vol. 51, pp. 4203–4215, 2005.

[2] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. of the ACM*, vol. 58, pp. 1–37, 2011.

[3] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–722, 2009.

[4] V. Chandrasekaran, S. Sanghavi, P. R. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM J. Optim.*, vol. 21, no. 2, pp. 572–596, 2011.

[5] F. Deutsch, *Best Approximation in Inner Product Spaces*, Springer-Verlag, second edition, 2001.

[6] A. Ganesh, K. Min, J. Wright, and Y. Ma, "Principal component pursuit with reduced linear measurements," Available at arXiv:1202.6445v1 [cs.IT], 2012.

[7] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models,* Springer, 2009.

[8] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," in *Proc. of ACM SIGCOMM*, Portland, OR, Aug. 2004.

[9] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. D. Kolaczyk, and N. Taft, "Structural analysis of network traffic flows," in *Proc. of ACM SIGMETRICS*, NY, USA, Jul. 2004.

[10] M. Mardani, G. Mateos, and G. B. Giannakis, "Recovery of low-rank plus compressed sparse matrices with application to unveiling traffic anomalies," *IEEE Trans. Info. Theory*, 2012, see also arXiv:1204.6537v1 [cs.IT].

[11] M. Mardani, G. Mateos, and G. B. Giannakis, "In-network sparsity regularized rank minimization: Applications and algorithms," *IEEE Trans. Signal Process.*, 2012, see also arXiv:1203.1570v1 [cs.MA].

[12] M. Mardani, G. Mateos, and G. B. Giannakis, "Dynamic anomalography: Tracking network anomalies via sparsity and low-rank," *IEEE J. Sel. Tpoics in Signal Process.*, 2012, see also arXiv:1208.4043v1 [cs.NI].

[13] M. Mardani and G. B. Giannakis, "Robust network traffic estimation via sparsity and low rank: identifiability and algorithms," *IEEE Trans. Info. Theory*, 2012 (submitted).

[14] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, pp. 227–234, 1995.

[15] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear-norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[16] Y. Zhang, Z. Ge, A. Greenberg, and M. Roughan, "Network anomography," in *Proc. of ACM SIGCOMM Conf. on Interent Measurement*, CA, USA, Oct. 2005.

[17] Q. Zhao, Z. Ge, J. Wang, and J. Xu, "Robust traffic matrix estimation with imperfect information: making use of multiple data sources," in *Proc. of IEEE/ACM SIGMETRICS/Performance*, Saint Malo, France, Jun. 2006.

[18] http://internet2.edu/observatory/archive/data-collections.html