Bayesian Continues-Time Tucker Decomposition

Shikai Fang, Akil Narayan, Robert M. Kirby, Shandian Zhe

The University of Utah

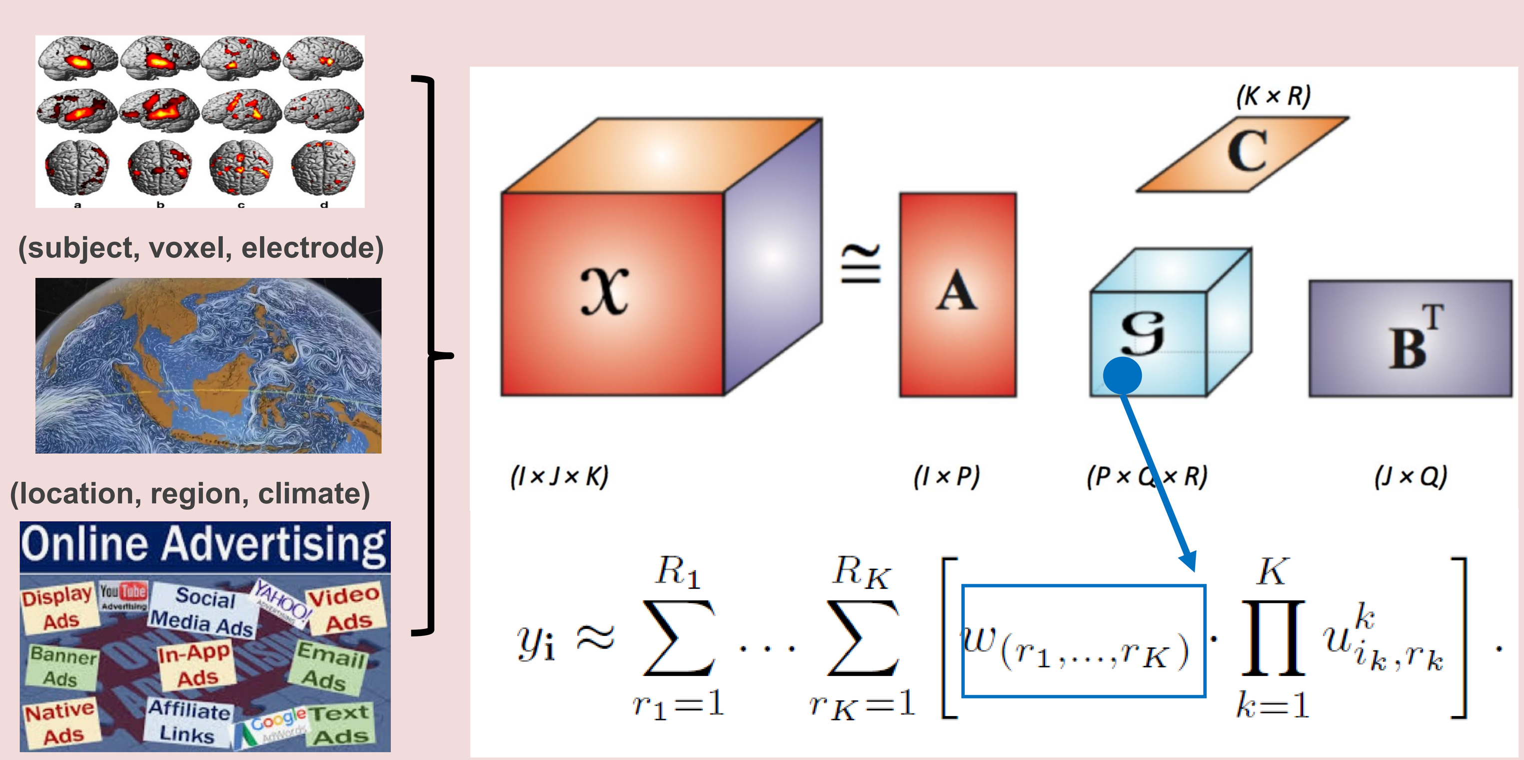
Contact: shikai.fang@utah.edu

ABSTRACT

Tensor decomposition is a dominant framework for multiway data analysis and prediction. Although practical data often contains timestamps for the observed entries, existing tensor decomposition approaches overlook or under-use this valuable time information. They either drop the timestamps or bin them into crude steps and hence ignore the temporal dynamics within each step or use simple parametric time coefficients. To overcome these limitations, we propose **Bayesian Continuous-Time Tucker Decomposition**. We model the **tensor-core of the classical Tucker decomposition as a time-varying function**, and place a **Gaussian process prior** to flexibly estimate all kinds of temporal dynamics. In this way, our model maintains the interpretability while is flexible enough to capture various complex temporal relationships between the tensor nodes. For efficient and high-quality posterior inference, we use the **stochastic differential equation (SDE) representation of temporal GPs** to build an equivalent **state-space prior**, which avoids huge kernel matrix computation and sparse/low-rank approximations. We then **use Kalman filtering**, **RTS smoothing**, and **conditional moment matching** to develop a scalable message passing inference algorithm. We show the advantage of our method in simulation and several real-world applications.

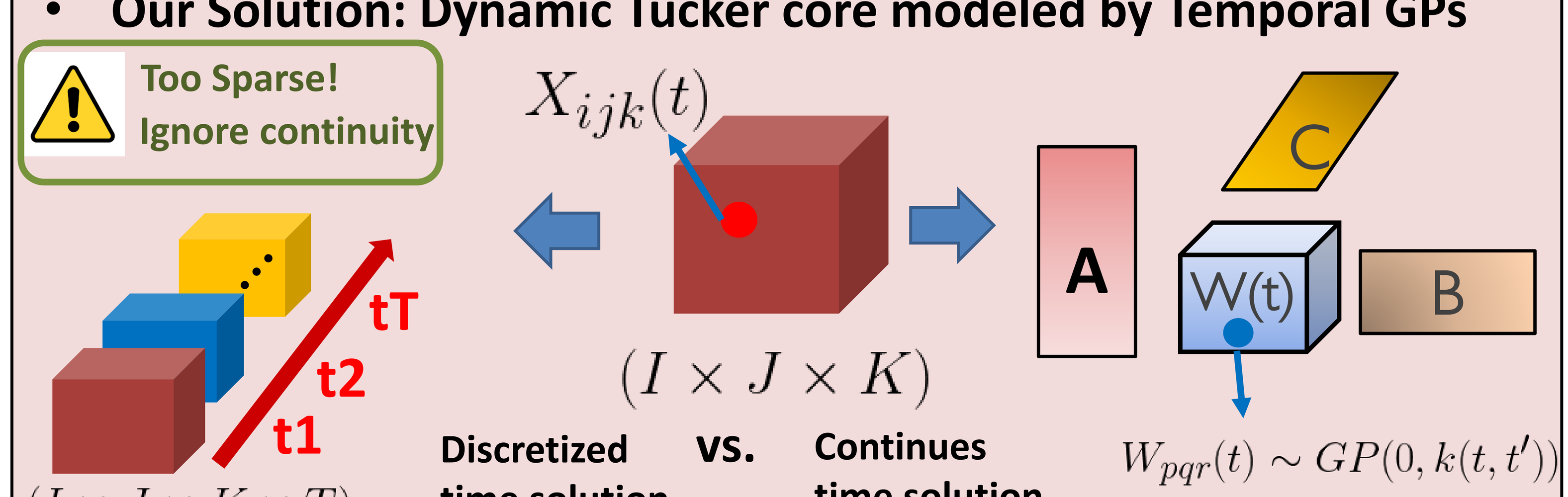
INTRODUCTION

- Background: **Tensor data and Tucker Decomposition**
- Goal: estimate **latent factors** to reconstruct tensor



$$y_i \approx \sum_{r_1=1}^{R_1} \cdots \sum_{r_K=1}^{R_K} \left[w_{(r_1,\ldots,r_K)} \cdot \prod_{k=1}^{K} u^k_{i_k, r_k} \right].$$

- **Challenge:** How to model **Temporal info** in tensor
- **Our Solution: Dynamic Tucker core modeled by Temporal GPs**



$X_{ijk}(t)$

$W_{pqr}(t) \sim GP(0, k(t, t'))$

METHODS

- **Joint Prob model**:

$$p(\mathcal{U}, \{\mathbf{w_r}\}_{\mathbf{r}}, \tau, \mathbf{y}) = \mathrm{Gam}(\tau \mid b_0, c_0) \prod_{k=1}^{K} \prod_{j=1}^{d_k} \mathcal{N}\left(\mathbf{u}_j^k \mid \mathbf{0}, \mathbf{I}\right) \times \prod_{\mathbf{r}=(1,\ldots,1)}^{(R_1,\ldots,R_K)} \mathcal{N}(\mathbf{w_r} \mid \mathbf{0}, \mathbf{K_r}) \times \prod_{n=1}^{N} \mathcal{N}\left(y_n \mid \mathrm{vec}(\mathcal{W}(t_n))^\top \left(\mathbf{u}^1_{i_{n_1}} \otimes \ldots \otimes \mathbf{u}^K_{i_{n_K}}\right), \tau^{-1}\right)$$

Temporal GP prior on Tucker Core

Priors of factors and noise

Gaussian Likelihood

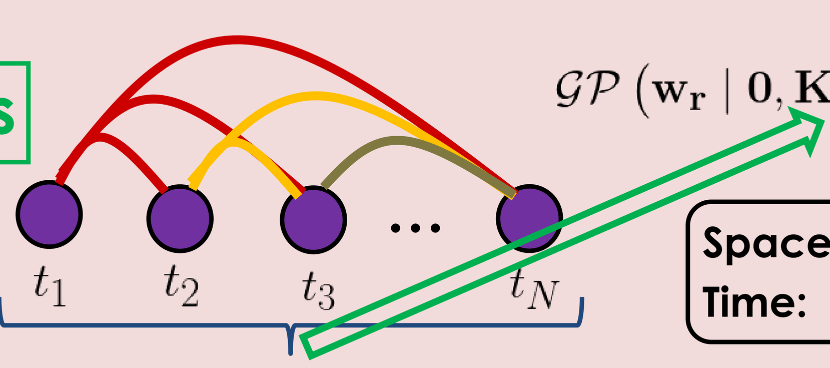
O(N^3) cost !!

- To **avoid crude low rank approx.**,

We apply the fact:

Temporal GPs

$\mathcal{GP}(\mathbf{w_r} \mid \mathbf{0}, \mathbf{K_r}(t, t'))$



stationary kernel (e.g. Matern25)

Space: $\mathcal{O}(N^2)$
Time: $\mathcal{O}(N^3)$

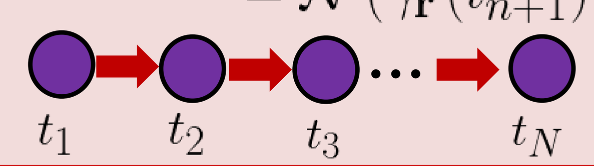
Linear Time-Invariant(LTI) SDE

$$\frac{\mathrm{d}\gamma_{\mathbf{r}}(t)}{\mathrm{d}t} = \mathbf{F}\gamma_{\mathbf{r}} + \mathbf{L}\xi(t)$$

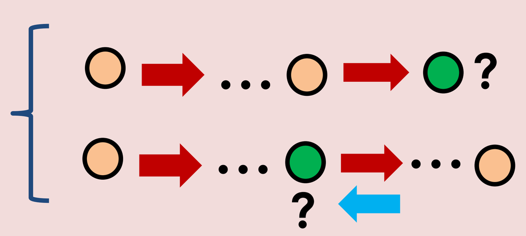
Solve Riccati Equation

discrete form

$$p\left(\gamma_{\mathbf{r}}(t_{n+1}) \mid \gamma_{\mathbf{r}}(t_n)\right) = \mathcal{N}\left(\gamma_{\mathbf{r}}(t_{n+1}) \mid \mathbf{A}_n \gamma_{\mathbf{r}}(t_n), \mathbf{Q}_n\right)$$



State Space Model (Gauss Markov Chain)



Kalman Filter

RTS Smoothing

Space: $\mathcal{O}(N)$
Time: $\mathcal{O}(N)$

- **Gaussian-Gamma Approx. on data likelihood** :

$$\mathcal{N}\left(y_n \mid (\mathbf{H}\bar{\gamma}_n)^\top \left(\mathbf{u}^1_{i_{n_1}} \otimes \ldots \otimes \mathbf{u}^K_{i_{nK}}\right), \tau^{-1}\right) \approx Z_n \prod_{k=1}^{K} \mathcal{N}\left(\mathbf{u}^k_{i_{n_k}} \mid \mathbf{m}^{k,n}_{i_{n_k}}, \mathbf{V}^{k,n}_{i_{n_k}}\right) \cdot \mathrm{Gam}(\tau \mid b_n, c_n) \times \mathcal{N}(\mathbf{H}\bar{\gamma}_n \mid \boldsymbol{\beta}_n, \mathbf{S}_n)$$

Approx. Msg of Factors & noise -> solve by *moment match*

Approx. Msg of SDE states /Tucker core -> solve by KF/RTS

- To enable ***tractable moment match***, we apply:

1. Conditional Expectation Propagation(CEP)

$$\mathbb{E}_{\bar{p}}[\phi(\mathcal{W})] = \mathbb{E}_{\bar{p}(\Theta_{\backslash \mathcal{W}})}\left[\mathbb{E}_{\bar{p}(\mathcal{W} \mid \Theta_{\backslash \mathcal{W}})} \phi(\mathcal{W}) \mid \Theta_{\backslash \mathcal{W}}\right]$$

2. Delta method:

$$\mathbb{E}_{q_{\mathrm{cur}}(\Theta_{\backslash \mathcal{W}})}\left[\mathbf{h}\left(\Theta_{\backslash \mathcal{W}}\right)\right] \approx \mathbf{h}\left(\mathbb{E}_{q_{\mathrm{cur}}}\left[\Theta_{\backslash \mathcal{W}}\right]\right)$$

ALGO ILLUSTARTION

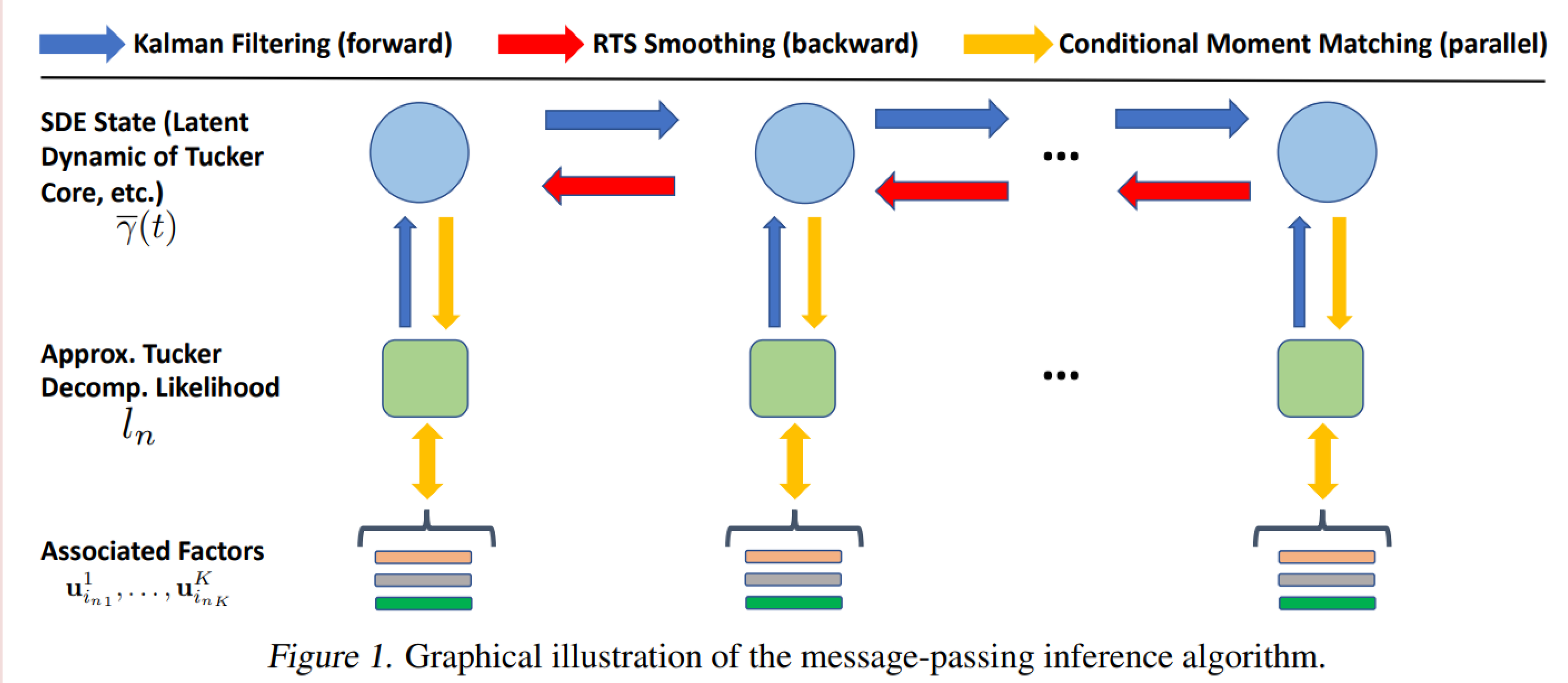
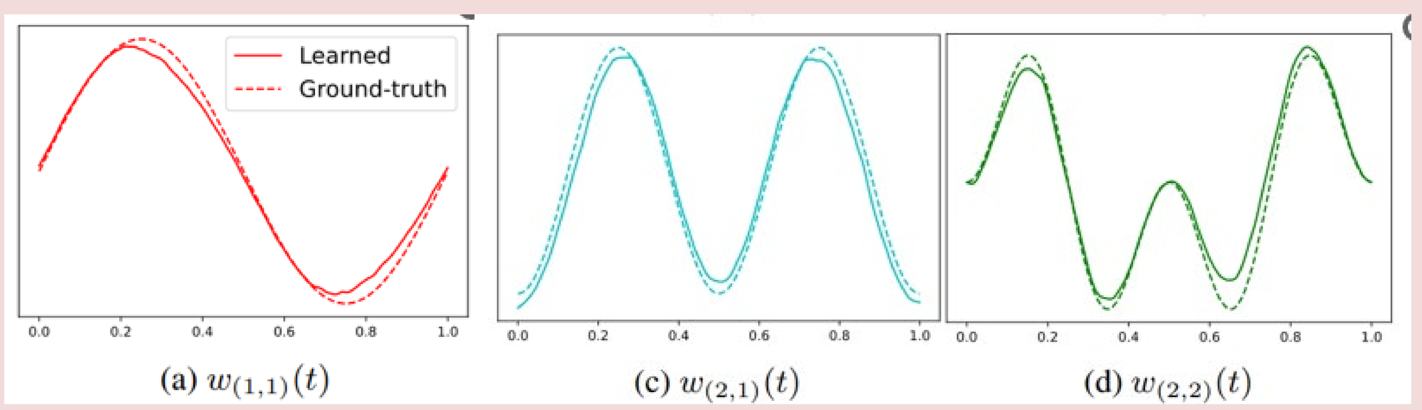


Figure 1. Graphical illustration of the message-passing inference algorithm.

Time cost: $\mathcal{O}(N\bar{R})$ **Space cost:** $\mathcal{O}\left(N\left(\bar{R}^2 + \sum_{k=1}^{K} R_k^2\right)\right)$

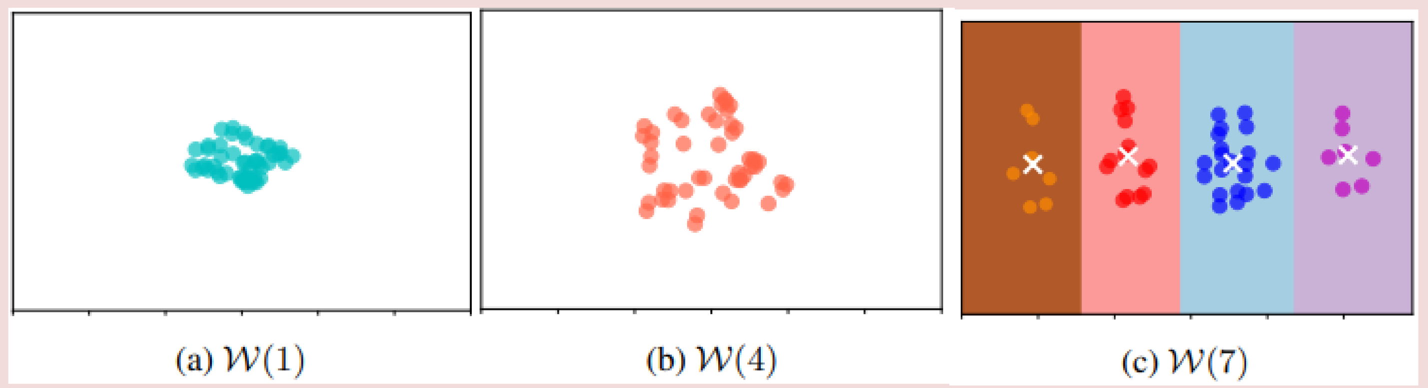
EXPERIMENTS

- **Captured Dynamics on Simu data**



(a) $w_{(1,1)}(t)$ (c) $w_{(2,1)}(t)$ (d) $w_{(2,2)}(t)$

- **Change of Temporal Patterns of real-world dataset (DBLP)**



(a) $\mathcal{W}(1)$ (b) $\mathcal{W}(4)$ (c) $\mathcal{W}(7)$

- **Predictive Performance**

RMSE	*MovieLens*	*AdsClicks*	*DBLP*
CT-CP	1.113 ± 0.004	1.337 ± 0.013	0.240 ± 0.007
CT-GP	0.949 ± 0.008	1.422 ± 0.008	0.227 ± 0.009
DT-GP	0.963 ± 0.008	1.436 ± 0.015	0.227 ± 0.007
DDT-GP	0.957 ± 0.008	1.437 ± 0.010	0.225 ± 0.006
DDT-CP	1.022 ± 0.003	1.420 ± 0.020	0.245 ± 0.004
DDT-TD	1.059 ± 0.006	1.401 ± 0.022	0.232 ± 0.09
BCTT	$\mathbf{0.922 \pm 0.002}$	$\mathbf{1.322 \pm 0.012}$	$\mathbf{0.214 \pm 0.009}$
MAE			
CT-CP	0.788 ± 0.004	0.787 ± 0.006	0.105 ± 0.001
CT-GP	0.714 ± 0.004	0.891 ± 0.011	0.092 ± 0.004
DT-GP	0.722 ± 0.008	0.893 ± 0.008	0.084 ± 0.003
DDT-GP	0.720 ± 0.003	0.894 ± 0.009	0.083 ± 0.001
DDT-CP	0.755 ± 0.002	0.901 ± 0.011	0.114 ± 0.002
DDT-TD	0.742 ± 0.006	0.866 ± 0.012	0.101 ± 0.001
BCTT	$\mathbf{0.698 \pm 0.002}$	$\mathbf{0.777 \pm 0.016}$	$\mathbf{0.084 \pm 0.001}$

RMSE	*MovieLens*	*AdsClicks*	*DBLP*
CT-CP	1.165 ± 0.008	1.324 ± 0.013	0.263 ± 0.006
CT-GP	0.965 ± 0.019	1.410 ± 0.015	0.227 ± 0.007
DT-GP	0.949 ± 0.007	1.425 ± 0.015	0.225 ± 0.008
DDT-GP	0.948 ± 0.005	1.421 ± 0.012	0.220 ± 0.006
DDT-CP	1.141 ± 0.007	1.623 ± 0.013	0.282 ± 0.011
DDT-TD	0.944 ± 0.003	1.453 ± 0.035	0.312 ± 0.072
BCTT	$\mathbf{0.895 \pm 0.007}$	$\mathbf{1.304 \pm 0.018}$	$\mathbf{0.202 \pm 0.009}$
MAE			
CT-CP	0.835 ± 0.006	0.792 ± 0.007	0.128 ± 0.001
CT-GP	0.717 ± 0.012	0.883 ± 0.016	0.092 ± 0.002
DT-GP	0.714 ± 0.005	0.886 ± 0.012	0.084 ± 0.001
DDT-GP	0.707 ± 0.004	0.882 ± 0.015	0.082 ± 0.003
DDT-CP	0.843 ± 0.003	1.082 ± 0.013	0.141 ± 0.004
DDT-TD	0.712 ± 0.002	0.903 ± 0.024	0.221 ± 0.047
BCTT	$\mathbf{0.679 \pm 0.001}$	$\mathbf{0.785 \pm 0.010}$	$\mathbf{0.080 \pm 0.001}$

ACKNOWLEDGEMENTS

The work has been supported by MURI AFOSR grant FA9550-20-1-0358, NSF IIS-1910983, NSF CAREER Award IIS-2046295 and NSF DMS-1848508