Issue 1

March 2025

# International Indian Statistical Association Newsletter

# Contents



Cover Photo: Group photograph at the IISA 2024 Conference, Kochi. December 2024.

## About Us

**Fundraising Committee**

Subrata Kundu
George Washington University

Srijan Sengupta
North Carolina State University

Debashis Mondal
Indian Institute of Technology, Kanpur

Soudeep Deb
Indian Institute of Management Bangalore

Hiya Banerjee
Eli Lilly and Co.

**Webinar Committee**

Hiya Banerjee
Eli Lilly and Co.

Himel Mallick
Cornell University

Kaushik Jana
ISI Delhi

Jyotishka Datta
Virginia Tech

Nilanjan Chakraborty
Missouri University of Science and Technology

Piyali Basak
Merck

## Message from President, IISA

Dear IISA Community, I would like to begin by expressing my heartfelt gratitude to each of you for being an integral part of the IISA community. It is an immense honor to serve as the President of IISA, and I am truly inspired by the enthusiasm and dedication of our members in advancing the field of statistics and data science worldwide.

As you all know, the IISA 2024 Annual Conference at the Cochin University of Science and Technology (CUSAT) was a resounding success. With over 550 participants from academia, government, and industry—hailing from India, the United States, and numerous other countries—the conference provided a vibrant platform for intellectual exchange. All presentations were conducted in person, and the event officially commenced on the afternoon of December 27, concluding at lunch on December 31.

This success would not have been possible without the generous sponsorship from Pfizer, Merck, the American Statistical Association, Johnson & Johnson, the ASA Biopharmaceutical Section, as well as funding from the U.S. National Science Foundation, the Indian Association for Statistics in Clinical Trials, and the National Board of Higher Mathematics of India. Their support helped keep registration fees relatively low, particularly for participants from India, and enabled travel support for many students and junior researchers. The conference also provided excellent networking opportunities, fostering meaningful professional and academic interactions. The local organizing committee at CUSAT played a crucial role in ensuring the event's smooth execution, and IISA extends its sincere appreciation for their efforts in making the conference a resounding success.

Looking ahead, our flagship event, the Annual Conference of IISA, will take place from June 12 to June 15, 2025, in Lincoln, Nebraska. The conference website is now live at https://www.intindstat.org/conference2025/index. This year's conference is co-organized with the Department of Statistics at the University of Nebraska–Lincoln (UNL), and IISA is deeply appreciative of their generous hospitality. The IISA Conference serves as a critical platform for fostering interactions between mathematical statisticians, applied statisticians, probabilists, and data scientists.

Diversity and student participation remain top priorities for IISA, and we actively seek participant travel grants and industry contributions to support these efforts. I extend my sincere thanks to Dr. Bodhisattva Sen, who is leading the Scientific Program Committee, along with Dr. Po-Ling Loh and Dr. Margaret Gamalo, for their dedication to curating a robust and engaging scientific program. This conference will showcase a wide range of topics spanning statistics, probability, and their applications. Furthermore, it serves as a platform for collaboration with International Chinese Statistical Association (ICSA), Korean International Statistical Society (KISS), and Caucus for Women (CWS).. I warmly invite all of you to join us in Lincoln, where we can share knowledge, build camaraderie, and create new opportunities together.

I would also like to express my gratitude to Dr. Cyrus Mehta for his generous contribution to fund our Early Career Award in Statistics and Data Science. We encourage our community to consider supporting IISA through donations, which will enable us to continue providing valuable opportunities for our members. If you are interested in contributing, please reach out to us.

Additionally, I am delighted to share that IISA has successfully revived its webinar series, which has been running since last year. I extend my gratitude to the Webinar Organizing Committee for their relentless efforts in curating relevant topics and distinguished speakers for our monthly sessions. It brings me great joy to witness global participation in these webinars, engaging with esteemed experts to whom we owe our sincere appreciation. You can watch the past presentations on our YouTube channel: https://www.youtube.com/playlist?list=PLo1imHbJHO1W3FBrgEeS7rSoMkXisjgdM.

IISA will also have a strong presence at the Joint Statistical Meetings (JSM) in Nashville. We are sponsoring three invited sessions and three topic-contributed sessions, with many more IISA members actively participating. IISA will also host a booth and organize a mixer session at JSM. If you plan to attend JSM, I encourage you to visit our booth, join the IISA Mixer, and explore opportunities to volunteer.

The Executive Committee and various subcommittees are working diligently to build a strong and vibrant IISA community - kudos to them for their efforts! I encourage all of you to consider volunteering in any capacity you can— Your support is essential for IISA's continued growth and success.

Your feedback is invaluable to us. Please feel free to share your thoughts and suggestions at president@intindstat.org as we strive to make IISA even better in the future.
Thank you!

— Hiya Banerjee
President, International Indian Statistical Association (IISA)
Director, Eli Lilly and Company

## Celebrating March – Women's History Month
## Interview with Dr. Madhumita Ghosh Dastidar

IISA had the pleasure to have a chat with Dr. Madhumita (Bonnie) Ghosh Dastidar, Head, RAND Statistics Group, and the 2024 President of the American Statistical Association. We thank Dr. Ghosh Dastidar for her time and for her inspiring words!

IISA **What progress have you seen in terms of gender diversity and inclusion in statistics and data science?**

Bonnie *I will first answer this from a personal perspective. After my PhD, I was hired by RAND, the oldest think tank in the U.S. The head of the group, at the time, was also the first female member of the RAND statistics group (I am the second). At present, I am serving as the past president of the American Statistical Association (ASA) next to the president and president-elect who are both women. This is a sign of progress. A key measure of gender diversity is the number of women who are in leadership positions within our professional associations and across the employment sectors. Yet under representation persist in certain areas of the profession, such as the proportion of women among tenured faculty.*

*To better understand these trends, the ASA has tabulated data from the Department of Education, National Center for Education Statistics (Integrated Postsecondary Education Data System Completions Survey) and NSF/NCSES Survey of Earned Doctorates. In 2023, women represented approximately 54% of the earned doctorates in biostatistics and 35% in statistics. (I should note that, given recent events, the availability of this data in the future is uncertain). These data suggest that women are actively pursuing degrees in statistics and data science, though gaps remain. Encouragingly, our field has a higher representation of women relative to STEM fields such as computer science. While the progress in our profession is meaningful, there is work to be done to ensure equitable representation across all levels of the profession.*

IISA **What initiatives or programs has ASA supported to empower women in these fields?**

Bonnie *Since 2016, the ASA has hosted the Women in Statistics and Data Science (WSDS) conference, which brings together professionals and students from academia, industry, and government to exchange ideas and share knowledge. Everyone is welcome to attend and participate in WSDS, which always features plenary presentations from prominent women leaders in our profession. The conference welcomes everyone but is designed to celebrate the advancement of women in statistics and data science. WSDS always features plenary presentations from prominent women leaders and fosters an environment that encourages participation from students and early-career professionals. One of its key goals is to build a strong community. Career challenges are inevitable, so having a supportive network is essential for navigating these challenges.*

*For the past several years, the ASA has celebrated Women's History Month every March by highlighting women who have made and continue to make contributions to the profession. By celebrating this diversity, we not only acknowledge the successes but also empower young women to find their unique path. Finally, the ASA Committee on Women in Statistics works to make members of ASA more aware of the common professional interests and issues facing women members of ASA. I had the honor of serving on this committee, where I co-authored two articles for Amstat News (the magazine of the ASA) to highlight the under representation of women in certain types of leadership roles such as journal editorial boards. The committee's efforts to promote women already in the statistics profession and to encourage more women to enter the field of statistics have had a profound impact.*

IISA **What advice would you give to young women considering a career in statistics or data science?**

Bonnie *The most important advice I can offer is to be strong, confident and advocate for yourself. Women in science are less likely than their male counterparts to receive authorship credit for the work they do, according to the findings of a study published in Nature. So, we need to be better at promoting ourselves, both within our organizations and in the profession. Also, women resist calling attention to their accomplishments when they work in groups with men, according to a series of studies in psychology. For too many women, the hardest part of being successful might be taking credit for the work that they do, especially when they work in groups. Let's be an ally for one another – encourage or compliment a friend or colleague, when you see exceptional work or scientific merit. That type of encouragement really helped build my self-confidence! Speaking of allies, it is important to find your community early. Connect with professional associations like the ASA and IISA. I cannot imagine my professional life without the mentors who have guided me. Mentors serve as role models and can offer a valuable sounding board for career decisions. Let's recognize our mentors' contribution – nominate them for the ASA Mentoring Award or simply express your gratitude.*

*We should also take every opportunity to speak up – be it on research teams, faculty meetings or in the board room. As statisticians and data scientists, we have the added challenge of communicating complex ideas to both technical and non-technical audiences. The ability to articulate ideas with clarity is a skill we can develop over time. This trait often sets successful data scientists apart and is a necessary skill to be an effective leader.*

*We are living in a rapidly evolving world, so I encourage everyone to stay adaptable. Our field also constantly evolves, so cultivate a growth mindset and continuously expand your skillset.*

*Remember that your path doesn't have to look like anyone else's. Each of your paths will be shaped by personal experiences, interests, and opportunities. STEM careers are not just about technical skills but also about curiosity, resilience, and the empowerment to make a real impact.*

IISA **How can IISA better support the advancement of women in these disciplines?**

Bonnie *By providing opportunities to support the advice I offered in the previous question. Professional associations must create a welcoming community that respects and celebrates the unique contributions of all our members.*

*Mentorship programs are essential at every career stage, but creating and sustaining successful programs can be challenging. I encourage professional societies to work together to ensure meaningful mentorship opportunities for women in statistics and data science.*

*We need to nominate more women to serve in high profile positions such as editorial boards and nominate them for scientific awards. Women are far too often recognized for service roles. Nomination committees within societies and organizations can help in addressing this imbalance.*

*Finally continue the good work you are doing to create conferences and activities that are welcoming and supportive.*

IISA **How can male allies contribute to fostering a more inclusive and equitable environment?**

Bonnie *All of us must commit to amplifying diverse perspectives in meetings and discussions. As we develop conference programs, we can be thoughtful and intentional about representation. In our professional environments, we should ensure that our colleagues' ideas and contributions are not overlooked. Those of us in leadership positions should sponsor our colleagues for high-visibility projects and provide opportunities for them to take on leadership roles. We can be supportive of equitable parental leave policies, flexible work arrangements, transparent hiring practices, and pay equity audits. Perhaps most important, we can all show up for events focused on inclusion because our presence signals the importance of these issues.*

IISA Thank you Bonnie for your time! We are sure our members would love reading your interview and learn from your wisdom.

Prof. Anil K. Bera, University of Illinois at Urbana-Champaign (UIUC), received a Lifetime Achievement Award in recognition of his extraordinary contribution to the field of Spatial Econometrics, Economics and Statistics, from the Indian Institute of Technology (IIT) - Indore.

Prof. Dipak Dey, University of Connecticut, received Lifetime Achievement Award from The Indian Bayesian Society and the International Advisory Committee of the International Conference on Recent Developments in the Techniques of Bayesian Paradigm, presented on January 6, 2025.

## History Corner: Statistics at IIT Bombay

by Prof. Sanjeev Sabnis, IIT Bombay

The four specializations in Computer Science, Statistics & OR, Pure Mathematics and Applied Mathematics were run by the Mathematics department at IIT Bombay from 1984-85 to 1994-95. In the subsequent academic year, a structural change took place. On one hand, it seemed appropriate to combine the specializations in Pure mathematics and in Applied mathematics to create a balanced and sound programme in Mathematics, and on the other hand, a strong need was felt to combine the specializations in Computer Science and in Statistics & OR to formulate an industry-oriented programme. Thus in place of the earlier four specializations, a so-called scholastic stream 'M.Sc. (Mathematics)' and a so-called utilitarian stream 'M.Sc. (Applied Statistics & Informatics)' were introduced. A great deal of flexibility was offered in the scholastic stream by making all courses in the 2nd, 3rd, and 4th semesters electives in Group A (Pure Mathematics), Group B (Applied Mathematics), and Group C (Statistics). Only the first semester courses Linear Algebra, Real Analysis I, Complex Analysis, Numerical Analysis I, and Differential Equations I were compulsory. The utilitarian stream was aimed at creating professional academicians who can manage and analyze large statistical data with tools of computer science.


Department of Mathematics, IIT Bombay

It had the following course structure. Semester I: Computer Programming & Utilization (CS), Discrete Structures (CS), Applied Linear Algebra (SI), Mathematical Modeling (SI), Mathematical Systems Theory (SI), Applied Probability & Statistics; Semester II: Data Structures & Algorithms (CS), Programming Lab (CS), Business Information Systems (CS), Numerical Methods (SI), Optimization (SI), Applied Stochastic Processes (SI); Semester III: Discrete Algorithms (SI), Finite Difference and Finite Element Methods (SI), Categorical Data Analysis & Regression (SI), Electives I & II; Semester IV: Stochastic Programming Applications (SI), Experimental Design (SI), Electives III & IV. In addition, a work visit of 4 weeks duration during the vacation months was a requirement. Over the years, some courses and labs offered by the Computer Science & Engineering department were replaced by redesigned courses and labs offered by the Mathematics Department. From 2024, the Department of Mathematics has started offering an M.Sc. in Statistics in lieu of an M.Sc. in Applied Statistics & Informatics. The Department of Mathematics also offers (i) a five-year BS in Mathematics with an M.Sc. in Statistics, (ii) an Inter Disciplinary Dual Degree Program (IDDDP) in Statistics for students already enrolled in a B.Tech. degree in various engineering departments across the institute, (iii) a Ph.D. degree.

IISA mourns the loss of Dr. Sayan Mukherjee, Humboldt Professor at Leipzig University and Fellow at the Max Planck Institute for Mathematics in the Sciences, who passed away on March 31, 2025 at the age of only 54. Dr. Mukherjee received the 2008 Young Researcher's Award from IISA. More can be found here: https://www.mis.mpg.de/news/loss-sayan-mukherjee

IISA mourns the loss of Dr. Riten Mitra, Associate Professor in Bioinformatics and Biostatistics at the University of Louisville, who passed away on March 13, 2025. More can be found here: https://louisville.edu/sphis/news/passing-of-dr-riten-mitra

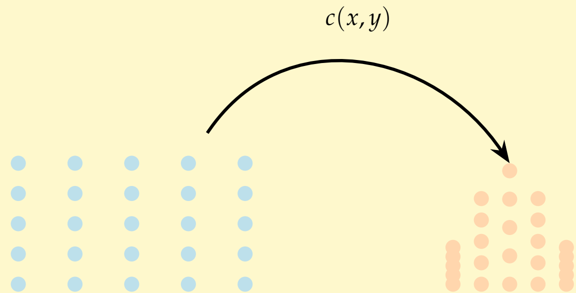
by Ananyapam De, TU Clausthal, Germany

# A Journey in Optimal Transport

## Introduction

The origins of optimal transport trace back to the late 18th century. In 1776, while presenting his memoiré at the French Académie Royale des Sciences, Gaspard Monge posed an intriguing problem:

> *Given two excavation sites $D_1$ and $D_2$ and a cost function that assigns a price for moving a unit mass from any point in $D_1$ to any point in $D_2$, what is the most cost-effective method to transfer mass from $D_1$ to $D_2$?*




Visualization of Monge's transport problem: transforming a flat distribution of mass (left) into a bell distribution (right) while minimizing the transport cost $c(x, y)$.

Although Monge did not resolve this challenge in his memoiré, his formulation laid the essential groundwork for modern optimal transport theory.

In this setting, the excavations are modeled as probability measures, $\rho_0$ and $\rho_1$, on a domain $\Omega \subset \mathbb{R}^d$, and the cost of moving a unit mass from $x$ to $y$ is given by a function $c(x, y)$. Monge's challenge was to find a transport map $T : \Omega \to \Omega$, satisfying the push-forward condition $T_\# \rho_0 = \rho_1$, and minimizing the total cost

$$\int_\Omega c(x, T(x))\, d\rho(x).$$

The easiest way to think of the problem is in terms of atomic items. In its simplest discrete form, optimal transport can be viewed as an assignment problem: among all possible configurations, which one is the best? This formulation is already quite restrictive, as it requires working with two sets of the same total size. However, despite its simplicity, this version of the problem is incredibly hard to solve.

Each set can be represented as a vector **a** that belongs to the probability simplex—the set of vectors whose components sum to 1:

$$\mathbf{a} \in \left\{ x = (x_1, \ldots, x_N) \in \mathbb{R}^N : \sum_{i=1}^{N} x_i = 1 \right\}. \tag{1}$$

If we denote by $\mathbf{C}_{i,j}$ the cost of moving an element from $i$ to $j$, the quantity we aim to minimize is $\sum_i \mathbf{C}_{i,\sigma(i)}$, where $\sigma$ is a permutation of the set $\{1, \ldots, N\}$.

In this form, optimal transport is fundamentally **combinatorial** and can be summarized as follows:

> *How can we assign every element $i$ of $\{1, \ldots, N\}$ to elements $j$ of $\{1, \ldots, N\}$ so as to minimize the cumulative cost of this assignment?*

The result of this search is called the **optimal assignment**. Since there are exactly $N!$ possible solutions to this problem, it extremely difficult to solve for large $N$. Note that this assignment is **not unique**, as illustrated below, where two elements are mapped to two others that together form the four corners of a 2D square with side length 1.

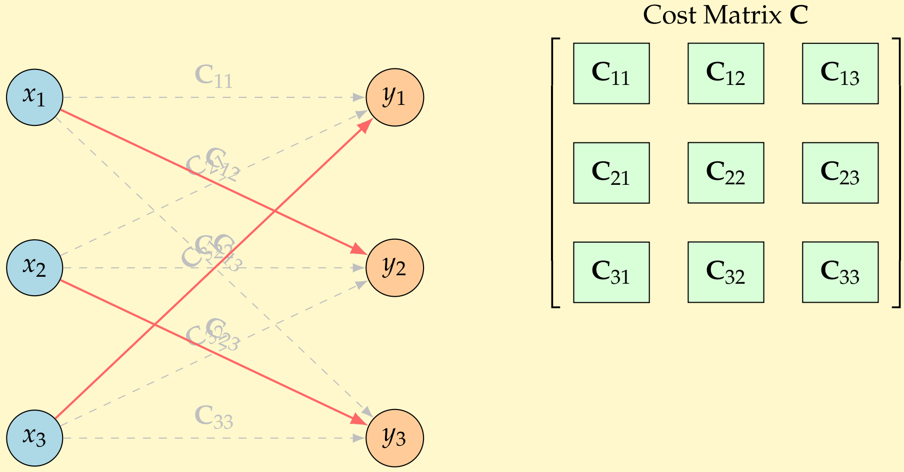



Illustration of the discrete optimal transport problem as an assignment task. Each element $x_i$ from the source set must be assigned to exactly one element $y_j$ from the target set while minimizing the total transport cost $\sum_i \mathbf{C}_{i,\sigma(i)}$. The dashed gray lines represent all possible assignments, while the red paths highlight an optimal assignment. The corresponding cost matrix $\mathbf{C}$ encodes the transport costs between pairs. This problem is computationally challenging. $N = 15$ yields 1.3076744e+12 possible combinations!.

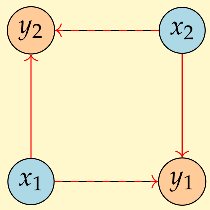



Illustration of non-unique optimal assignments within a single square. The pastel blue nodes represent $x_1$ and $x_2$, while the pastel orange nodes represent $y_1$ and $y_2$. The dashed and solid arrows illustrate different valid assignments.

## Monge Problem

A major limitation of using two equally sized histograms or piles is that it restricts the flexibility of the mapping. By generalizing the above definition to a broader class of histograms, we arrive at the Monge problem, where multiple points $x_i$ can map to the same $y_j$, with their weights being summed.

In this case, the mapping between inputs and outputs is no longer a permutation but rather a surjective map $T$. If the points $x_1, \ldots, x_n$ have weights $\mathbf{a} = (a_1, \ldots, a_n)$ and the points $y_1, \ldots, y_m$ have weights $\mathbf{b} = (b_1, \ldots, b_m)$, then $T$ must satisfy:

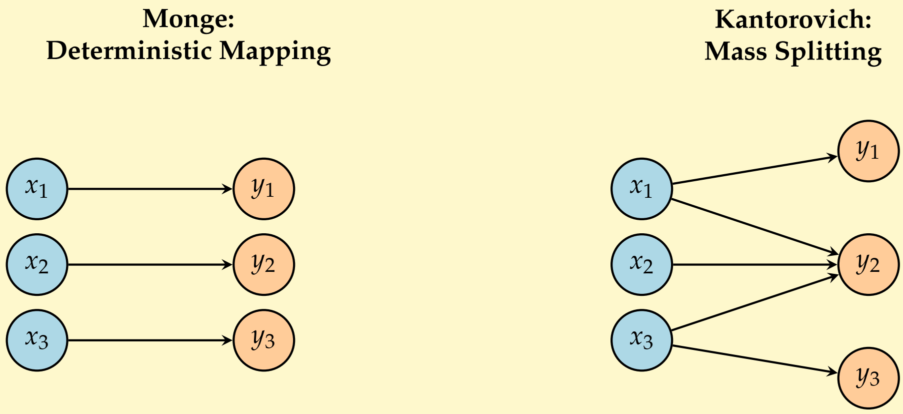
$$\forall j \in \{1, \ldots, m\}, \quad b_j = \sum_{i:T(x_i)=y_j} a_i. \tag{2}$$

With this formulation, the problem is only well-posed if the mass conservation constraint is satisfied. However, it remains challenging to solve because we are still assigning elements $x_i$ to elements $y_j$.

## Kantorovitch Relaxation

While Monge's formulation is elegant, it is encumbered by significant limitations. Even after extending the original problem, requiring maps $T$ with $T_{\#}\rho_0 = \rho_1$ results in a non-convex set, and an optimal transport map may not exist or be unique. This challenge inspired Leonid Kantorovich to introduce a relaxation that revolutionized optimal transport theory.

Kantorovich's key insight was to relax the deterministic requirement of the mapping. Instead of forcing each source point $x_i$ to map to a single target, mass may be **split** and distributed among several targets. This idea leads to the concept of *couplings*: joint probability measures $\gamma$ on $\Omega \times \Omega$ whose marginals are $\rho_0$ and $\rho_1$. That is, for every




Comparison between Monge's deterministic transport map (left) and Kantorovich's relaxed formulation allowing mass splitting (right). In Monge's setting, each point $x_i$ is mapped to a unique $y_j$, whereas Kantorovich's formulation permits distributing mass from a single $x_i$ across multiple $y_j$, leading to a convex optimization problem.

measurable set $A \subset \Omega$,

$$\gamma(A \times \Omega) = \rho_0(A) \quad \text{and} \quad \gamma(\Omega \times A) = \rho_1(A).$$

This approach of mass splitting, where a single point $x$ can send parts of its mass to several different points $y$ transforms the problem into a convex one.

The optimal transport problem is then reformulated as

$$\inf_{\gamma \in \Pi(\rho_0, \rho_1)} \int_{\Omega \times \Omega} c(x, y)\, d\gamma(x, y),$$

where $\Pi(\rho_0, \rho_1)$ is the set of all couplings with the prescribed marginals. This reformulation not only ensures the existence of an optimal solution under very general conditions but also leads to a dual formulation that introduces the so-called *Kantorovich potentials*.

The dual formulation is particularly elegant. Instead of directly minimizing over couplings, one maximizes over a pair of functions (potentials) $\varphi$ and $\psi$, defined on $\Omega$, subject to the constraint

$$\varphi(x) + \psi(y) \leq c(x, y) \quad \text{for all } x, y \in \Omega.$$

The dual problem can then be written as

$$\sup_{\varphi, \psi} \left\{ \int_\Omega \varphi\, d\rho_0 + \int_\Omega \psi\, d\rho_1 \ : \ \varphi(x) + \psi(y) \leq c(x, y) \text{ for all } x, y \in \Omega \right\}.$$

Under appropriate conditions—for instance, when the cost is quadratic and $\rho_0$ is absolutely continuous—the optimal potential $\varphi$ is convex (up to an additive constant) and its gradient yields the unique optimal transport map.

In the discrete setting, Kantorovich's relaxation replaces the permutation $\sigma$ with a coupling matrix

$$\mathbf{P} = (\mathbf{P}_{ij}) \in \mathbb{R}_+^{n \times m},$$

where each entry $\mathbf{P}_{ij}$ represents the fraction of mass moved from source $i$ to destination $j$. The set of all admissible transport plans is then given by

$$\mathbf{U}(\mathbf{a}, \mathbf{b}) = \left\{ \mathbf{P} \in \mathbb{R}_+^{n \times m} : \mathbf{P}\, \mathbb{1}_m = \mathbf{a}, \quad \mathbf{P}^T\, \mathbb{1}_n = \mathbf{b} \right\},$$

which enforces that the total mass leaving each source matches $\mathbf{a}$ and the mass arriving at each destination matches $\mathbf{b}$.

An attractive feature of this formulation is its inherent symmetry. If $\mathbf{P} \in \mathbf{U}(\mathbf{a}, \mathbf{b})$ describes a transport plan from $\mathbf{a}$ to $\mathbf{b}$, then its transpose $\mathbf{P}^T$ naturally belongs to $\mathbf{U}(\mathbf{b}, \mathbf{a})$, effectively reversing the roles of sources and targets.

The optimal transport cost is expressed by the linear program

$$L_{\mathbf{C}}(\mathbf{a},\mathbf{b}) = \min_{\mathbf{P}\in\mathbf{U}(\mathbf{b},\mathbf{a})} \langle \mathbf{C},\mathbf{P}\rangle = \min_{\mathbf{P}\in\mathbf{U}(\mathbf{b},\mathbf{a})} \sum_{i,j} \mathbf{C}_{ij}\,\mathbf{P}_{ij},$$

where the cost matrix $\mathbf{C}$ quantifies the expense of transporting mass from source $i$ to destination $j$. Since the constraints define a convex polytope via $m+n$ equality constraints (or equivalently $2(m+n)$ inequalities), this reformulation transforms the problem from one of combinatorial optimization into a convex optimization problem which are much simpler to deal with.

In addition to the primal formulation, the Kantorovich relaxation admits an elegant dual formulation involving what are known as *Kantorovich potentials*. In the discrete setting, we introduce dual variables $\varphi_i$ for each source and $\psi_j$ for each destination. The dual problem is then formulated as

$$\max_{\varphi\in\mathbb{R}^n,\,\psi\in\mathbb{R}^m} \left\{ \sum_{i=1}^{n} \varphi_i\, a_i + \sum_{j=1}^{m} \psi_j\, b_j \;:\; \varphi_i + \psi_j \le \mathbf{C}_{ij} \quad \forall\, i,j \right\}.$$

Thus, the dual formulation not only enhances our theoretical understanding of the transport problem but also serves as a foundation for numerous efficient computational algorithms like simplex algorithm, dual ascent methods, and the Auction algorithm.

## Benamou-Brenier Formulation

Having developed a solid intuition through the discrete setting, we now shift our focus to continuous probability densities—a transition that naturally reveals a rich geometric framework. In this context, optimal transport is built around a cost function $c(x,y)$ that quantifies the "expense" of moving a unit of mass from point $x$ to point $y$. Although the abstract theory allows for a variety of cost functions, much of the intuitive and geometric insight emerges when we choose

$$c(x,y) = \|x-y\|^2,$$

the squared Euclidean distance between $x$ and $y$ in $\mathbb{R}^d$.

This particular choice is not only mathematically natural but also leads us directly to one of the most important objects in the theory: the 2-Wasserstein metric $W_2$. Specifically, when the cost is taken to be the squared Euclidean distance, one defines the metric $W_2$ on the space of probability measures with finite second moments, denoted by

$$\mathcal{P}_2(\Omega) = \left\{ \rho : \quad \int_{\Omega} \|x\|^2\, d\rho(x) < \infty \right\}.$$

The metric space $(\mathcal{P}_2(\Omega), W_2)$ then captures a natural notion of "distance" between probability measures, reflecting how one measure can be morphed into another with minimal transportation cost.

A transformative breakthrough in optimal transport theory was achieved by Benamou and Brenier in 1991, who recast the problem from a static matching between source and target measures into a dynamic fluid-mechanical formulation. There are two complementary perspectives on fluid flows: the Lagrangian perspective which emphasizes particle trajectories, and the Eulerian perspective which tracks the evolution of the fluid density. Suppose that $X_0 \sim \rho_0$ and that $t \mapsto X_t$ evolves according to the ODE

$$\dot{X}_t = v_t(X_t).$$

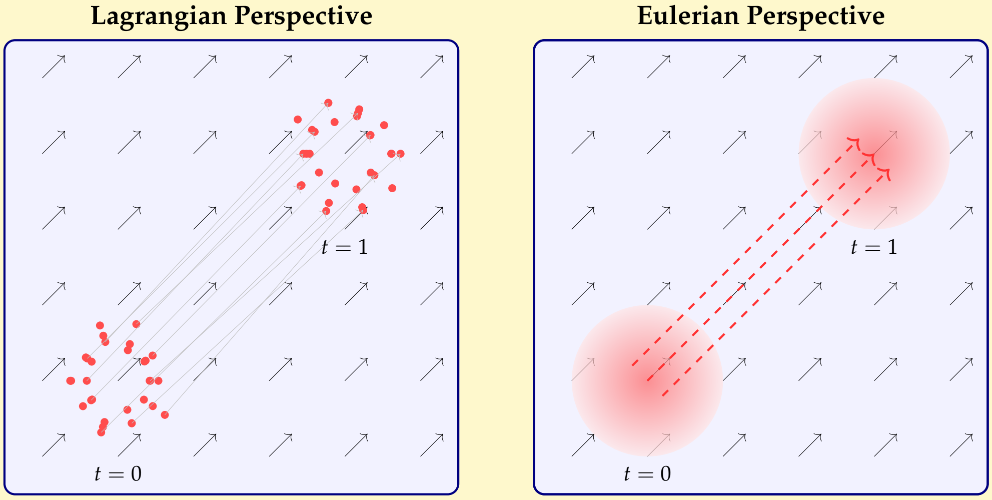
Here, $(v_t)_{t\ge 0}$ is a family of vector fields, i.e., mappings $\mathbb{R}^d \to \mathbb{R}^d$. Since the ODE describes the evolution of the particle trajectory, it is the Lagrangian description of the dynamics. The corresponding Eulerian description is the continuity equation.

**Theorem.** Let $t \mapsto v_t$ be a family of vector fields and suppose that the random variables $t \mapsto X_t$ evolve according to $\dot{X}_t = v_t(X_t)$. Then, the law $\rho_t$ of $X_t$ evolves according to the continuity equation

$$\partial_t \rho_t + \operatorname{div}(\rho_t v_t) = 0.$$

The upshot of this theorem is that every nice curve of measures $t \mapsto \rho_t$ can be interpreted as the fluid flow along a family of vector fields, i.e., we can find vector fields $t \mapsto v_t$ such that the continuity equation holds.

However, note that this is not unique: if $\partial_t \rho_t + \operatorname{div}(\rho_t v_t) = 0$ and for each $t$, $w_t$ is a vector field satisfying $\operatorname{div}(\rho_t w_t) = 0$, then the continuity equation also holds with the new vector fields $\tilde{v}_t = v_t + w_t$.




The Benamou-Brenier formulation of optimal transport reframes the problem as a dynamic flow of probability densities. The Lagrangian perspective (left) tracks individual particle trajectories $X_t$ evolving under a velocity field $v_t(X_t)$, while the Eulerian perspective (right) describes the evolution of the probability density $\rho_t$ via the continuity equation $\partial_t \rho_t + \operatorname{div}(\rho_t v_t) = 0$. The two views are mathematically equivalent, linking optimal transport to fluid dynamics.

We will show how to pick a distinguished choice of vector fields $t \mapsto v_t$ which can be described in two equivalent ways. First, among all vector fields making the continuity equation hold true, we can choose $v_t$ to minimize $\int \|v_t\|^2 d\rho_t$, which has the physical interpretation of minimizing kinetic energy. Second, we can choose $v_t$ to be the gradient of a function; we provide some intuition to see why gradient vector fields should be optimal.

Since the continuity equation is equivalent to requiring that for all test functions $\varphi : \mathbb{R}^d \to \mathbb{R}$, it holds that

$$\partial_t \int \varphi \, d\rho_t = \int \langle \nabla \varphi, v_t \rangle \, d\rho_t.$$

In this expression, the vector field $v_t$ only enters through inner products with gradients. Equivalently, if we define

$$S := \{\nabla \psi \mid \psi : \mathbb{R}^d \to \mathbb{R}\}$$

as a subspace of $L^2(\rho_t)$, then we can decompose

$$L^2(\rho_t) = S \oplus S^\perp$$

(with the understanding that one should technically take the closure of $S$, but we will ignore this detail). Decomposing $v_t$ as this direct sum,

$$v_t = \nabla \psi_t + w_t$$

with$w_t$ orthogonal to $S$ in $L^2(\rho_t)$, we see that replacing $v_t$ by $\nabla \psi_t$, preserves the continuity equation while potentially reducing the norm $\|v_t\|_{L^2(\rho_t)}$. This reasoning supports the optimality of choosing $v_t$ as a gradient.

We interpret $v_t$ as the velocity vector along a curve $\rho_t$ in the space $\mathcal{P}_{2,ac}(\mathbb{R}^d)$ of absolutely continuous probability measures in a Riemannian framework. Here, the tangent space at $\rho$ is defined as

$$T_\rho \mathcal{P}_{2,ac}(\mathbb{R}^d) := \overline{\{\nabla \psi : \psi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\rho)},$$

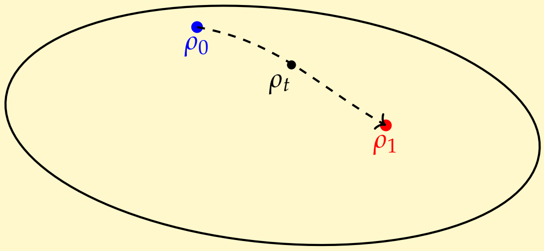
with the norm $\|v\|_{L^2(\rho)}$ given by the inner product induced by the $L^2(\rho)$ metric. If the continuity equation

$$\partial_t \rho_t + \operatorname{div}(\rho_t v_t) = 0$$

holds with $v_t \in T_{\rho_t} \mathcal{P}_{2,ac}(\mathbb{R}^d)$, then $v_t$ is the tangent vector to the curve. This $L^2(\rho)$ metric endows $\mathcal{P}_{2,ac}(\mathbb{R}^d)$ with a Riemannian-like structure that is compatible with the 2-Wasserstein distance $W_2$.

Let $\rho_0, \rho_1 \in \mathcal{P}_{2,ac}(\mathbb{R}^d)$. Then,

$$W_2(\rho_0, \rho_1) = \inf \left\{ \int_0^1 \|v_t\|_{L^2(\rho_t)} \, dt : \partial_t \rho_t + \operatorname{div}(\rho_t v_t) = 0 \right\}.$$




**Wasserstein Space** $\mathcal{P}_2(\mathbb{R}^d)$

This reveals that optimal transport paths can be viewed as geodesics in the space $\mathcal{P}_2(\Omega)$. The infimum is attained as follows. Let $X_0 \sim \rho_0$ and $X_1 \sim \rho_1$ be optimally coupled random variables, and define $X_t = (1-t)X_0 + tX_1$. Let $\rho_t$ denote the law of $X_t$. Then, the map $t \mapsto \rho_t$ represents the unique constant-speed geodesic connecting $\rho_0$ and $\rho_1$. Quite beautiful indeed! What began as a simple question about transporting soil has currently grown into a profound theory intertwining analysis, geometry, and probability. This progression not only deepens theoretical understanding but also drives the development of practical algorithms in machine learning, statistics, and applied mathematics, marking it as an ever-expanding and vibrant area of research.

## Aren't you a member of IISA yet?

# Join Us!

## Become a Member of our Exclusive Community

**Are you ready to be a part of something extraordinary?**

We are thrilled to announce that membership is now open for IISA! As a member, you'll have access to a wide range of benefits and opportunities. Join us and be a part of the extraordinary Statistics community!

Student membership is FREE!

Whether you are a professional looking for networking opportunities, an enthusiast passionate about Statistics, or someone seeking personal growth and development, our community welcomes you with open arms.

## Don't miss out on this incredible chance to make a difference!

Visit our website at www.intindstat.org to learn more about our organization, the membership benefits, and how to join. Together, let's create a brighter future!

## Join now and be part of something extraordinary!