Cartography for Cognitive Networks

Acknowledgments: Prof. G. Mateos, MURI Grant No. AFOSR FA9550-10-1-0567
NSF Grants ECCS-1202135, AST-1247885, ECCS-1002180

Georgios B. Giannakis

A Coruña, Spain
June 22, 2014
Network traffic growth

Projected IP traffic in Exabytes/month

IP traffic is growing explosively

“Smart” devices multiply traffic

- Communication networks today
  - Large-scale interconnection of “smart” devices
  - Commercial, consumer-oriented, heterogeneous

Service diversification

Residential services

By 2016, Online Music will have 1.7 billion users, ranking as the highest penetrated service among all residential service categories (TV & Internet).

- 63% in 2011
- 79% in 2016

Voice over IP (VoIP) will be the fastest growing residential Internet service from 2011 to 2016.
- ▲ 10.6% CAGR

By 2016, Video on Demand (VoD) will be the highest penetrated residential TV service with 258 million subscribers.
- 14% of digital TV households.

Digital TV will be the fastest growing service across all residential service categories (TV & Internet) from 2011 to 2016.
- ▲ 13.8% CAGR

The average household will have 2.5 fixed devices/connections by 2016.

Mobile services

By 2016, Mobile SMS will have 4.1 billion users, ranking as the highest penetrated service among all consumer mobile services.

- 74% in 2011
- 90% in 2016

Mobile Video will be the fastest growing consumer mobile service from 2011 to 2016.
- ▲ 42.9% CAGR

Mobile Banking & Commerce will be the second fastest growing consumer mobile service from 2011 to 2016.
- ▲ 42.7% CAGR

By 2016, Mobile Social Networking will be the second highest penetrated consumer mobile service with 2.4 billion users.
- 53% of consumer mobile users.

By 2016, mobile consumers will have an average of 1.6 devices per person.

Dynamic network cartography

- Accurate network diagnosis and statistical analysis tools
  - Seamless end-user experience in dynamic environments
  - Secure and stable network operation

- **Network cartography**: succinct depiction of the *network state*
  - Tool for statistical modeling, monitoring and management
  - Offers situational awareness of the network landscape

---

Tutorial outlook

- Dynamic network delay and traffic cartography
  - Map network state via limited measurements
  - Monitor network health

- Dynamic anomalography for IP networks
  - Reveal where and when traffic anomalies occur
  - Leverage sparse anomalies and low-rank traffic

- RF cartography for cognition at the PHY
  - Map ambient RF power in space-time-frequency
  - Identify “crowded” regions to be avoided
General context: NetSci analytics

- Online social media
- Internet
- Clean energy and grid analytics
- Biological networks
- Robot and sensor networks
- Square kilometer array telescope

**General tools:** process, analyze, and learn from large pools of network data
Roadmap

- Dynamic network delay cartography
  - Kriged Kalman filter predictor
  - Optimal network sampling
  - Empirical validation: Internet2 and NZ-AMP data

- Unveiling network anomalies via sparsity and low rank

- Network-wide link count prediction

- RF cartography for cognition at the PHY

- Conclusions and future research directions
Why monitor delays?

Motivating reasons
- Assess network health
- Fault diagnosis
- Network planning

Application domains
- Old 8-second rule for WWW
- Content delivery networks
- Peer-to-peer networks
- Multiuser games
- Dynamic server selection
Research issues and goal

- Few tools are widely supported, e.g., traceroute, ping

- Additional tools from CAIDA\(^1\)
  - Require software installation at intermediate routers
  - Useless if intermediate routers not accessible

Desiderata: infer delays from a limited number of end-to-end measurements only!

\(^1\)Cooperative Association for Internet Data Analysis. [Online]. www.caida.org
Problem statement

- Consider a network graph with links, nodes, and paths

- Challenges
  - Overhead: \( \# \text{ paths} \approx O(\# \text{ nodes}^2) \)
  - Heavily congested routers may drop packets

- Q: Can fewer measurements suffice?
  - Most paths tend to share a lot of links [Chua’06]

- Inference task
  - Measure \( y_s \) on subset \( S \subset P \)
  - Predict \( y_{\bar{s}} \) on remaining paths \( \bar{S} : = \ P \setminus S \)
Network Kriging prediction

Given \( V_{ss} := \text{cov}(y_s) \), \( V_{\bar{s}s} := \text{cov}(y_{\bar{s}}, y_s) \), universal Kriging:

\[
\hat{y}_{\bar{s}} = V_{\bar{s}s} V_{ss}^{-1} y_s
\]

To obtain \( V_{ss}, V_{\bar{s}s} \) adopt a linear model for path delays

\[
y = Gx
\]

\[
[G]_{pl} = \begin{cases} 
1 & \text{link } l \in \text{path } p \\
0 & \text{otherwise}
\end{cases}
\]

\[
y = \begin{bmatrix} y_s \\ y_{\bar{s}} \end{bmatrix} = \begin{bmatrix} S \\ \bar{S} \end{bmatrix} Gx
\]

\[
\text{cov}(y) = \begin{bmatrix} V_{ss} & V_{s\bar{s}} \\ V_{\bar{s}s} & V_{\bar{s}\bar{s}} \end{bmatrix} = \begin{bmatrix} S \\ \bar{S} \end{bmatrix} G \Sigma G^T \begin{bmatrix} S^T \\ \bar{S}^T \end{bmatrix}
\]

Sampling matrix \( S \) known (selected via heuristic algorithms)

Spatio-temporal prediction

- Wavelet-based approach [Coates’07]
  - Diffusion wavelet matrix constructed using network topology
  - Can capture temporal correlations, but for $\tau$ time slots
  - High complexity ($O(\tau^3 P^3)$) $\Rightarrow$ cannot have $\tau > 10$

- Q: Should the same set of paths be measured per time slot?
  - Load balancing? Measurement on random paths?

- Prior art does not jointly offer
  - Spatio-temporal inference with online path selection, at low complexity

Simple delay model

- Delay measured on path $p \in \mathcal{P}$

$$y_p(t) = \chi_p(t) + \nu_p(t) + \epsilon_p(t)$$

Component due to traffic queuing:
- Random-walk with noise cov. $C_\eta$
  $$\chi(t) = \chi(t-1) + \eta(t)$$

Component due to processing, transmission, propagation:
- Traffic independent, temporally white, w/ cov. $C_\nu = \alpha G G^T$

Measurement noise i.i.d. over paths and time with known variance

$$\mathbb{E}[\epsilon_p(t)\epsilon_p^T(t)] = \sigma^2$$

Kriged Kalman Filter: Formulation

Path measured on subset $S \in \mathcal{P}$

$$y_s(t) = S(t)\chi(t) + \nu_s(t) + \epsilon_s(t)$$

$$\nu_s(t) := S(t)\nu(t)$$

KKF:

$$\chi(t) = \chi(t-1) + \eta(t)$$

$$y_s(t) = S(t)\chi(t) + \nu_s(t) + \epsilon_s(t)$$

Goal: Given history $\mathcal{H}(t) := \{y_s(\tau)\}_{\tau=1}^t$ find $\hat{y}_s(t)$
KKF updates

- State and covariance recursions
  \[
  \hat{\chi}(t) := \mathbb{E}^*[\chi(t)|\mathcal{H}(t)] \\
  = \hat{\chi}(t - 1) + K(t)(y_s(t) - S(t)\hat{\chi}(t - 1)) \\
  M(t) := \mathbb{E}[(\hat{\chi}(t) - \chi(t))(\hat{\chi}(t) - \chi(t))^T] \\
  = (I - K(t)S(t))(M(t - 1) + C_\eta)
  \]

- KKF gain
  \[
  K(t) := (M(t - 1) + C_\eta)S^T(t) \left( S(t)(M(t - 1) + C_\eta + C_\nu)S^T(t) + \sigma^2 I \right)^{-1}
  \]

- Kriging predictor
  \[
  \hat{y}_s(t) = \bar{S}(t)\hat{\chi}(t) + \bar{S}(t)C_\nu S^T(t) \left[ S(t)C_\nu S^T(t) + \sigma^2 I \right]^{-1} (y_s(t) - S(t)\hat{\chi}(t))
  \]
Which paths to measure?

- KKF can model and track network-wide delays
  - Practical sampling of paths? Optimal measurements? Criterion?

- Error covariance matrix
  
  \[
  M^Y_s(t) := \mathbb{E} \left\{ \left( y_s(t) - \hat{y}_s(t) \right) \left( y_s(t) - \hat{y}_s(t) \right)^T \right\} \\
  = \sigma^2 I_s + \sigma^2 \tilde{S}(t) \left[ \Phi^{-1} + S^T(t)S(t) \right]^{-1} \tilde{S}^T(t) \\
  \Phi = \left( M(t-1) + C_\eta + C_\nu \right) / \sigma^2
  \]

- Online experimental design: minimize \( \log \det(M^Y_s(t)) =: -f_t(S) \)

  \[
  S^*(t) := \arg \max_{S \subseteq \mathcal{P}} f_t(S) \\
  \text{subject to } |S| = S
  \]

- Log-det: D-optimal design (entropy of a Gaussian r. v.)
Greedy algorithm

**Algorithm**

Repeat $S$ times

$$S \leftarrow S \cup \arg \max_{p \notin S} \delta_S(p)$$

- Submodular + monotonic $\Rightarrow$ greedy solution $\left(1 - \frac{1}{e}\right)$ optimal [Nemhauser’78]

- Increments can be evaluated efficiently: $O(PS^3)$ with $P \gg S$
  - Operational complexity can be reduced further [Krause’11]

- Can be modified to handle cases when
  - Each node measures delay on all paths – which $S$ nodes to choose?
  - All nodes measure delay on only one path – which path to choose?

Empirical validation: Internet2

- Internet2 backbone
  - 72 paths
  - Lightly loaded

- One-way delay measurements using OWAMP

- Measurements every minute for 3 days in July 2011 ~ 4500 samples

- Training phase employed to estimate \( C, \eta, \alpha \)
  - Empirical estimates; see e.g., [Myers’76]
  - Techniques modified to handle measurements on subset of paths
  - First 1000 samples used for training; 50 random paths used for training

Data: http://internet2.edu/observatory/archive/data-collections.html
Network delay cartography (Internet2)
Normalized MSPE (Internet2)

Normalized MSPE \( := \frac{1}{T(P - S)} \sum_{t=1}^{T} ||\hat{y}_s(t) - y_s(t)||^2 \)
Empirical validation: NZ-AMP

- Delays measured on NZ-AMP, part of NLANR project
  - 186 paths, heavily loaded network

- Measurements every 10 minutes during August 2011 ~ 4500 samples

- Round-trip times measured using ICMP, paths via scamper

Data: http://erg.wand.net.nz
Normalized MSPE (NZ-AMP)

Random path selection

“Optimal” path selection
Scatter plots (NZ-AMP)

\[ \hat{y}_{\tilde{S}}(t) \text{ vs. } y_{\tilde{S}}(t) \forall t \]
\[ S = 30 \]
Takeaways

- Spatio-temporal inference useful for network health monitoring
- Dynamic network delay cartography via Kriged Kalman filtering
- Near-optimal path selection by utilizing submodularity
- Empirical validation on Internet2 and NZ-AMP datasets
Roadmap

- Dynamic network delay cartography

- Unveiling network anomalies via sparsity and low rank
  - Traffic modeling and identifiability
  - (De-) centralized and online algorithms
  - Numerical tests

- Network-wide link count prediction

- RF cartography for cognition at the PHY

- Conclusions and future research directions
Traffic anomalies

- Backbone of IP networks

- **Traffic anomalies**: changes in origin-destination (OD) flows
  - Failures, transient congestions, DoS attacks, intrusions, flooding

- **Motivation**: Anomalies $\Rightarrow$ congestion $\Rightarrow$ limits end-user QoS provisioning

**Objective**: Measuring superimposed OD flows per link, identify anomalies by leveraging **sparsity** of anomalies and **low-rank** of traffic.
Model

- Graph $G(N, L)$ with $N$ nodes, $L$ links, and $F$ flows ($F >> L$)

(as) Single-path per OD flow $z_{f,t}$

- Packet counts per link $l$ and time slot $t$

$\epsilon \{0, 1\}$

- Matrix model across $T$ time slots: $Y = R(Z + A) + V$

$L \times T$

$L \times F$
Low rank of traffic matrix

\[ Y = R(Z + A) + V \]

- **Z**: traffic matrix has **low rank**, e.g., [Lakhina et al‘04]

**Data:** http://math.bu.edu/people/kolaczyk/datasets.html
Sparsity of anomaly matrix

\[ Y = R(Z + A) + V \]

- **A**: anomaly matrix is **sparse** across both time and flows
Problem statement

Given $Y$ and routing matrix $R$, identify sparse $A$ when $Z$ is low rank.

R fat but $X$ still low rank

$$\{\hat{X}, \hat{A}\} = \arg \min_{\{X,A\}} \frac{1}{2} \|Y - X - RA\|_F^2 + \lambda_1 \|A\|_1 + \lambda_* \|X\|_*$$

(P1)

Low-rank $\Rightarrow$ sparse vector of SVs $\Rightarrow$ nuclear norm $\| \cdot \|_*$ and $\ell_1$ norm
Prior art

- Anomaly identification
  - Change detection on per-link time series [Brutlag’00], [Casas et al’10]
  - Spatial PCA [Lakhina et al’04]
  - Network anomography [Zhang et al’05]

- Rank minimization with the nuclear norm, e.g., [Recht-Fazel-Parrilo’10]
  - Matrix decomposition [Candes et al’10], [Chandrasekaran et al’11]

\[ M = L_0 + S_0 \]

Principal Component Pursuit

\[
\begin{align*}
\min_{L,S} & \quad \|L\|_* + \lambda \|S\|_1 \\
\text{s. to} & \quad M = L + S
\end{align*}
\] (PCP)
Challenges and importance

\[ Y = X + RA + V \]

- RA not necessarily sparse and R fat \(\Rightarrow\) PCP not applicable

- Important special cases
  - \( R = I \): matrix decomposition with PCP [Candes et al’10]
  - \( X = 0 \): compressive sampling with basis pursuit [Chen et al’01]
  - \( X = C_{Lx\rho}W'_{\rho xT} \) and \( A = 0 \): PCA [Pearson 1901]
  - \( X = 0, R = D \) unknown: dictionary learning [Olshausen’97]
# Exact recovery

- Noise-free case

\[ \mathbf{Y} = \mathbf{X}_0 + \mathbf{R} \mathbf{A}_0 = \mathbf{U} \Sigma \mathbf{V}' + \mathbf{R} \mathbf{A}_0 \]

\[ r = \text{rank} [\mathbf{X}_0], \quad s = \| \mathbf{A}_0 \|_0 \]

\[
\min_{\{\mathbf{X}, \mathbf{A}\}} \|\mathbf{X}\|_* + \lambda \|\mathbf{A}\|_1
\]

s.t. \[ \mathbf{Y} = \mathbf{X} + \mathbf{R} \mathbf{A} \] (P0)

**Q:** Can one recover sparse \( \mathbf{A}_0 \) and low-rank \( \mathbf{X}_0 \) exactly?

**A:** Yes! Under certain conditions on \( \{\mathbf{X}_0, \mathbf{A}_0, \mathbf{R}\} \)

**Theorem:** Given \( \mathbf{Y} \) and \( \mathbf{R} \), assume every row and column of \( \mathbf{A}_0 \) has at most \( k < s \) non-zero entries, and \( \mathbf{R} \) has full row rank. If \( \text{C1) - C2)} \) hold, then with \( \lambda \in (\lambda_{\min}, \lambda_{\max}) \) (P0) exactly recovers \( \{\mathbf{X}_0, \mathbf{A}_0\} \)

- \( \text{C1)} \)

\[ (1 - \mu(\Phi, \Omega_R))^2 (1 - \delta_k(\mathbf{R})) > \alpha \]

- \( \text{C2)} \)

\[ \lambda_{\min} := \beta \| \mathbf{R}' \mathbf{U} \mathbf{V}' \|_\infty < \lambda_{\max} := \sqrt{s^{-1}[\gamma^{-1} - \mu(\Phi, \Omega_R) \sqrt{r(1 + \delta_k(\mathbf{R}))}]} \]

Intuition

Exact recovery conditions satisfied if

- $r$ and $s$ are sufficiently small
- Nonzero entries of $A_0$ are “sufficiently spread out”
- Incoherent rank and sparsity-preserving subspaces
- $R$ satisfies a restricted isometry property

Remarks

- Amplitude of non-zero entries of $A_0$ irrelevant
- Conditions satisfied for certain random ensembles w.h.p.
Numerical validation

- **Setup**
  
  \[ L=105, \; F=210, \; T = 420 \]
  
  \[ R \sim \text{Bernoulli}(1/2) \]
  
  \[ X_0 = RPQ', \; P, Q \sim \mathcal{N}(0, 1/FT) \]
  
  \[ a_{ij} \in \{-1,0,1\} \text{ w.p. } \{\pi/2, \; 1-\pi, \; \pi/2\} \]

- **Relative recovery error**

  \[ \epsilon = \frac{\| \hat{A} - A_0 \|_F}{\| A_0 \|_F} \]
In-network processing

- Spatially-distributed link count data

**Centralized:** $\sum \text{Local link counts per agent} + \text{nominal data matrix}$

**Decentralized:** $\sum \text{Local link counts per agent} + \text{nominal data matrix}$

- Local processing and single-hop communications

**Goal:** Given local link counts per agent, unveil anomalies in a distributed fashion by leveraging low-rank of the nominal data matrix and sparsity of the outliers.

**Challenge:** $\| \cdot \|_{*}$ not separable across rows (links/agents)
Separable regularization

Key property

\[ \|X\|_* := \min_{\{C, W\}} \frac{1}{2} \left\{ \|C\|_F^2 + \|W\|_F^2 \right\}, \text{ s.t. } X = CW' \]

Separable formulation equivalent to (P1)

\[ \min_{\{C, W, A\}} \frac{1}{2} \|Y - CW' - RA\|_F^2 + \lambda_1 \|A\|_1 + \frac{\lambda_*}{2} \left\{ \|C\|_F^2 + \|W\|_F^2 \right\} \quad \text{(P2)} \]

- Nonconvex; less variables: \( LT \Rightarrow \rho(L + T) \)

**Proposition 3:** If \( \{\tilde{C}, \tilde{W}, \tilde{A}\} \) stat. pt. of (P2) and \( \|Y - \tilde{C}\tilde{W}' - R\tilde{A}\| \leq \lambda_* \), then \( \{\tilde{X} := \tilde{C}\tilde{W}', \tilde{A} := \tilde{A}\} \) is a global optimum of (P1).
Distributed algorithm

- Alternating-direction method of multipliers (ADMM) solver for (P2)
  - Method [Glowinski-Marrocco’75], [Gabay-Mercier’76]
  - Learning over networks [Schizas-Ribeiro-Giannakis’07]

Consensus-based optimization

Attains centralized performance

Benchmark: PCA-based methods

- **Idea:** anomalies increase considerably \( \text{rank}(\mathbf{Y}) \)

**Algorithm**

i) Form subspace \( \mathcal{N} \) via \( r \)-dominant left singular vectors of \( \mathbf{Y} \) (resp. \( \mathcal{N}^c \))

ii) Infer anomalies from \( \mathcal{P}_{\mathcal{N}^c}(\mathbf{Y}) \)

- Assumes knowledge of \( r := \text{rank}(\mathbf{X}) \)

- **[Lakhina et al’04]** For \( t = 1, \ldots, T \) \( \| \mathcal{P}_{\mathcal{N}^c}(\mathbf{y}_t) \|_2 \geq_{H_0}^{H_1} \tau \)

- **[Zhang et al’05]** Sparse anomalies \( \hat{\mathbf{A}} = \arg \min_{\mathcal{P}_{\mathcal{N}^c}(\mathbf{Y})=\mathbf{R}\mathbf{A}} \| \mathbf{A} \|_1 \)
Synthetic data

- Random network topology
  - $N=20$, $L=108$, $F=360$, $T=760$
  - Minimum hop-count routing

Detection probability

- PCA-based method, $r=5$
- PCA-based method, $r=7$
- PCA-based method, $r=9$
- Proposed method, per time and flow

False alarm probability

$P_f = 10^{-4}$
$P_d = 0.97$
Internet2 data

- Real network data
  - Dec. 8-28, 2008
  - N=11, L=41, F=121, T=504

Detection probability vs. False alarm probability

- [Lakhina04], rank=1
- [Lakhina04], rank=2
- [Lakhina04], rank=3
- Proposed method
- [Zhang05], rank=1
- [Zhang05], rank=2
- [Zhang05], rank=3

Data: http://www.cs.bu.edu/~crovella/links.html
Dynamic anomalography

- Construct an estimated map of anomalies in real time

- Streaming data model:

  \[ \mathcal{P}_{S_t}(y_t) = \mathcal{P}_{S_t}(x_t + R_t a_t + v_t), \ t = 1, 2, \ldots \]
  \[ x_t := R_t z_t \]

**Goal:** Given \( \{ \mathcal{P}_{S_i}(y_i), R_i \}_{i=1}^t \) estimate \( (x_t, a_t) \) online when \( \{ x_t \} \) is in a low-dimensional space and \( \{ a_t \} \) is sparse

- (Robust) subspace tracking

  - Projection approximation (PAST) [Yang’95]
  - Missing data: GROUSE [Balzano et al’10], PETRELS [Chi et al’12]
  - Outliers: [Mateos-Giannakis’10], GRASTA [He et al’11]

- Compressed “outliers” challenge identifiability

Online estimator

- **Challenge:** $\| \cdot \|_*$ not separable across columns (time) $\Rightarrow x_t = Cw_t$

- **Approach:** regularized exponentially-weighted LS formulation

$$\min_{\{C,W,A\}} \sum_{t=1}^{T} \beta^{t-\tau} \left[ \frac{1}{2} \| \mathcal{P}_{S_t} (y_t - Cw_t - R_t a_t) \|^2_F + \frac{\lambda_*}{2} \| C \|^2_F + \frac{\lambda_*}{2} \| w_t \|^2_F + \lambda_1 \| a_t \|_1 \right]$$
Delay cartography

- Network distance prediction [Liau et al’12]
- **Approach**: distributed low-rank matrix completion
- Internet2 data (Aug 18-22,2011)
  - End-to-end latency matrix
  - $N=9$, $L=T=N$; 20% missing data

Takeaways

- Unveiling network traffic anomalies via convex optimization
  - Leveraging sparsity and low rank

- Reveal when and where anomalies occur

- Exact recovery of low-rank plus compressed sparse matrices

- Distributed/online algorithms with guaranteed performance
Roadmap

- Dynamic network delay cartography

- Unveiling network anomalies via sparsity and low rank

- Network-wide link count prediction
  - Semi-supervised learning for traffic maps
  - Batch and online processing
  - Empirical validation: Internet2 data

- RF cartography for cognition at the PHY

- Conclusions and future research directions
A commuting conundrum

- **Objective:** map a “good” route for packet delivery
  - Measure traffic at few roads/links only

- **Application domains**
  - Transportation networks [Gastner-Newman’04]
  - Communication networks [Soule et al’05]
  - Sensor networks [Abrams et al’04]
Model

- Graph $G(N, L)$ with $N$ nodes, $L$ links, and $F$ flows ($F \gg L$)

(as) Single-path per OD flow $z_{f,t}$

- Packet counts per link $l$ and time slot $t$

$$y_{l,t} = \sum_{f=1}^{F} r_{l,f} x_{f,t} + v_{l,t}$$

$\in (0, 1)$

- Incomplete, noisy measurements on a subset of links $l \in S_t$

$$\mathcal{P}_{S_t}(y_t) = \mathcal{P}_{S_t}(Rx_t + v_t)$$

$$[\mathcal{P}_{S_t}(y_t)]_l = \begin{cases} y_{l,t}, & l \in S_t \\ 0, & l \not\in S_t \end{cases}$$
**Problem statement**

**Goal:** Given $P_{S_t}(y_{t'})$, $R$ and historical data $\{P_{S_t}(y_t)\}_{t=1}^{T}$, find $\hat{y}_t, t' > T$

**Prior art**
- Traffic estimation $\hat{y}_{t'} = R \hat{x}_{t'}$ [Zhang et al’05]
- Kriging [Chua et al’06], plus traffic modeling [Vaughn et al’10]
- Topology-driven basis expansion [Crovella-Kolaczyk’03], [Coates et al’07]

**Impact**
- Ability to handle missing data
- Online prediction capturing spatio-temporal correlations
- Computationally-efficient link traffic prediction

Data-driven model of link counts

- **Sparse** representation of link counts: $y_t = Ds_t$
  \[ L \times Q, \ (L \leq Q) \]

- **Notation:**
  \[ D := \{D = [d_1, \ldots, d_Q] \in \mathbb{R}^{L \times Q} : \|d_q\| \leq 1, \ \forall q\} \]
  \[ S := [s_1, \ldots, s_T] \]

**Dictionary Learning (DL) [Olshausen-Field’97]**

Given $Y = [y_1, \ldots, y_T]$, find dictionary (basis) $D$ and sparse $S$

\[
(\hat{S}, \hat{D}) = \arg \min_{S,D \in \mathcal{D}} \|Y - DS\|_F^2 + \lambda \|S\|_1
\]

- **Q:** How about DL from incomplete data $\{P_{\mathcal{S}_t}(y_t)\}_{t=1}^T$?
Capturing spatial link dependence

- Auxiliary graph $G$ with vertices = links in $G$
  - Edge weights $w_{l,l'} = \text{number of OD flows common to links } l, l'$
  - Adjacency matrix: $W = RR'$, graph Laplacian $L = \text{diag}(W1_L) - W$

- Cost function to learn $D$

$$C_t(D, s_t) := \|P_s(y_t - Ds_t)\|^2_2 + \lambda_1\|s_t\|_1 + \lambda_2 s'_t D'LDs_t$$

- Regularizers affect sparsity and smoothness over $G$

$$s'_t D'LDs_t = \frac{1}{2} \sum_{l=1}^{L} \sum_{l'=1}^{L} w_{l,l'} (x_{l,t} - x_{l',t})^2$$
**Semi-supervised DL**

**Semi-supervised Dictionary Learning (SSDL)**

Given $\{P_{S_t}(y_t)\}_{t=1}^T$, find dictionary (basis) $D$ and sparse $S$

$$\hat{(S, D)} = \arg \min_{S, D \in D} \sum_{t=1}^T C_t(D, s_t)$$

- SSDL biconvex, block-coordinate descent (BCD) solver
  - Update $\{s_t\}_{t=1}^T$ via parallel entry-wise soft-thresholding
  - Update each $D$ via QP + projection onto the Euclidean ball

**Proposition:** BCD’s iterates converge to a stationary point of SSDL
Link load prediction

- Given $\mathcal{P}_{S_{t'}}(y_{t'})$ and learnt dictionary $\hat{\mathbf{D}}$, solve

$$\hat{s}_{t'} := \arg \min_{s} \| \mathcal{P}_{S_{t'}}(y_{t'} - \hat{\mathbf{D}}s) \|_2^2 + \lambda_1 \|s\|_1 + \lambda_2 s'\hat{\mathbf{D}}'\mathbf{L}\hat{\mathbf{D}}s$$

  - Captures sparsity of $s_{t'}$ and smoothness of link loads over $\mathcal{G}$

- Predict $y_{t'}$ based on $\hat{s}_{t'}$

$$\hat{y}_{t'} = \hat{\mathbf{D}}\hat{s}_{t'}$$

  - Scaling factor $(1 + \lambda_2)$ reduces bias in $\hat{y}_{t'}$ [Zou-Hastie’05]
Batch processing summary

**TRAINING PHASE**

\[ \min_{s_t} C_t(D[k], s_t) \]

\[ \{PS_t(\overline{y}_t)\}_{t=1}^T \]

\[ S_t[k] \]

\[ \min_{s_{t'}} T \sum_{t=1}^T C_t(D, s_{t'}[k]) \]

\[ D[k + 1] \]

\[ \hat{D} \]

\[ \min_{s_{t'}} C_{t'}(\hat{D}, s_{t'}) \]

\[ \hat{s}_{t'} \]

\[ \hat{x}_{t'} = \hat{D}\hat{s}_{t'} \]

\[ \hat{x}_{t'} \]
Test case: Internet2

- Internet2 measurement archive
  - $L=54$, $T=2000$

  Training phase – 30 links measured
  Operational phase – 30 links measured

- Prediction improves as link load increases
Prediction error (Internet2)

- Normalized prediction error:  \[ \text{NPE} := \frac{1}{Lt_0} \sum_{\tau=1}^{t_0} \| y_{\tau} - \hat{y}_{\tau} \|_2^2 \]
  
  - \( Q \) = number of columns of \( D \); \( t_0 = 2000 \)

- Gravity-based [Zhang et al’05]; Diffusion wavelets [Coifman-Maggioni’07]

- SSDL outperforms competing alternatives
Online processing

- Capture temporal correlations on \( \{s_\tau\} \)

\[
C^t_\beta(D_t, s) := \sum_{\tau=1}^{t} \beta^{t-\tau} \|P_{S_\tau}(y_\tau - D_t s)\|_2^2 + \lambda_1 \|s\|_1 + \lambda_2 s' D'_t LD_t s
\]

- Given \( P_{S_t}(y_t) \) and dictionary \( D_t \), solve \( s_t := \arg \min_s C^t_\beta(D_t, s) \)

- Predict \( y_t \) based on \( s_t \)

\[
\hat{y}_t = (1 + \lambda_2) D_t s_t
\]

- Dictionary update

\[
D_{t+1} = \arg \min_{D \in D} \frac{1}{t} \sum_{\tau=1}^{t} C^{\tau}_\beta(D, s_\tau)
\]
Real-time prediction (Internet2)

- $Q=60$, different values of the forgetting factor $\beta$
  - Measure traffic at 30 links only

- SSDL-based tracker outperforms diffusion wavelets
Takeaways

- Prediction of network processes from incomplete observations
  - Link count prediction based on dictionary learning

- Spatial correlation of link counts via Laplacian regularization
  - Semi-supervised learning

- Online algorithms capturing temporal correlations
Roadmap

- Dynamic network delay cartography
- Unveiling network anomalies via sparsity and low rank
- Network-wide link count prediction
- RF cartography for cognition at the PHY
  - Interference spectrum cartography
  - Channel gain cartography
- Conclusions and future research directions
What is a cognitive radio?

- Fixed radio
  - *Policy-based*: parameters set by operators

- Software-defined radio (SDR)
  - *Programmable*: can adjust parameters to intended link

- Cognitive radio (CR)
  - *Intelligent*: sense the environment & learn to adapt [Mitola’00]

- **Cognizant transceiver**: sensing
- **Agile transmitter**: adaptation
- **Intelligent DRA**: decision making
  - Radio reconfiguration decisions
  - Spectrum access decisions
Spectrum scarcity problem

- Fixed spectrum access policies
  - Useful radio spectrum pre-assigned

Inefficient occupancy

US FCC
Dynamical access under user hierarchy

- Primary users (PUs) versus secondary users (SUs/CRs)

**Spectrum underlay**
- Restriction on transmit power levels
- Operation over ultra wide bandwidths

**Spectrum overlay**
- Constraints on when and where to transmit
- Avoid interference to Pus via sensing and adaptive allocation
Cooperative sensing for efficient sharing

- Multiple CRs jointly detect the spectrum [Ganesan-Li’06][Ghasemi-Sousa’07]

- **Benefits of cooperation**
  - Spatial diversity gain mitigates multipath fading/shadowing
  - Reduced sensing time and local processing
  - Ability to cope with hidden terminal problem

- **Limitation**: existing approaches do not exploit space-time dimensions

Source: Office of Communications (UK)
Cooperative PSD cartography

- **Idea:** CRs collaborate to form a spatial map of the RF spectrum

**Goal:** Find PSD map $\Phi(x, f)$ across space $x \in \mathbb{R}^2$ and frequency $f \in \mathbb{R}$

- **Specifications:** coarse approx. suffices
- **Approach:** basis expansion of $\Phi(x, f)$

Modeling

- Transmitters
  \( T_{x_s}, \ s = 1, \ldots, N_s \)

- Sensing CRs
  \( C_{R_r}, \ r = 1 : N_r \)

- Frequency bases
  \( b_{\nu}(f), \ \nu = 1 : N_b \)

- Sensed frequencies
  \( f_k, \ k = 1 : K \)

\[ \Phi_s(f) = \sum_{\nu=1}^{N_b} \theta_{s\nu} b_{\nu}(f) \]

- Sparsity present in space and frequency
Space-frequency basis expansion

- Find $\theta_{s\nu} \Rightarrow$ Tx-power of source $s$ over frequency band $\nu$
- Data $\phi_{r\nu} \Rightarrow$ Rx-power at cognitive radio $CR_r$
- Estimate sparse $\theta$ to find PSD at $CR_r$

\[ \phi = \Gamma \theta + e \]

\[ \hat{\theta} = \arg \min_{\theta} \| \phi - \Gamma \theta \|_2^2 + \lambda \| \theta \|_1 \]  

Sparsity-promoting regularization
Distributed recursive implementation

- **Consensus-based** approach
  - Solve locally
    
    $$\hat{\theta} = \arg \min_{\theta_r} ||\phi_r - \Gamma_r \theta_r||_2^2 + \frac{\lambda}{M} ||\theta_r||_1$$
    
    s.to
    $$\theta_r = \theta_{r'}, \quad \forall r' \in \mathcal{N}_r$$

- Constrained optimization using ADMM

- Exchange of local $\theta_r$ estimates

- Scalability
- Robustness
- Lack of infrastructure

- Centralized
  - Fusion center

- Decentralized
  - Ad-hoc
RF spectrum cartography

- 5 sources
- $N_s = 121$ candidate locations, $N_r = 50$ CRs

As a byproduct, Lasso localizes all sources via variable selection
Simulated test: PSD map estimation

- Centralized sensing
- No fading
- $N_s = 25$

2 CR Tx

$\Phi_i(f)$

$\Phi_m(f)$

4 CR Rx

Rx signal 1

Rx signal 2

Rx signal 3

Rx signal 4
Distributed consensus with fading

- Starting from a local estimate, sensors reach **consensus**

- "True" Tx spectrum

- Sensed at the $t^{th}$ consensus step

$\Phi_i(f)$
Spline-based PSD cartography

Q: How about shadowing?

Path-loss

Shadowing

A: Basis expansion with coefficient functions

\[ \Phi(x, f) = \sum_{\nu=1}^{N_b} g_\nu(x) b_\nu(f) \]

\(g_\nu(x)\): unknown dependence on spatial variable \(x\)

Frequency basis expansion

- PSD of Tx source $s \in \{1, \ldots, N_s\}$ is $\Phi_s(f)$

$$\Phi_s(f) = \sum_{\nu=1}^{N_b} \theta_{s\nu} b_{\nu}(f)$$

Basis functions
- Accommodate prior knowledge $\Rightarrow$ raised-cosine
- Sharp transitions (regulatory masks) $\Rightarrow$ rectangular, non-overlapping
- Overcomplete basis set (large $N_b$) $\Rightarrow$ robustness
Spatial PSD model

- Spatial loss function $l_s(x): \mathbb{R}^2 \rightarrow \mathbb{R} \Rightarrow \text{Unknown}$

- BEM:
  $$\Phi(x, f) = \sum_{\nu=1}^{N_b} g_\nu(x) b_\nu(f)$$
  $$\left(g_\nu(x) := \sum_{s=1}^{N_s} \theta_{sv} l_s(x)\right)$$

  ➢ Per sub-band factorization in space and frequency (indep. of $N_s$)

- Goal: estimate PSD atlas as $\hat{\Phi}(x, f) = \sum_{\nu=1}^{N_b} \hat{g}_\nu(x) b_\nu(f)$
Nonparametric basis pursuit

Available data:
\( \mathcal{X} := \{x_r\}_{r=1}^{N_r} \) location of CRs
\( \mathcal{F} := \{f_n\}_{n=1}^{N} \) measured frequencies

Observations \( \{\varphi_{rn}\} \)

\( \tilde{g}_1(x) \)

\( \tilde{g}_2(x) \)

\[ \{\tilde{g}_\nu\}_{\nu=1}^{N_b} = \arg \min_{\{g_\nu \in \mathcal{S}\}} \sum_{r=1}^{N_r} \sum_{n=1}^{N} \left( \varphi_{rn} - \sum_{\nu=1}^{N_b} g_\nu(x_r)b_\nu(f_n) \right)^2 \]

\[ + \lambda \sum_{\nu=1}^{N_b} \int_{\mathbb{R}^2} \| \nabla^2 g_\nu(x) \|_F^2 dx + \mu \sum_{\nu=1}^{N_b} \| [g_\nu(x_1), \ldots, g_\nu(x_{N_r})]' \|_2 \] (P1)

- Avoid overfitting by promoting smoothness
- Nonparametric basis selection \( (\tilde{g}_\nu \neq 0 \Rightarrow b_\nu(f) \text{ selected}) \)
Thin-plate splines solution

Proposition 1: Estimates \( \{ \hat{g}_\nu \}_{\nu=1}^{N_b} (P1) \) are thin-plate splines [Duchon’ 77]

\[
\hat{g}_\nu(x) = \sum_{r=1}^{N_r} \hat{\beta}_{\nu r} K(||x - x_r||) + \hat{\alpha}'_{\nu 1} x + \hat{\alpha}_{\nu 0}
\]

where \( K(\rho) \) is the radial basis function \( K(\rho) = \rho^2 \log(\rho) \), and

\[
\hat{\beta}_\nu := [\hat{\beta}_{\nu 1}, \ldots, \hat{\beta}_{\nu N_r}] \in B := \left\{ \beta : \sum_r \beta_r = 0, \sum_r \beta_r x_r = 0, x_r \in \mathcal{X} \right\} .
\]

- Unique, closed-form, finitely-parameterized minimizers!
- **Q1**: How to estimate \( \{ \alpha_\nu, \beta_\nu \}_{\nu=1}^{N_b} \) based on \( \varphi \)?
- **Q2**: How does (P1) perform basis selection?
Lassoing bases

- (P1) equivalent to group Lasso estimator [Yuan-Lin’ 06]
  - Matrices (X and F dependent)
  i) $T := \begin{bmatrix} 1 & x'_1 \\ \vdots & \vdots \\ 1 & x'_{N_r} \end{bmatrix} = [Q_1 \; Q_2][R' \; 0]'$, ii) $[K]_{rl} := K(||x_r - x_l||)$, iii) $[B]_{n\nu} := b_{\nu}(f_n)$

**Proposition 2:** Minimizers $\{\hat{\alpha}_\nu, \hat{\beta}_\nu\}_{\nu=1}^{N_b}$ (P1) are fully determined by

$$\hat{\zeta} := \arg \min_{\zeta} ||y - X\zeta||_2^2 + \mu \sum_{\nu=1}^{N_b} ||\zeta_{\nu}||_2$$

with $y := [\varphi \; 0]'$, $X := \begin{bmatrix} B \otimes I \\ I \otimes F(\lambda, T, K, B) \end{bmatrix}$

as $[\hat{\beta}'_{\nu}, \hat{\alpha}'_{\nu}]' = \text{bdiag}(Q_2, I)[KQ_2 \; T]^{-1}\hat{\zeta}_{\nu}$.

- Group Lasso encourages sparse factors $\hat{\zeta}_{\nu}$
  - Full-rank mapping: $\hat{\zeta}_{\nu} = 0 \Rightarrow \hat{g}_{\nu}(x) \equiv 0$
Simulated test

- $N_s = 2$ sources; raised cosine pulses
- $N_r = 50$ sensing CRs, $N = 64$ sampling frequencies
- $N_b = (2 \times 15 \times 2) = 60$; (roll off x center frequency x bandwidth)
Numerical test IEEE 802.11

- $N_b = 14$
- PUs $N_s = 2$

- $N_r = 100$ CRs

Maps estimated under fading + shadowing + overlapping bases
Real RF data

- IEEE 802.11 WLAN activity sensed

$N_r = 166$ CRs

- Frequency bases identified
- Maps recovered and extrapolated
Semi-supervised DL for PSD maps

- Signal model \( \pi_t = G_t p_t \Rightarrow \pi = Ds \)

  - Rx-power measured by a few CRs \( \mathcal{P}_{S_t}(y_t) = \mathcal{P}_{S_t}(\pi_t + v_t) \)

- Batch formulation

  - Online algorithm via exponentially weighted criterion

\[ \min_{s_t} C_t(D[k], s_t) \]

\[ \min_{\|d_k\| \leq 1} \sum_{t=1}^{T} C_t(D[s_t[k]]) \]

\[ \min_{s_{t'}} C_{t'}(\hat{D}, s_{t'}) \]

\[ \hat{x}_{t'} = \hat{D} \hat{s}_{t'} \]

Numerical tests
Recap: PHY sensing via RF cartography

- **Power spectral density (PSD) maps**
  - Capture ambient power in space-time-frequency
  - Can identify “crowded” regions to be avoided

- **Channel gain (CG) maps**
  - Time-frequency channel from any-to-any point
  - CRs adjust Tx power to min. PU disruption

---

Channel gain cartography

- CG after averaging small-scale fading (dB)
  \[ G_{x\rightarrow y}(t) := \Gamma_{x\rightarrow y}(t) + s_{x\rightarrow y}(t) \]

- State-space model for shadowing
  \[ s_{x\rightarrow y}(t) := \bar{s}_{x\rightarrow y}(t) + \tilde{s}_{x\rightarrow y}(t) \]

Approach: spatial LMMSE interpolation (Kriging) + KF for tracking channel dynamics

Payoffs: tracking PU activities; accurate interference models; efficient resource allocation

Outlook: jointly optimal PHY CR sensing and access

---

Any-to-any CG estimation

- Shadowing model-free approach
  - Slow variations in shadow fading
  - Low-rank any-to-any CG matrix $\hat{G}$

**Approach:** low-rank matrix completion

$$\min_{C,W} \left\| P_S(G - CW') \right\|_F^2 + \lambda (\|C\|_F^2 + \|W\|_F^2)$$

**Payoffs:** global view of any-to-any CGs; real-time propagation metrics; efficient resource allocation

**Outlook:** kernel-based extrapolator for missing CR-to-PU measurements, or future time intervals

PU power and CR-PU link learning

- Reduce overhead in any-to-any CG mapping
  - Learn CGs only between CRs and PUs
  - Online detection of active PU transmitters

**Approach:** blind dictionary learning

\[
\min_{G,P} \| \Pi - GP \|_F^2 + \lambda_1 \| P \|_1
\]

**Payoffs:** tracking PU activities;
  efficient resource allocation

**Outlook:** missing data due to limited sensing;
  distributed and robust algorithms

---

Takeaways

- PHY layer spatiotemporal sensing via RF cartography
  - Space-time-frequency view of interference and channel gains

- Identify idle bands across space
  - Aid dynamic spectrum access policies

- PU/source localization and tracking

- Parsimony via sparsity and distribution via consensus
  - Lasso, group Lasso on splines, and method of multipliers
Roadmap

- Dynamic network delay cartography
- Unveiling network anomalies via sparsity and low rank
- Network-wide link count prediction
- RF cartography for cognition at the PHY
- Conclusions and future research directions
The big picture ahead...

- **Network cartography**: succinct depiction of the network state
- **Vision**: use *atlas* to enable spatial re-use, hand-off, localization, Tx-power tracking, resource allocation, health monitoring, and routing
Concluding summary

- Dynamic network cartography
  - Framework to construct maps of the dynamic network state
  - Real-time, distributed scalable algorithms for large-scale networks

- Global state mapping from incomplete and corrupted data
  - Path delay and link traffic maps
  - Prompt and accurate identification of traffic anomalies
  - PHY layer sensing in wireless CR networks via RF cartography

- Statistical SP toolbox
  - Sparsity-cognizant learning, low-rank modeling
  - Kriged Kalman filtering of dynamical processes over networks
  - Semi-supervised dictionary learning
  - Distributed optimization via the ADMM

Thank you!
Questions?

University of Minnesota
http://spincom.umn.edu

Dr. J. A. Bazerque
UofM

Dr. E. Dall’Anese
UofM

Dr. P. A. Forero
SPAWAR

Dr. S. J. Kim
UofM

M. Mardani
UofM

Prof. K. Rajawat
IIT Kanpur