Combining physics-informed machine learning with transportation science methods to tackle large-scale urban mobility problems

Carolina Osorio
Google Research, HEC Montreal

Martin Mladenov, Chao Zhang, Chih-wei Hsu, Sanjay Ganapathy Subramaniam, Quan Nguyen, Suyash Vishnoi, Akhil Shetty, Iveel Tsogsuren
Yechen Li, Neha Arora, Andrew Tomkins, Craig Boutilier

Google Research

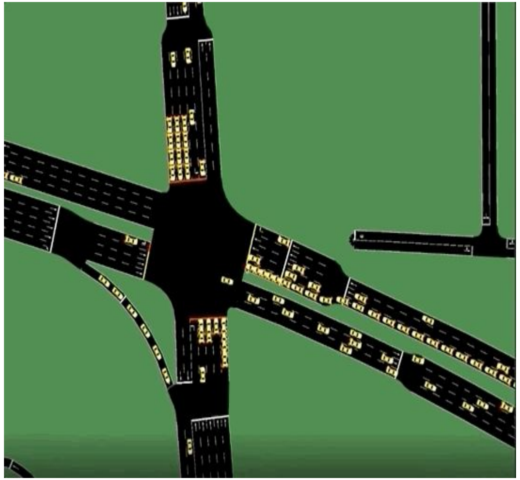
Big Data, AI and Transportation Planning Applications
May 30, 2025, UW, Seattle WA

Physics-informed Traffic Optimization

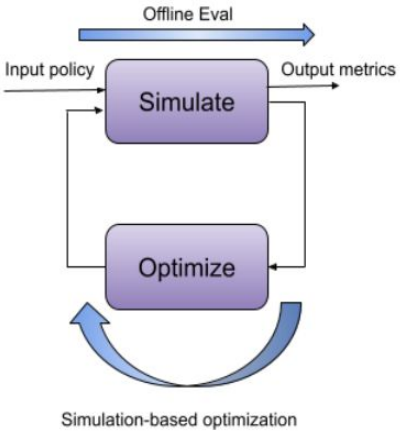
- Digital twins of mobility systems
- Physics-informed algorithms:
  - Sample efficiency
  - Dimensionality reduction
  - Variance reduction
- Future research opportunities

Simulation-based optimization for traffic operations and planning

Building high quality traffic simulations to evaluate traffic management policies and use simulation-based optimization to suggest interventions to optimize transportation networks for sustainability, safety and reduction of congestion.



Simulation of car traffic near an intersection in Salt Lake City



Urban mobility simulation + optimization finds best configurations for infrastructure and policy

Google Research

Carolina Osorio

Simulation Platform

Google Maps
Routing Algorithms

Mobility Demand
(from Census, city
DOTs, 3P data
providers)

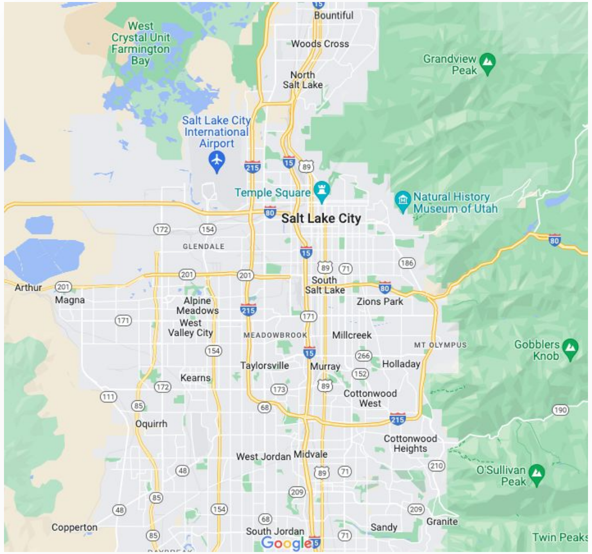
Google Maps
Network Data

Road sensor data
(from city DOTs)

SUMO
[Simulation of
Urban MObility]
based traffic
simulator

Traffic
outputs
(travel time,
speed,
volume,
emissions)

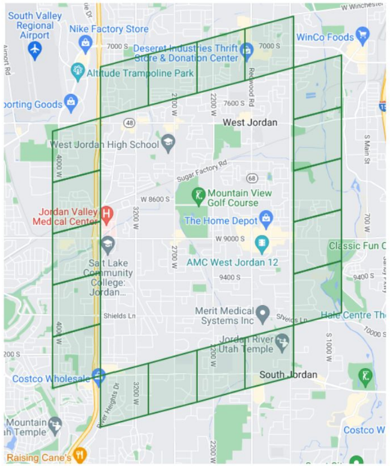
Calibration /
Optimization
Algorithms

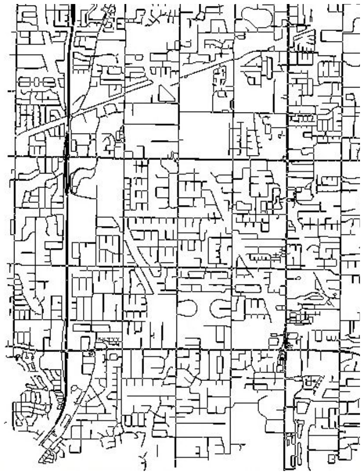


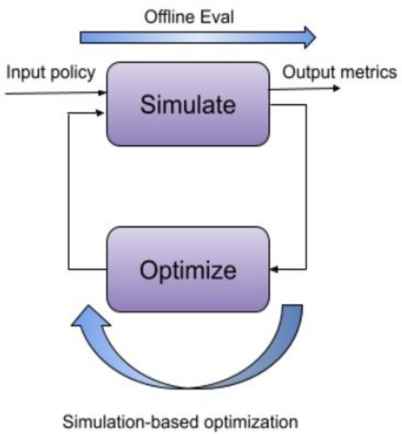
Google Research

Carolina Osorio

OD demand calibration







$$\min_{x} \quad f(x) = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} w_i \left(v_i^{\mathrm{GT}} - E[v_i(x, u_1; u_2)]\right)^2$$
$$\text{s.t.} \quad 0 \le x \le x_U$$

- Goal: find a travel demand input that, once simulated, yields simulated traffic statistics that match ground truth field data.
- Traffic statistics: segment counts, segment speeds.
- Regularization
- Inverse problem
- Challenges: dimensionality, stochasticity, compute-time, non-differentiability

Google Research

(1) Achieving sample efficiency through metamodeling

Google Research

Sample efficiency: metamodeling

$$\min_{x} \quad f(x) = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} w_i \left(v_i^{\mathrm{GT}} - E[v_i(x, u_1; u_2)]\right)^2$$
$$\text{s.t.} \quad 0 \le x \le x_U$$

$\longleftrightarrow$

$$\min_{x} \quad m_k(x; \beta_k) = \beta_{k,0} f_A(x) + \phi(x; \beta_k)$$
$$\text{s.t.} \quad f_A(x) = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} w_i \left(v_i^{\mathrm{GT}} - v_i^{\mathrm{a}}\right)^2$$
$$\phi(x; \beta_k) = \beta_{k,1} + \sum_{z \in \mathcal{Z}} \beta_{k,z+1} x_z$$
$$\boxed{\begin{aligned} & v_i^{\mathrm{a}} = v_i^{\min} + (v_i^{\max} - v_i^{\min}) \left(1 - \left(\frac{q_i}{q_i^{\max}}\right)^{\alpha_i^1}\right)^{\alpha_i^2} \quad \forall i \in \mathcal{I} \\ & q = Ax \end{aligned}}$$
$$0 \le x \le x_U$$

- Combining macroscopic + microscopic models
- Macro: differentiable and compute-efficient
- Micro: high-resolution

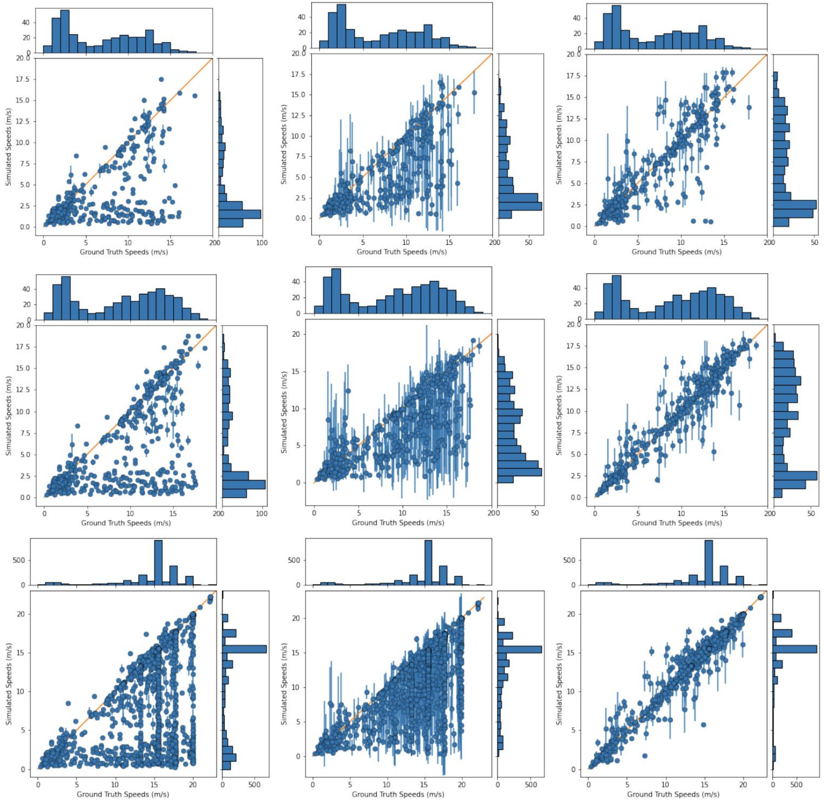
Calibration on:

Path travel times:	Zhang et al. (2024) IEEE ITSC
Segment speeds:	Vishnoi et al. (2023) ACM SIGSPATIAL
Probe segment counts:	Alanqary et al. (2025) TRISTAN

Google Research

Sample efficiency: metamodeling

- Salt Lake City
- Segment speeds calibration
- 62-dimensional instance
- 100 sequential simulations



Vishnoi, S., Tsogsuren, I., Arora, N., Osorio, C. (2023) ACM SIGSPATIAL

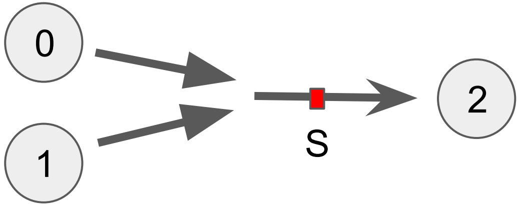
Google Research

(2) Achieving scalability through dimensionality reduction

Google Research

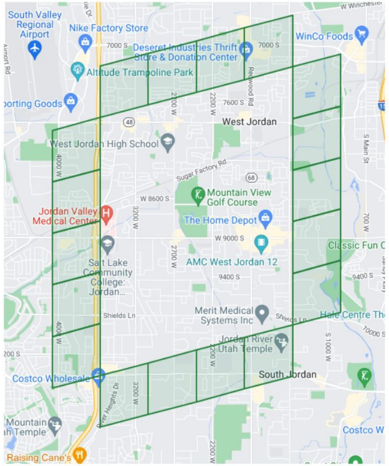
OD demand calibration

$$
\begin{aligned}
\min_{x} \quad & f(x) = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} w_i \left(v_i^{\mathrm{GT}} - E[v_i(x, u_1; u_2)]\right)^2 \\
\text{s.t.} \quad & 0 \leq x \leq x_U
\end{aligned}
$$



- Traditional modeling approach:
  - Discretize space into disjoint zones, traffic analysis zones (TAZes)
  - Consider the set of plausible pairs of zones. High dimensional set !
- OD calibration problem is underdetermined
  - Level of underdetermination typically increases with road network size
- **Should we be solving the problem in this high-dimensional space ?**

- Find lower dimensional spaces to:
  - Enhance computational efficiency of existing calibration algorithms
  - Reduce underdetermination



Google Research

Scalability: dimensionality reduction

- Can we develop sample-efficient transportation-relevant dimensionality reduction techniques ?

- A: **assignment matrix**, network loading map
- Use a low-rank approximation of the assignment matrix A
- Singular value decomposition of A
- Select a submatrix of V of column-dimension k that captures the most information in A
- This low-rank approximation of A minimizes the L2 norm and the Frobenius norm of the difference from A
- Maps from the high d-dimensional OD space to the low k-dimensional space, designed to
    - **Preserve traffic-specific information from the assignment matrix A**
    - **Is sample-efficient: approach does not require any simulation data**
- We theoretically prove that this method reduces the underdetermination of the problem

$$A\,\theta = s$$

$$A = U\,\Sigma\,V^\top$$

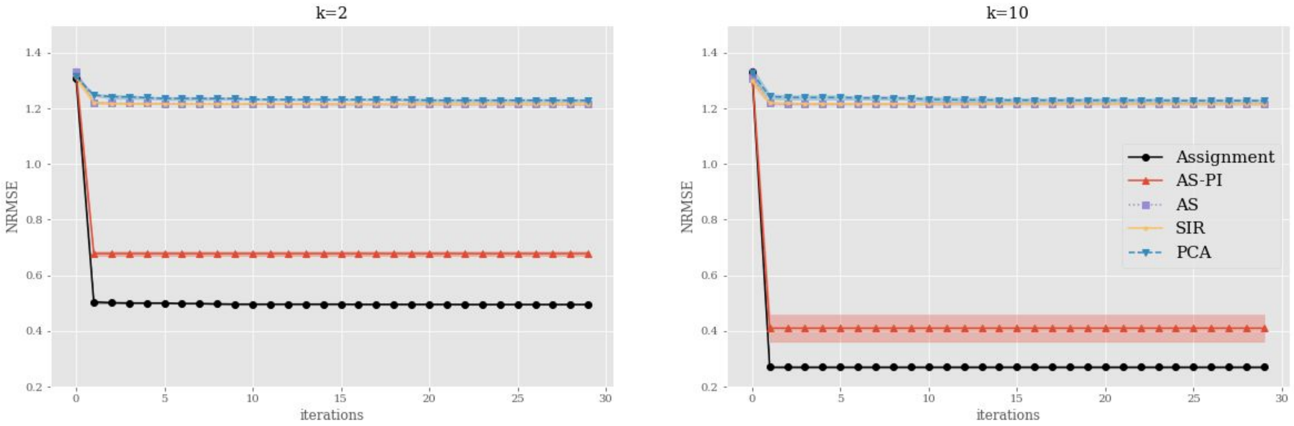
$$\hat{V} \in R^{d\times k}$$

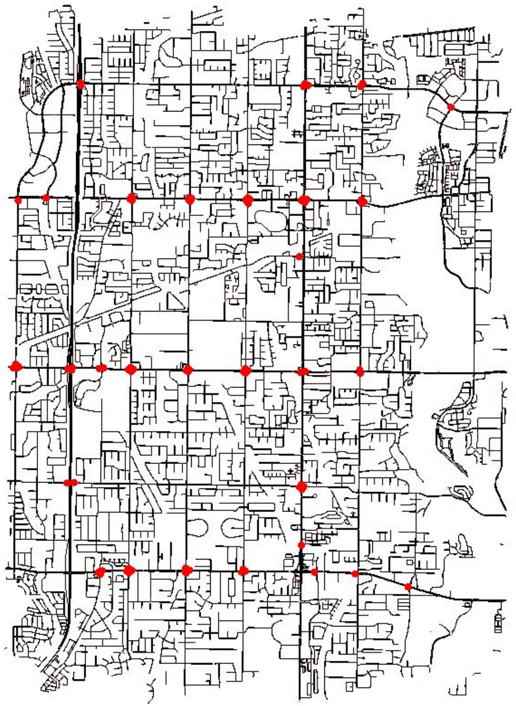
Google Research

Scalability: dimensionality reduction

- 62-dimensional instance
- 79 count sensors
- Optimizer: metamodel, Osorio (2019)

- Dimensionality reduction (DR)
  - PCA principal component analysis, Pearson (1901)
  - SIR: sliced inverse regression, Li et al. (1991)
  - Active subspace, Constantine et al. (2014)
  - Physics-informed active subspace, Nguyen and Osorio (2023)
  - Assignment: Physics-informed metamodel DR, Nguyen and Osorio (2023)



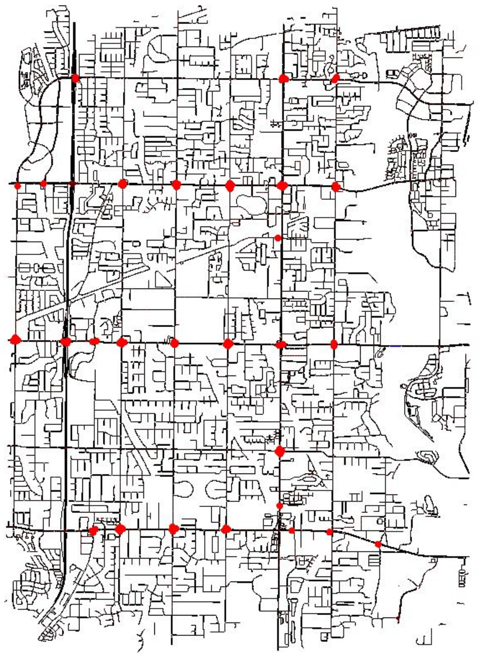


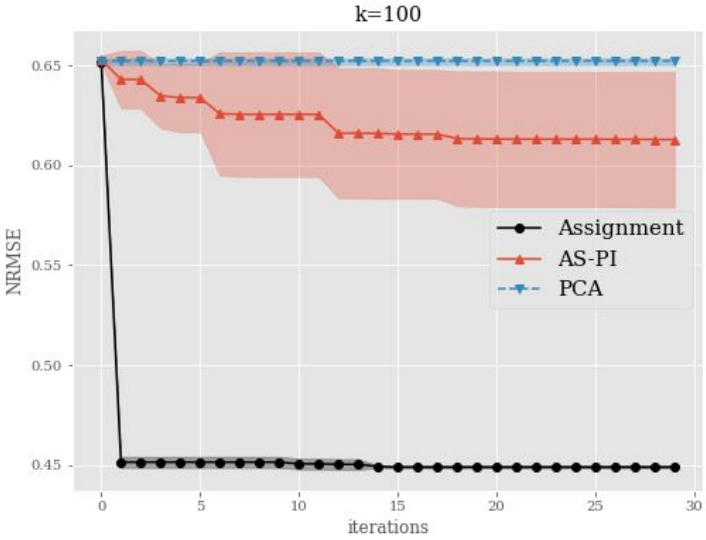
Salt Lake City

Google Research

Scalability: dimensionality reduction

- 2449-dimensional instance
- 76 sensors
- Count data from Utah DOT





- Lack of sample efficiency
  - SIR: sliced inverse regression: requires initial sample of size at least d (OD dimension)
  - Active subspace: requires derivatives (simulation-based or surrogate-based)

Google Research

(3) Achieving scalability through variance reduction

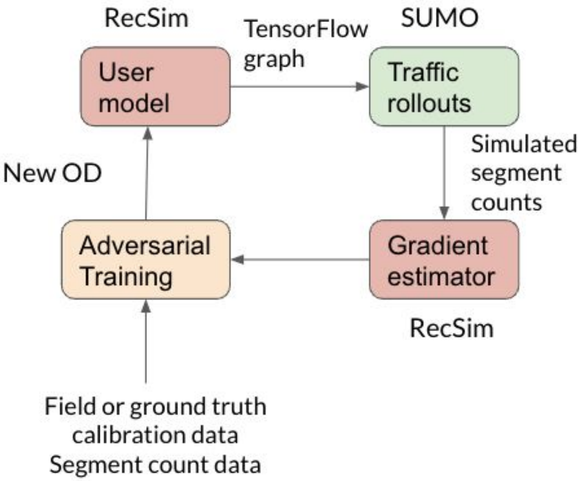
Google Research

Scalability challenges: gradient estimation

- Adversarial formulation
  - Mladenov et al. (2022) ACM SIGSPATIAL
- Gradients can be estimated by a scoring function estimator:

$$\nabla_\omega \mathbb{E}_{\theta\sim\pi_\omega, o\sim s(\theta)}\left[\mathrm{loss}(o)\right] = \mathbb{E}_{\theta\sim\pi_\omega, o\sim s(\theta)}\left[\mathrm{loss}(o)\frac{\nabla_\theta \pi_\omega(\theta)}{\pi_\omega(\theta)}\right]$$

$$\approx \frac{1}{n}\sum_i \mathrm{loss}(o_i)\frac{\nabla \pi_\omega(\theta_i)}{\pi_\omega(\theta_i)} = \frac{1}{n}\sum_i \mathrm{loss}(o_i)\nabla \log \pi_\omega(\theta_i)$$



- The variance of the estimator may grow exponentially with the number of ODs (dimensi
- Mitigation:
  - variance reduction technique
  - metamodel-based control variates

Field data	$o \in O$
Parameter space:	$\theta \in \Theta$
Simulator:	$s_\theta \in \Delta(O)$

Google Research

Variance reduction techniques

- In Monte Carlo simulation the error goes as: $\sigma / \sqrt{n}$
- Reduce variance by:
  - Increasing n
  - <u>Decreasing σ</u>

- Control variates
  - Goal: estimate $E[X] = \nu$

  - Consider a r.v. Y with known mean μ
  - Choose Y such that:
    - $E[X - (Y-\mu)] = \nu$
    - $Var(X - (Y-\mu)) < Var(X)$

  - Simulate: $X - \lambda(Y-\mu)$
  - With an optimal choice of λ: $Var(X - \lambda(Y-\mu)) = Var(X) (1 - \rho^2)$
    ρ : correlation of X and Y
  - Amount of variance reduction depends on how well the two correlate.

- Define metamodel-based control variates

Google Research

Metamodel-based variance reduction

$$\nabla_\omega \mathbb{E}_{\theta\sim\pi_\omega, o\sim s(\theta)}[\text{loss}(o)] \approx \frac{1}{n}\sum_i \text{loss}(o_i)\nabla \log \pi_\omega(\theta_i)$$

$$\nabla_\omega \mathbb{E}_{\theta\sim\pi_\omega, o\sim s(\theta)}[\text{loss}(o)] = \frac{1}{n}\sum_i (\text{loss}(o_i) - \lambda \text{loss}(\mathrm{s_i}))\, \nabla \log \pi_\omega(\theta_i) + \lambda\mu$$

1. Sample ODs θ_i from π_ω
2. Simulate θ_i to obtain counts o_i
3. Compute metamodel counts s_i : $A\,\theta = s$

4. New gradient estimator based on metamodel error

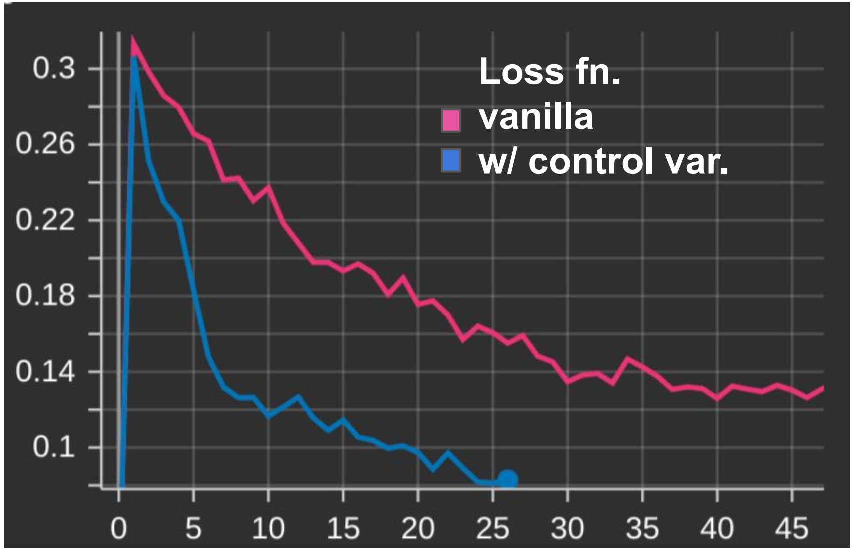
$$\nabla_\omega \mathbb{E}_{\theta\sim\pi_\omega, o\sim s(\theta)}[\text{loss}(o)] = \frac{1}{n}\sum_i (\text{loss}(o_i) - \lambda \text{loss}(\mathrm{s_i}))\, \nabla \log \pi_\omega(\theta_i) + \lambda\mu$$

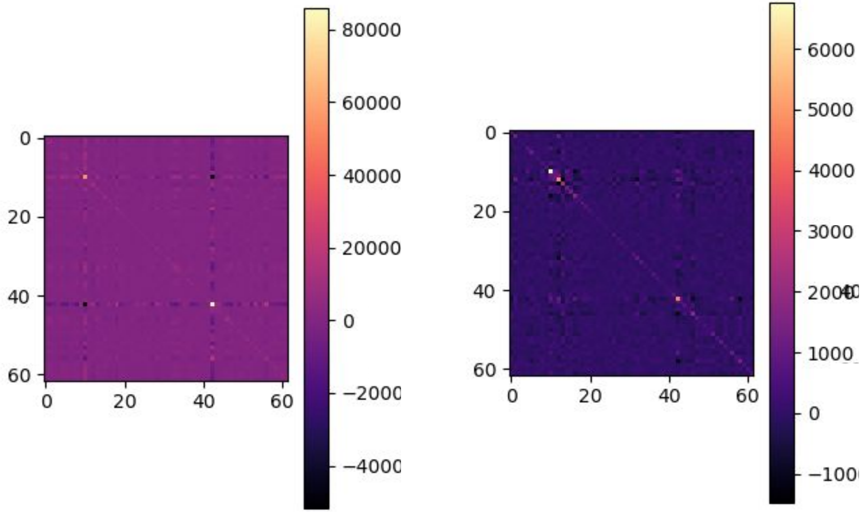
- μ : mean of the metamodel gradient estimator (can be computed on a much larger sample)
- λ : scaling parameter

Google Research

Metamodel-based variance reduction

62-dimensional Salt Lake City case study





- 2x faster convergence
- Variance reduction on the order of 10x is often observed.

Gradient covariance matrix: vanilla vs. control variate

Google Research

Opportunities

- Travel demand calibration: Traffic physics + metamodeling enables scalability through:
  - Enhanced sample efficiency
  - Dimensionality reduction
  - Variance reduction
- Treating the transportation system and data as a gray-box rather than a black-box

- Big-data + deep learning models are not plug-and-play
  - Lack sample efficiency, lack scalability
  - There is great potential to enhance them for transportation applications through the use of simple transportation science ideas

- Most transportation stakeholders live in a data sparse, and/or often compute sparse, environment: Importance of designing data-efficient and sample-efficient methods relevant for transportation

- Data-efficiency
  - How can we make the most out of our existing field data?
    - Use of higher-order statistics from, spatially or temporally, sparse field data
  - How can we use it to:
    - Quantify uncertainty, and derive more resilient and robust mobility solutions ?
      - Quantification of model parameter uncertainty: Greisemer et al. (2024) NeurIPS
    - Reduce the level undetermination or ill-posedness of fundamental transportation problems?

Google Research

Thank You

Google Research