Optimal Transport and Reinforcement Learning to Manage Police Risk

Evan Misshula

April 8, 2025

1 Introduction

Existence of Police Misconduct as a Public Problem

1.1 Existence and Cost of Bad Police Behavior

Police misconduct remains a significant and well-documented challenge for law enforcement agencies in the United States and internationally. Empirical studies consistently demonstrate patterns of abuse, ranging from excessive use of force to psychological violence and systemic bias. For example, Schwartz (2020) analyzes litigation costs associated with police misconduct and argues that the financial burden placed on municipalities undermines incentives for institutional accountability. Similarly, Hickman, Piquero, and Garner (2008) present national estimates of non-lethal force by police, documenting its prevalence and raising concerns about underreporting.

The fiscal impact is equally staggering. In 2024 alone, New York City paid $206 million in legal settlements related to NYPD misconductits highest yearly payout since 2018 (Legal Aid Society, 2025). Over six years, NYC disbursed more than $750 million addressing officer misconduct. These trends are not confined to the U.S.; the Metropolitan Police in London saw legal costs rise to £15 million in 2024, reflecting growing public scrutiny (Financial Times, 2025).

Beyond financial burdens, police misconduct imposes intangible societal costs. Research into racial bias and the "adultification" of Black children in England and Wales found systemic mischaracterizations that result in disproportionate and unjust police interactions (Independent Office for Police Conduct, 2024). In the U.S., a Department of Justice (2024) investigation

into the Memphis Police Department revealed systemic constitutional violations, including discriminatory practices and excessive force.

These examples highlight not only the existence but also the structural nature of police misconduct. Kane and White (2009), in a seminal study of NYPD officers, found that career-ending misconduct is rarely isolatedit tends to be predictable and linked to earlier indicators.

1.2 Problem of Predicting Police Misconduct

Scholars have employed a wide array of methods to understand the origins, patterns, and consequences of police misconduct. Early research emphasized descriptive surveys. For instance, Hickman et al. (2008) developed a national-level understanding of excessive use of force through officer and civilian surveys. Walker (2005) further contextualized these behaviors within systems of accountability, describing how oversight mechanisms shape officer conduct.

Sociological and organizational models added a structural perspective. Smith (2004) examined how organizational policies and department size correlate with use-of-force incidents. Psychological and behavioral approaches also emerged. For example, recent studies have compared officers with histories of serious misconduct to matched comparators to identify distinguishing traits (PubMed, 2022). This line of work suggests that some misconduct is predictable through psychometric assessment.

The role of peer networks has also been emphasized. Investigating the effect of social exposure, network-based studies show that officers are more likely to engage in misconduct if they work closely with colleagues who have already been accused of similar behavior (Schwartz, 2020; Rozema & Schanzenbach, 2019).

With advances in computational modeling, scholars now use predictive analytics and machine learning to forecast misconduct. Stoddard, Fitzpatrick, and Ludwig (2024) show that prior complaints and deployment patterns can be used to identify at-risk officers. Similarly, Ensign et al. (2017) examine how feedback loops in predictive policing systems may exacerbate racial disparities and lead to increased targeting of marginalized communities.

Collectively, these approaches-from institutional theory to predictive modelingoffer a layered understanding of the causes and costs of misconduct and provide a foundation for reform and early intervention strategies.

1.2.1 Traditional Predictive Models

Statistical modeling techniques have been foundational in the fields of statistics, economics, and sociology, providing essential tools for analyzing relationships between variables and predicting outcomes. In the realm of criminal justice, these models have been employed to understand crime patterns, evaluate policy impacts, and forecast criminal behavior. However, their application is accompanied by limitations, particularly concerning metrics like R^2 and Mean Squared Error (MSE), especially when dealing with out-of-sample data and large, unstructured datasets such as those derived from body-worn cameras or textual records.

1. Linear Regression

   Linear regression estimates the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. In criminal justice, it's employed to assess factors influencing crime rates or the effectiveness of interventions. For instance, studies have used linear regression to explore socioeconomic determinants of crime (Weisburd & Britt, 2009). However, the reliance on R^2 as a measure of fit can be misleading. A high R^2 indicates that the model explains a large proportion of variance in the dependent variable, but it doesn't confirm the model's predictive accuracy on new data. Moreover, linear regression assumes a linear relationship and is sensitive to outliers, which can distort results. When applied to large, unstructured datasets, such as video or text from body-worn cameras, linear regression becomes impractical due to its inability to capture complex, nonlinear patterns inherent in such data.

2. Logistic Regression

   Logistic regression is utilized when the dependent variable is binary, modeling the probability of a particular outcome. In criminal justice, it's often used to predict recidivism or the likelihood of reoffending (Berk, 2012). While logistic regression doesn't produce an R^2 value analogous to linear regression, pseudo-R^2 measures exist but are interpreted differently. The model's performance is typically evaluated using metrics like accuracy, sensitivity, and specificity. Logistic regression assumes a linear relationship between independent variables and the log odds of the dependent variable, which may not hold in complex real-world scenarios. Additionally, it struggles with large-scale unstructured data, as it cannot effectively process the high dimensionality and variability present in such datasets.

3. Time Series Analysis

   Time series analysis examines data points collected or recorded at specific time intervals to identify trends, cycles, and seasonal variations. In criminal justice, it's applied to forecast crime rates and allocate resources accordingly (Corman et al., 1987). Metrics like Mean Squared Error (MSE) are used to assess model accuracy, but they are scale-dependent and can be influenced by outliers. Time series models assume that past patterns will continue, which may not account for sudden changes or anomalies. Furthermore, these models are less effective with large, unstructured datasets, as they require structured, time-indexed data and cannot easily incorporate the complexity of information from sources like body-worn cameras or textual data.

4. Structural Equation Modeling (SEM)

   SEM combines factor analysis and multiple regression to examine complex relationships among variables, including latent constructs. In criminal justice, SEM has been used to study theoretical constructs like the relationship between social factors and criminal behavior (Gau, 2010). Fit indices such as the Root Mean Square Error of Approximation (RMSEA) and Comparative Fit Index (CFI) evaluate model adequacy. However, SEM requires large sample sizes and assumes multivariate normality, which can be restrictive. Its application to large, unstructured datasets is limited due to challenges in specifying models that can accommodate the complexity and variability of such data.

5. Hierarchical Linear Modeling (HLM)

   HLM, or multilevel modeling, analyzes data with nested structures, such as individuals within neighborhoods. It's particularly useful in criminal justice for assessing contextual effects on individual behaviors (Raudenbush & Bryk, 2002). While HLM provides insights into variance at multiple levels, interpreting R^2 in this context is complex, as it doesn't partition variance neatly across levels. The model assumes normality and homoscedasticity at each level, which may not be realistic. Additionally, HLM is not well-suited for large, unstructured datasets, as it relies on clearly defined hierarchical structures and cannot efficiently process the high-dimensional data typical of sensor outputs or textual records.

6. Limitations in Handling Large, Unstructured Data

Traditional statistical models like those discussed are designed for structured data with clearly defined variables and relationships. They struggle with large, unstructured datasets due to several reasons:

- **Dimensionality**: Unstructured data often have a vast number of features, making traditional models computationally inefficient and prone to overfitting.
- **Nonlinearity**: The relationships in unstructured data are often nonlinear and complex, which linear models cannot capture effectively.
- **Data Preprocessing**: Transforming unstructured data (e.g., text, images) into a format suitable for traditional models requires extensive preprocessing, which can lead to information loss.

As a result, alternative methods such as machine learning algorithms, including deep learning, are more appropriate for analyzing unstructured data from sources like body-worn cameras or textual records.

1.3 Intuition for the inadequacy of traditional methods

Rather than thinking in the most abstract case first, let's take a familiar example linear regression. In simple linear regression we are trying to the best linear relationship between two variables. We are minimizing:

$$L(\hat{\beta}, x, y) = \arg\min_{\beta}(y - \hat{\beta} \cdot x)^2. \tag{1}$$

In many contexts the *loss function* above is also defined:

$$L(\hat{\beta}, x, y) = \hat{\epsilon} \tag{2}$$

In the Equations 1 and 2, I am assumeing y is a $1 \times n$ scaled and mean 0 vector of our dependent values, $\hat{\beta}$ is a scalar real valued constant. The dependent value is also mean 0 which saves us from having to worry about constants.

What is important here is that the relationship between the *scaled* or *regularized* independent and dependent variables is *linear* of *affine* if we did not rescale the variables. From a geometric point of view when we use multi-variable regression we continue to specify a linear relationship but in for dimension greater than 1, we are now specifing a hyper-plane in the number of dimensions of our independent variable usueally, d.

$$L(\hat{\beta}, x, y) = \arg\min_{\beta}(y - \hat{\beta}'x)'(y - \hat{\beta}'x) \tag{3}$$

Notice that in Equation 3, I have expicitly used the notation from linear algebra. Here y is a $1 \times n$ vector of observations where n is the number of observations while x is a $n \times d$ vector of observations where d is the number of variables while again n is the number of observations.

We can extend estimation in this manner by looking at *polynomial estimation*. Here we can approximate a a conftinuous function by taking polynmials of any number of variables. However Hastie (2016) and Bishop (2006) document that these methods suffer from instability.

1.4 More recent approaches

Risk modeling has traditionally relied on conventional estimation methodssuch as maximum likelihood estimation or linear regressionthat assume a fixed underlying structure and static parameters. While these methods have long provided a basis for decision-making, they can struggle to capture the dynamic and often non-stationary nature of risk in complex systems. In response, advanced methods in risk modeling have emerged that leverage frameworks from reinforcement learning, decision theory, and adversarial modeling to address these limitations.

1.5 Conventional Model Estimation

In fact, conventional model estimation typically involves fitting a pre-specified parametric model to historical data. Such approaches, though computationally efficient and theoretically grounded in statistical inference (Hastie, Tibshirani, & Friedman, 2009), often fall short in environments where risk factors evolve over time. Their static nature can lead to suboptimal performance when the system dynamics change, as these methods do not inherently incorporate mechanisms for continuous adaptation or learning.

1.6 Markov Decision Processes and Reinforcement Learning

Advanced risk modeling increasingly adopts a framework based on Markov Decision Processes and reinforcement learning. Markov Decision Processes offer a formal structure for modeling sequential decision-making under uncertainty, where the current state encapsulates all relevant past information. Reinforcement learning builds on this by allowing an agent to learn optimal policies through trial and error. As Sutton and Barto (2018) explain, "reinforcement learning enables agents to learn optimal policies through direct interaction with the environment" (p. 42). This dynamic approach permits

continuous adjustment of risk assessments as new data becomes availablea significant advantage over conventional static models.

1.7 Policy Evaluation and Optimization

Within the reinforcement learning framework, policy evaluation and optimization are central components. Policy evaluation involves estimating the value functionthe expected cumulative rewardfrom a given state under a particular policy. Optimization then iteratively improves the policy based on these evaluations. This cycle contrasts sharply with conventional methods that typically require a one-shot estimation process. The iterative nature of reinforcement learning allows for fine-tuning in real time, accommodating the uncertainties inherent in risk environments. Such adaptability is crucial for scenarios where risk profiles may change abruptly or evolve gradually over time.

1.8 Actor-Critic Methods

Actor-critic methods represent a hybrid approach that combines the strengths of both policy-based and value-based methods. The "actor" proposes actions based on a parameterized policy, while the "critic" evaluates these actions using a value function. According to Konda and Tsitsiklis (2000), actor-critic algorithms "combine the benefits of policy gradient methods and value function approximation, leading to more stable and efficient learning" (p. 93). This dual mechanism not only improves convergence properties but also enhances the robustness of the model when faced with the uncertainties typical in risk management scenarios.

1.9 Generative Adversarial Networks (GANs)

Originally introduced for tasks in image and text generation, Generative Adversarial Networks (GANs) have been adapted to risk modeling to simulate complex data distributions. In a GAN framework, two modelsa generator and a discriminatorengage in a competitive process that refines the ability of the generator to produce realistic data. Goodfellow et al. (2014) described GANs as "a novel framework for estimating generative models via an adversarial process" (p. 2672). In risk modeling, this adversarial process can be used to explore a wide range of potential risk scenarios, providing insights that are not readily available through traditional estimation methods.

1.10 Summary of modern estimation methods

In contrast to conventional model estimationwhich often relies on static assumptions and single-step fittingadvanced methods in risk modeling offer a dynamic and adaptive approach. By leveraging frameworks such as Markov Decision Processes, reinforcement learning, actor-critic methods, and GANs, modern risk models can continuously learn from data, accommodate changing environments, and better capture complex risk dynamics. These advanced techniques not only provide more robust estimates in real-world applications but also open new avenues for exploring and mitigating risk in uncertain and volatile domains.

2 Introduction to Markov Decision Making and Reinforcement Learning

In reinforcement learning, decision-making under uncertainty is modeled using a Markov Decision Process (MDP). An MDP is defined by the tuple:

$$(\mathcal{S}, \mathcal{A}, P, R, \gamma) \tag{4}$$

where:

- $\mathcal{S}$: set of all possible states. Example: $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$
- $\mathcal{A}$: set of all possible actions. Example: $\mathcal{A} = \{a_1, a_2, \ldots, a_M\}$
- $P(s' \mid s, a)$: probability of transitioning to state s' from state s after taking action a:

$$P(s' \mid s, a) = \Pr(s_{t+1} = s' \mid s_t = s, a_t = a) \tag{5}$$

- $R(s, a)$: reward function that gives the immediate reward received after taking action a in state s
- $\gamma \in [0, 1]$: discount factor that weights future rewards

3 Policy and Value Functions

A policy π maps states to a probability distribution over actions:

$$\pi(a \mid s) = \Pr(a_t = a \mid s_t = s) \tag{6}$$

State-Value Function $V^\pi(s)$: expected cumulative discounted reward when starting in state s and following policy π:

$$V^\pi(s) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \,\middle|\, s_0 = s \right] \tag{7}$$

Action-Value Function $Q^\pi(s, a)$: expected return starting from state s, taking action a, and then following policy π:

$$Q^\pi(s, a) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \,\middle|\, s_0 = s, a_0 = a \right] \tag{8}$$

4 Bellman Equations

Bellman equation for $V^\pi(s)$:

$$V^\pi(s) = \sum_{a \in \mathcal{A}} \pi(a \mid s) \left[R(s, a) + \gamma \sum_{s' \in \mathcal{S}} P(s' \mid s, a) V^\pi(s') \right] \tag{9}$$

Bellman optimality equation for $V^*(s)$:

$$V^*(s) = \max_{a \in \mathcal{A}} \left[R(s, a) + \gamma \sum_{s' \in \mathcal{S}} P(s' \mid s, a) V^*(s') \right] \tag{10}$$

Bellman optimality equation for $Q^*(s, a)$:

$$Q^*(s, a) = R(s, a) + \gamma \sum_{s' \in \mathcal{S}} P(s' \mid s, a) \max_{a' \in \mathcal{A}} Q^*(s', a') \tag{11}$$

5 Reinforcement Learning Objective

Reinforcement learning methods seek a policy π^* that maximizes the expected return:

$$J(\pi) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \right] \tag{12}$$

This involves:

- **Exploration**: sampling unknown actions to improve knowledge

- **Exploitation**: using current estimates to select high-value actions
- **Policy Evaluation**: estimating V^π or Q^π
- **Policy Improvement**: updating π using the estimated values

6 Summary of set up of Sequention Markov Decisionmaking

This outlines the formal structure and notation of an MDP and how reinforcement learning can be applied to find optimal policies by evaluating and improving value functions iteratively. We have set this up as a stochastic optimization problem, so we are primed to use stichastic calculus, and algebraic manipulation to work towards a solution. However, there is also another important element. The surface of possible actions and the surface of expected rewards are manifolds.

7 Introduction

In reinforcement learning , agents interact with environments characterized by high-dimensional state spaces, while the action and reward spaces are often of much lower dimensionality. This disparity suggests that the agent's decision-making process operates on a lower-dimensional manifold within the higher-dimensional state space. Information geometry provides a mathematical framework to understand and exploit this structure by modeling the space of probability distributions as a Riemannian manifold, where the Fisher information metric defines the geometry.

8 Information Geometry and the Fisher Information Metric

Information geometry is an interdisciplinary field that applies differential geometric methods to study probability theory and statistics. It conceptualizes statistical models as geometric structures, providing valuable insights into the relationships between different probability distributions.

8.0.1 Manifolds in Information Geometry:

In this context, a **manifold** is a mathematical space that locally resembles Euclidean space but may have a more complex global structure. Specifically, a **statistical manifold** is a smooth manifold where each point represents a distinct probability distribution. This framework allows for the application of geometric concepts to analyze statistical models.

8.0.2 Metrics and Riemannian Metrics:

A **metric** on a manifold provides a way to measure distances between points, enabling the quantification of differences between probability distributions. When this metric varies smoothly across the manifold and defines an inner product on the tangent space at each point, it is termed a **Riemannian metric**. This structure facilitates the application of calculus on manifolds, allowing for the measurement of angles, lengths, and volumes, and the definition of concepts like curvature.

8.0.3 Fisher Information Metric:

A pivotal Riemannian metric in information geometry is the **Fisher information metric**. For a family of probability distributions parameterized by $\boldsymbol{\theta} = (\theta^1, \theta^2, \ldots, \theta^d)$, the Fisher information matrix $g_{ij}(\boldsymbol{\theta})$ is defined as:

$$g_{ij}(\boldsymbol{\theta}) = \mathbb{E}\left[\frac{\partial \log p(X;\boldsymbol{\theta})}{\partial \theta^i}\frac{\partial \log p(X;\boldsymbol{\theta})}{\partial \theta^j}\right],$$

where $p(X;\boldsymbol{\theta})$ is the probability density function of the random variable X parameterized by $\boldsymbol{\theta}$, and the expectation is taken with respect to $p(X;\boldsymbol{\theta})$. This metric provides a measure of the sensitivity of the probability distribution to changes in the parameters, effectively capturing the local curvature of the statistical manifold.

8.1 Characteristics of the Spaces into which the manifolds are emebeded

Furtermore, information geometry extends the principles of differential geometry to the study of probability distributions, treating them as points on a manifold. This geometric framework provides profound insights into the structure of statistical models. Central to this framework are concepts such as affine connections, duality, and dually flat spaces, which are instrumental in understanding the geometric properties of statistical manifolds.

8.2 Affine Connections and Duality

An **affine connection** on a differentiable manifold M is a tool that allows for the differentiation of vector fields on the manifold. Formally, it is a bilinear map $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ satisfying:

1. $\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z$
2. $\nabla_X (fY) = f\nabla_X Y + (Xf)Y$

for all vector fields $X, Y, Z \in \mathfrak{X}(M)$ and smooth functions f, g on M. This structure enables the comparison of tangent vectors at different points, facilitating the definition of parallel transport and covariant differentiation

In information geometry, two specific affine connections are of particular interest:

1. **e-Connection (Exponential Family):** Associated with the natural parameters of an exponential family of probability distributions, the e-connection aligns with the geometry induced by the Fisher information metric.
2. **m-Connection (Mixture Family):** Corresponding to the expectation parameters in a mixture family, the m-connection reflects the dualistic nature of statistical manifolds, capturing the geometry from the perspective of mixture parameters.

These dual connections are fundamental in understanding the geometric structure of statistical models, as they provide complementary perspectives on the curvature and connection properties of the manifold.

8.3 Dually Flat Spaces

A **dually flat space** is a statistical manifold that is flat with respect to both the e-connection and the m-connection. This dual flatness implies the existence of dual affine coordinate systems and a potential function whose Hessian defines the Riemannian metric. In such spaces, the e- and m-connections are mutually dual relative to the Fisher information metric, leading to elegant geometric properties and simplifications in statistical inference

The significance of dually flat spaces lies in their ability to generalize Euclidean space properties to the realm of information geometry. They facilitate the development of efficient algorithms for statistical estimation and inference by leveraging the natural duality between the e- and m-connections

8.3.1 Importance in Machine Learning:

In machine learning, understanding the geometric structure of parameter spaces is crucial for developing efficient algorithms. The Fisher information metric allows practitioners to account for the underlying geometry of statistical models, leading to more natural and efficient optimization techniques. For instance, natural gradient descent leverages this metric to navigate the parameter space more effectively than standard gradient descent, resulting in improved convergence properties.

8.3.2 Application to Complex Problems like Police Misconduct:

When addressing complex societal issues such as police misconduct, machine learning models can be employed to analyze patterns and predict outcomes. By representing the space of possible behaviors or incidents as a statistical manifold equipped with the Fisher information metric, one can gain deeper insights into the relationships between different variables and their influence on outcomes. This geometric perspective facilitates the development of models that are more sensitive to the nuances of the data, potentially leading to more accurate predictions and informed policy decisions.

In summary, the concepts of manifolds, metrics, and Riemannian metrics are foundational in information geometry and play a significant role in the study and solution of complex machine learning problems. They provide a structured framework for understanding the geometry of statistical models, leading to more effective analysis and interpretation of data.

9 Natural Gradient Descent in Reinforcement Learning

Traditional gradient descent methods update parameters in the direction of the steepest descent in the parameter space, which may not be optimal on a curved manifold. Natural gradient descent addresses this by taking into account the manifold's curvature, updating parameters in the direction of the steepest descent in the space of probability distributions. The natural gradient $\tilde{\nabla} J(\boldsymbol{\theta})$ is given by:

$$\tilde{\nabla} J(\boldsymbol{\theta}) = G(\boldsymbol{\theta})^{-1} \nabla J(\boldsymbol{\theta}), \tag{13}$$

where $J(\boldsymbol{\theta})$ is the objective function, $\nabla J(\boldsymbol{\theta})$ is the standard gradient, and $G(\boldsymbol{\theta})$ is the Fisher information matrix. This approach ensures that

updates are made in a manner that respects the underlying geometry of the parameter space, leading to more efficient and stable convergence in learning algorithms.

In the context of reinforcement learning, natural gradient methods have been applied to policy optimization, resulting in algorithms like the natural policy gradient. These methods leverage the geometric structure to guide policy updates, improving convergence rates and stability in learning processes.

10 Dimensionality Considerations in Action and Reward Spaces

The recognition that the action and reward spaces often have much lower dimensionality than the state space has significant implications for policy optimization. By focusing on the lower-dimensional manifold where decision-making occurs, algorithms can be designed to operate more efficiently, reducing computational complexity without sacrificing performance. This insight aligns with findings that, under certain conditions, the dimensionality of the manifold of reachable states is at most the dimensionality of the action space plus one.

11 Summary of connection to Informational Geometry

Incorporating information geometry into reinforcement learning provides a principled approach to addressing the challenges posed by high-dimensional state spaces. By leveraging the geometric structure of the space of probability distributions, algorithms can achieve more efficient and stable learning, particularly when the action and reward spaces are of much lower dimensionality than the state space. Natural gradient methods exemplify this approach, offering a geometrically informed alternative to traditional gradient descent techniques in the optimization of policies within reinforcement learning frameworks.

12 Optimal Tranport

Another geometric framework relevant to reinforcement learning and probabilistic modeling is **optimal transport (OT)**. Originating from problems

posed by Gaspard Monge (1781) and later reformulated by Leonid Kantorovich (1942), the OT framework provides a principled way to measure the distance between probability distributions, accounting for the geometry of the underlying space. This framework has found powerful applications in modern machine learning, particularly through the **Wasserstein distance**, which can be interpreted as a Riemannian metric on the space of probability measures.

13 The Monge Problem

The **Monge formulation** of optimal transport seeks a transport map $T : \mathcal{X} \to \mathcal{Y}$ that pushes a probability distribution μ onto another distribution ν, minimizing the total cost of transport. Mathematically:

$$\inf_{T:T_{\#}\mu=\nu} \int_{\mathcal{X}} c(x, T(x))\, d\mu(x), \tag{14}$$

where $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$ is a cost function (often $c(x, y) = \|x - y\|^p$), and $T_{\#}\mu$ denotes the pushforward measure of μ by T. Monge's problem is nonlinear and may not always have a solution.

14 The Kantorovich Relaxation

Kantorovich introduced a relaxed version of Monge's problem using **transport plans** $\gamma \in \Pi(\mu, \nu)$, where $\Pi(\mu, \nu)$ denotes the set of all joint distributions with marginals μ and ν. The **Kantorovich problem** becomes:

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\, d\gamma(x, y). \tag{15}$$

This relaxation turns the problem into a linear program over a convex set, guaranteeing the existence of a solution under mild conditions.

15 Brenier and Alexandrov's Contributions

In the theory of optimal transport, the foundational contributions of Brenier, Alexandrov, and Rockafellar provide critical insights into the existence, uniqueness, and structure of optimal transport maps, particularly when the cost function is quadratic.

Let μ and ν be probability measures on $\mathbb{R}^d$ with finite second moments, and assume that μ is absolutely continuous with respect to the Lebesgue

measure. Consider the quadratic cost function $c(x, y) = \frac{1}{2}\|x-y\|^2$. Brenier's theorem asserts the existence of a unique optimal transport map $T : \mathbb{R}^d \to \mathbb{R}^d$ that pushes forward μ to ν, and this map is characterized as the gradient of a convex function $\phi : \mathbb{R}^d \to \mathbb{R}$:

$$T(x) = \nabla\phi(x).$$

This result connects optimal transport with convex analysis and PDEs, making it amenable to efficient numerical computation. Similarly, Alexandrov developed deep geometric insights related to the curvature of convex surfaces and measure-preserving maps, which underpin the geometric intuition of OT.

Alexandrov's theorem pertains to the regularity properties of convex functions. It states that any convex function $\phi : \mathbb{R}^d \to \mathbb{R}$ is twice differentiable almost everywhere. In the context of optimal transport, this implies that the convex potential ϕ from Brenier's theorem is not only differentiable but possesses a Hessian $D^2\phi(x)$ at almost every point $x \in \mathbb{R}^d$. This regularity is crucial for understanding the geometric properties of the transport map and ensures that the map $T(x) = \nabla\phi(x)$ behaves well in a measure-theoretic sense.

16 Convexity and Rockafellars Theorem

Rockafellar's theorem (1970) provides a characterization of cyclically monotone operators. It states that a set-valued operator $T : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is the subdifferential of a convex, lower semicontinuous, and proper function if and only if it is maximally cyclically monotone. In the framework of optimal transport, this theorem implies that if a transport map T is cyclically monotone, there exists a convex function ϕ such that $T = \nabla\phi$. This result underpins the uniqueness and structure of optimal transport maps in Monge's problem when the cost function is convex, linking the concepts of monotonicity in optimization to the geometry of transport maps.

This connects optimal transport directly to convex optimization, opening the door for first-order and variational optimization methods in learning.

Collectively, these theorems form the theoretical backbone of optimal transport theory, elucidating the deep interplay between convex analysis, geometry, and the structure of optimal mappings.

17 Wasserstein Distance and Riemannian Geometry

The **Wasserstein distance** $W_p(\mu, \nu)$, particularly the 2-Wasserstein distance W_2, quantifies the cost of transporting mass between μ and ν over a metric space:

$$W_2^2(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} \|x - y\|^2 \, d\gamma(x, y). \tag{16}$$

In the space of probability measures $\mathcal{P}_2(\mathbb{R}^d)$, the 2-Wasserstein distance defines a Riemannian metric, as shown in the works of Otto and others. This endows the space of distributions with a geometry where gradient flows and natural transport paths (geodesics) can be computed and used for modeling and optimization.

18 Applications in Reinforcement Learning and Machine Learning

In reinforcement learning, optimal transport provides tools for:

- Regularizing policy updates via Wasserstein penalties,
- Comparing distributions of trajectories or value functions,
- Enforcing smooth interpolations between policies.

Moreover, Wasserstein distances have been used in generative modeling (e.g., Wasserstein GANs) and imitation learning, where geometry-aware loss functions improve sample efficiency and stability.

19 Optimal Transport Computational Challenges and Solutions

Optimal transport (OT) theory focuses on determining the most cost-effective way to transform one probability distribution into another, minimizing a predefined cost function. This problem is computationally intensive due to its formulation as a linear programming task involving constraints that ensure the preservation of mass during transportation. The complexity arises from the need to solve for a transport plan that satisfies these constraints while

minimizing the transportation cost, leading to significant computational demands, especially as the size of the distributions increases.

The classical OT problem involves solving a linear program with a cost matrix that grows quadratically with the number of support points in the distributions. Traditional methods, such as the network simplex algorithm, can be computationally prohibitive for large-scale problems due to their super-cubic time complexity.

19.0.1 Sinkhorn Distances and Entropic Regularization

To address these computational challenges, entropic regularization techniques have been introduced. By adding an entropy term to the OT problem, the objective function becomes strictly convex, facilitating more efficient optimization. This approach leads to the formulation of Sinkhorn distances, which can be computed using the Sinkhorn-Knopp algorithm. This iterative method alternates between normalizing the rows and columns of the transport matrix, converging to an approximate solution of the regularized OT problem. The computational complexity of the Sinkhorn algorithm is significantly lower than that of traditional methods, making it more suitable for large-scale applications.

19.0.2 Bregman Projections

Bregman projections offer another approach to solving regularized OT problems. These methods iteratively project onto the set of coupling matrices that satisfy the marginal constraints, effectively handling the entropic regularization term. The iterative Bregman projection algorithm has been shown to converge to the optimal solution efficiently, providing a practical alternative to the Sinkhorn algorithm.

19.0.3 Accelerated Gradient Descent Methods

Recent advancements have explored the use of accelerated gradient descent methods to further improve the efficiency of solving OT problems. By approximating the Kantorovich dual potential using smooth functions, these methods can leverage fast proximal gradient algorithms to achieve precise estimates of the OT cost with reduced computational complexity.

20 Risks to privacy

Monitoring police behavior using sensor equipment to collect Big Data introduces several ethical considerations that must be carefully addressed:

The deployment of sensor technologies, such as body cameras and facial recognition systems, can infringe upon the privacy rights of both officers and the public. Continuous surveillance may lead to a "chilling effect," where individuals alter their behavior due to the perception of being constantly watched, potentially impacting civil liberties (Brayne, 2017).

Machine Learning and big data analytics in policing can perpetuate existing biases if the underlying data reflects historical inequalities. For instance, predictive policing algorithms may disproportionately target marginalized communities, leading to over-policing and reinforcing systemic discrimination (Selbst, 2018).

The use of complex algorithms and data-driven tools necessitates transparency to ensure accountability. However, many of these systems operate as "black boxes," making it difficult for the public and oversight bodies to understand and scrutinize their decision-making processes (Ferguson, 2017).

The collection of data through sensor equipment often occurs without explicit consent from those being monitored. This raises concerns about individual autonomy and the right to control one's personal information (Brayne, 2017).

There is often a lack of comprehensive legal frameworks governing the use of Big Data technologies in policing. This absence can lead to ethical dilemmas and potential abuses of power, highlighting the need for clear policies and regulations (Ferguson, 2017).

The accumulation of large datasets on individuals poses significant security risks. Unauthorized access or misuse of this data can have severe consequences, including wrongful arrests or the exposure of sensitive personal information (Selbst, 2018).

Addressing these ethical considerations requires implementing robust policies that promote transparency, ensure accountability, protect privacy, and prevent discrimination. Engaging with community stakeholders and establishing independent oversight mechanisms are also crucial steps in the ethical deployment of sensor technologies in policing.

21 Conclusion

The optimal transport framework offers powerful geometric insights into the structure of probability distributions. Through Monges and Kantorovichs formulations, convex optimization, and the Brenier map, OT connects naturally with information geometry. The Wasserstein distance, as a Riemannian metric on probability spaces, enables more principled and efficient algorithms in reinforcement learning and beyond.

22 References

References

Amari, S.-i. (1998). Natural gradient works efficiently in learning. *Neural Computation*, *10*(2), 251–276. https://doi.org/10.1162/089976698300017746

Amari, S., & Nagaoka, H. (2000). *Methods of information geometry* (Vol. 191). American Mathematical Society. https://bookstore.ams.org/mmono-191

Berk, R. A. (2012). *Criminal justice forecasts of risk: A machine learning approach*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4614-3085-8

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer. https://www.microsoft.com/en-us/research/wp-content/uploads/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf

Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, *44*(4), 375–417. https://doi.org/10.1002/cpa.3160440402

Corman, H., Joyce, T., & Lovitch, N. (1987). Crime, deterrence, and the business cycle in new york city: A var approach. *The Review of Economics and Statistics*, *69*(4), 695–700. https://doi.org/10.2307/1935968

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems 26*, 2292–2300. https://papers.nips.cc/paper/4927-sinkhorn-distances-lightspeed-computation-of-optimal-transport

Gau, J. M. (2010). Basic principles and practices of structural equation modeling in criminal justice and criminology research. *Journal of Criminal Justice Education*, *21*(2), 136–151. https://doi.org/10.1080/10511251003693694

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 2672–2680.

Guo, W., Ho, N., & Jordan, M. I. (2020). Fast algorithms for computational optimal transport and wasserstein barycenter. *Proceedings of the 37th International Conference on Machine Learning*, *108*, 8036–8046. https://proceedings.mlr.press/v108/guo20a/guo20a.pdf

Hastie, T., Tibshirani, R., & Friedman, J. (2009a). *The elements of statistical learning: Data mining, inference, and prediction* (2nd). Springer. https://doi.org/10.1007/978-0-387-84858-7

Hastie, T., Tibshirani, R., & Friedman, J. (2009b). *The elements of statistical learning: Data mining, inference, and prediction* (2nd). Springer.

Konda, V. R., & Tsitsiklis, J. N. (2000). Actor-critic algorithms. *Advances in Neural Information Processing Systems*, *12*.

Li, G., Chen, Y., Huang, Y., Chi, Y., Poor, H. V., & Chen, Y. (2023). Fast computation of optimal transport via entropy-regularized extragradient methods. *arXiv preprint arXiv:2301.13006*. https://arxiv.org/abs/2301.13006

Mai, V. V., Lindbäck, J., & Johansson, M. (2021). A fast and accurate splitting method for optimal transport: Analysis and implementation. *arXiv preprint arXiv:2110.11738*. https://arxiv.org/abs/2110.11738

Martens, J. (2014). New insights and perspectives on the natural gradient method. *arXiv preprint arXiv:1412.1193*. https://arxiv.org/abs/1412.1193

Nielsen, F. (2020). An elementary introduction to information geometry. *Entropy*, *22*(10), 1100. https://doi.org/10.3390/e22101100

Otto, F. (2001). The geometry of dissipative evolution equations: The porous medium equation. *Communications in Partial Differential Equations*, *26*(1-2), 101–174. https://doi.org/10.1081/PDE-100002243

Pascanu, R., & Bengio, Y. (2013). Revisiting natural gradient for deep networks. *arXiv preprint arXiv:1301.3584*. https://arxiv.org/abs/1301.3584

Peyré, G., & Cuturi, M. (2019a). Computational optimal transport. *Foundations and Trends in Machine Learning*, *11*(5–6), 355–607. https://doi.org/10.1561/2200000073

Peyré, G., & Cuturi, M. (2019b). Computational optimal transport: With applications to data science. *Foundations and Trendső in Machine Learning*, *11*(5-6), 355–607. https://doi.org/10.1561/2200000073

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd). Sage. https://books.google.com/books?id=eE-CAAAAIAAJ

Rockafellar, R. T. (1970). *Convex analysis* (Vol. 28). Princeton University Press. https://press.princeton.edu/books/paperback/9780691015866/convex-analysis

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.

Tiwari, S., Gottesman, O., & Konidaris, G. (2022). On the geometry of reinforcement learning in continuous state and action spaces. *arXiv preprint arXiv:2301.00009*. https://arxiv.org/abs/2301.00009

Villani, C. (2008). *Optimal transport: Old and new* (Vol. 338). Springer-Verlag. https://doi.org/10.1007/978-3-540-71050-9

Weisburd, D., & Britt, C. (2009). *Statistics in criminal justice* (3rd). Springer Science & Business Media. https://doi.org/10.1007/978-0-387-89015-8

Zanette, A., Kochenderfer, M. J., & Brunskill, E. (2019). Almost horizon-free structure-aware best policy identification with a generative model. *arXiv preprint arXiv:1901.09018*. https://arxiv.org/abs/1901.09018