

Matching and Optimal Transport

Regularización cuadrática heterogénea y congestión
Defensa de Tesis

Marcelo Gallardo

Asesor: Prof. Jorge Chávez

Heterogeneous quadratic regularization and congestion in optimal transport,
R&R, DEF (Springer)

Jurado: Johel Beltrán & Marcelo Flamarion

Departamento de Matemáticas – PUCP

Lima, 2026

- 1 Introducción
- 2 Regularización cuadrática heterogénea
- 3 El modelo
- 4 Propiedades del problema de optimización
- 5 Algoritmo
- 6 Estática comparativa
- 7 Aplicaciones
- 8 Extensiones
- 9 Referencias

Introducción

- Matching [Gale and Shapley, 1962, Roth and Sotomayor, 1990, Roth, 1982, Abdulkadiroğlu and Sönmez, 2003, Echenique and Yenmez, 2015, Echenique et al., 2023]
- Transporte óptimo [Villani, 2009, Galichon, 2016, Galichon, 2021, Dupuy and Galichon, 2022, Carlier et al., 2020].

- Dos grupos: X e Y .
- Matching/asignación π .
- Caso discreto $X = \{x_1, \dots, x_N\}$, $Y = \{y_1, \dots, y_L\}$, $I = \{1, \dots, N\}$ y $J = \{1, \dots, L\}$ los índices.

$$\min_{\pi \in \Pi(\mu, \nu)} \sum_{i=1}^N \sum_{j=1}^L c_{ij} \pi_{ij}$$

con

$$\Pi(\mu, \nu) = \left\{ \pi_{ij} \geq 0 : \sum_{j=1}^L \pi_{ij} = \mu_i \quad \forall i \in I, \quad \sum_{i=1}^N \pi_{ij} = \nu_j \quad \forall j \in J \right\}. \quad (1)$$

$$\mathcal{P}_\varepsilon : \quad \min_{\pi \in \Pi(\mu, \nu)} \sum_{i=1}^N \sum_{j=1}^L c_{ij} \pi_{ij} - \varepsilon H(\pi), \quad (2)$$

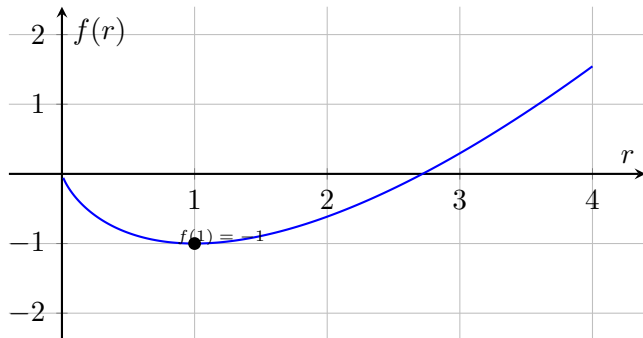
donde

$$H(\pi) = - \sum_{i,j} h(\pi_{i,j}) \quad (3)$$

$$h(r) = \begin{cases} r(\ln(r) - 1), & \text{if } r > 0 \\ 0, & \text{if } r = 0 \\ +\infty, & \text{if } r < 0. \end{cases}$$

π^* más densa.

Gráfica de $f(r) = r \ln r - r$



Modelo cuadrático generalizado discreto

Regularización cuadrática heterogénea

Sobre el politopo de transporte $\Pi(\mu, \nu)$ de la ec. (1):

$$\mathcal{P}_1 : \quad \min_{\pi \in \Pi(\mu, \nu)} \sum_{i=1}^N \sum_{j=1}^L d_{ij} + c_{ij} \pi_{ij} + a_{ij} \pi_{ij}^2. \quad (4)$$

- $d_{ij} \geq 0$: costo fijo del emparejamiento (base administrativa o distancia).
- $c_{ij} > 0$: costo marginal constante (preferencias del par (i, j)).
- $a_{ij} > 0$: **congestión heterogénea** $\Rightarrow$ costo marginal creciente en π_{ij} .
- Marginales rígidas: $\pi \in \Pi(\mu, \nu)$, con $\pi_{ij} \in \mathbb{R}_+$ (no $\mathbb{Z}_+$).

QP convexo separable con restricciones lineales. Generaliza la regularización cuadrática $\mathcal{P}_Q$ (que impone $a_{ij} = \varepsilon$ homogéneo).

- El término $\sum_{i,j} a_{ij} \pi_{ij}^2$ penaliza la **saturación**: mientras más masa recibe un centro, más costoso se vuelve (convexidad estricta).
- La **heterogeneidad** a_{ij} captura fricciones par-específicas (acceso desigual, cuellos de botella) que $\mathcal{P}_Q$ no puede reflejar.
- Al aumentar a_{ij} , la masa se redistribuye fuera de los pares congestionados: explica por qué el mejor candidato no siempre llega a su opción ideal.
- Convexo $\Rightarrow$ tratable con solvers estándar de optimización cuadrática.

Supuesto

$N = L = M$ y $\mu_i = \nu_j$ (igual cardinalidad).

Supuesto

Cada grupo i tiene opción preferida única ζ_i , con $\zeta_i \neq \zeta_k$ (preferencias alineadas).

Supuesto

$\tilde{c}_i := \min_{j \neq \zeta_i} c_{ij} > c_{i\zeta_i} + a_{i\zeta_i} \mu_i^2 (1 - 1/M)$ (separación).

Supuesto

$a_{i\zeta_i} \leq a_{ij}$ para todo j (mínima congestión en ζ_i).

Teorema

Bajo los Supuestos 2.1–2.4, el óptimo entero de $\mathcal{P}_1$ es

$$\pi_{ij}^* = \begin{cases} \mu_i & \text{si } j = \zeta_i, \\ 0 & \text{en otro caso,} \end{cases} \quad \pi^* \in \Pi(\mu, \nu).$$

Con preferencias alineadas y separación suficiente, la congestión no distorsiona el emparejamiento: cada grupo va a su opción preferida.

Limitación de $\mathcal{P}_1$: marginales rígidas $\Rightarrow$ sistema singular, sin resolución analítica sistemática.

Idea

Relajar $\sum_j \pi_{ij} = \mu_i$ y $\sum_i \pi_{ij} = \nu_j$ como penalizaciones cuadráticas:

$$\epsilon_i \left(\sum_j \pi_{ij} - \mu_i \right)^2, \quad \delta_j \left(\sum_i \pi_{ij} - \nu_j \right)^2,$$

combinadas con la congestión vía $\alpha \in (0, 1)$.

- μ, ν : de masas fijas a objetivos del planificador (desviaciones permitidas y penalizadas).
- F estrictamente convexa y coerciva sobre $\mathbb{R}_+^{NL}$ (problema irrestricto).
- $\nabla F = 0 \Rightarrow A\pi^* = b$ con A invertible: serie de Neumann y algoritmo Sherman–Morrison $O((N+L)(NL)^2)$.

Esto conduce al modelo $\mathcal{P}_{CP}$.

Modelo cuadrático con penalización

El modelo

- Penalización: [Izmailov and Solodov, 2023].
- Congestión: término cuadrático.

$$\begin{aligned} \mathcal{P}_{CP}: \min_{\pi_{ij} \geq 0} & \underbrace{\alpha \sum_{i=1}^N \sum_{j=1}^L \varphi(\pi_{ij}; \theta_{ij})}_{\text{Costo directo del matching}} + \\ & \underbrace{(1 - \alpha) \left[\sum_{i=1}^N \epsilon_i \left(\sum_{j=1}^L \pi_{ij} - \mu_i \right)^2 + \sum_{j=1}^L \delta_j \left(\sum_{i=1}^N \pi_{ij} - \nu_j \right)^2 \right]}_{\text{Costos de los objetivos sociales}} \\ & =: F(\pi; \theta, \alpha, \epsilon, \delta, \mu, \nu) \end{aligned} \tag{5}$$

donde $\epsilon_1, \dots, \epsilon_N$, $\delta_1, \dots, \delta_L$ y $\mu_1, \dots, \mu_N$, $\nu_1, \dots, \nu_L$ son todos no negativos, y

$$\varphi(\pi_{ij}; \theta_{ij}) = d_{ij} + c_{ij}\pi_{ij} + a_{ij}\pi_{ij}^2. \tag{6}$$

- $\sum_{\substack{1 \leq i \leq N \\ 1 \leq j \leq L}} c_{ij} \pi_{ij}$
- $\sum_{\substack{1 \leq i \leq N \\ 1 \leq j \leq L}} a_{ij} \pi_{ij}^2$
- $\mu_1, \dots, \mu_N$
- $\nu_1, \dots, \nu_L$
- $\epsilon_1, \dots, \epsilon_N$
- $\delta_1, \dots, \delta_L$
- $\alpha \in (0, 1), \alpha \rightarrow 1/2$

$$\sum_{j=1}^L \pi_{ij} = \mu_i \quad \Rightarrow \quad \epsilon_i \left[\sum_{j=1}^L \pi_{ij} - \mu_i \right]^2, \quad \epsilon_i > 0,$$
$$\sum_{i=1}^N \pi_{ij} = \nu_j \quad \Rightarrow \quad \delta_j \left[\sum_{i=1}^N \pi_{ij} - \nu_j \right]^2, \quad \delta_j > 0.$$

- Excesos de demanda u oferta.
- Mayor flexibilidad: α, ϵ, δ .
- Más sencillez en la resolución (CP0).

Nota: voy a tratar de manera indistinta $\mathcal{M}_{N \times L}$ y $\mathbb{R}^{NL}$ usando el isomorfismo (vectorizando por filas no columnas).

Propiedades del problema

Propiedades del problema de optimización

Definición

$F : \mathbb{R}^n \rightarrow \mathbb{R}$ es **coerciva** si $F(x) \rightarrow +\infty$ cuando $\|x\| \rightarrow \infty$. Su **conjunto de subnivel** de altura α es $\{x : F(x) \leq \alpha\}$.

- Si F es continua y coerciva, cada conjunto de subnivel es **cerrado y acotado**, luego **compacto**.
- Una función continua sobre un compacto no vacío **alcanza su mínimo** (Weierstrass).
- Por eso, coercividad + continuidad garantizan **existencia** de minimizador, aun cuando el dominio no sea acotado.

- **Existencia:** F es convexa, continua y coerciva sobre el cerrado convexo $\mathbb{R}_+^{NL}$. La coercividad hace compactos los conjuntos de subnivel, así que el ínfimo se alcanza (Weierstrass).
- **Unicidad:** consecuencia de la convexidad estricta de F (suma de $\sum_{i,j} \varphi(\pi_{ij}; \theta_{ij})$ estrictamente convexa y $N+L$ penalizaciones convexas).
- **Optimización sobre $\mathbb{R}_+^{NL}$** (métodos de redondeo):
 - Para cuadráticas separables con restricciones lineales existen cotas de proximidad entre soluciones enteras y continuas [Hochbaum and Shanthikumar, 1990, Park and Boyd, 2017, ?].
 - La desviación se controla vía el número de condición del Hessiano H :

$$\|\pi_{\mathbb{Z}}^* - \pi^*\|_{\infty} \leq O(\vartheta(H)), \quad \vartheta(H) = \frac{\lambda_{\max}(H)}{\lambda_{\min}(H)}.$$

- Ajustando los parámetros se controla esta cota.

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \pi_{ij}} = \frac{1}{2} & \left(\varphi'(\pi_{ij}^*; \theta_{ij}) + 2\epsilon_i \left(\sum_{\ell=1}^L \pi_{i\ell}^* - \mu_i \right) \right. \\ & \left. + 2\delta_j \left(\sum_{k=1}^N \pi_{kj}^* - \nu_j \right) \right) - \gamma_{ij}^* = 0, \quad \forall (i, j) \in I \times J. \end{aligned}$$

- γ_{ij} : multiplicador asociado a la condición de no negatividad.
- En el caso $\gamma_{ij}^* = 0$, se cumple que $\nabla F(\pi^*) = 0$.
- Esto implica $A\pi^* = b$, donde:

$$\begin{aligned} A &= \underbrace{\text{Diag}(a_{11}, a_{12}, \dots, a_{NL})}_D + \underbrace{\text{Diag}(\epsilon_1, \dots, \epsilon_N) \otimes \mathbf{1}_{L \times L}}_E \\ &\quad + \underbrace{\mathbf{1}_{N \times N} \otimes \text{Diag}(\delta_1, \dots, \delta_L)}_F, \\ b &= \left[\epsilon_1 \mu_1 + \delta_1 \nu_1 - \frac{c_{11}}{2}, \dots, \epsilon_N \mu_N + \delta_L \nu_L - \frac{c_{NL}}{2} \right]^T. \end{aligned}$$

$$\begin{aligned}
D &= \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{12} & \cdots & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & a_{NL} \end{bmatrix} \\
E &= \begin{bmatrix} \epsilon_1 & 0 & \cdots & 0 \\ 0 & \epsilon_2 & \cdots & \vdots \\ \vdots & & \ddots & \\ 0 & & \cdots & \epsilon_N \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & \vdots \\ \vdots & & \ddots & \\ 1 & \cdots & & 1 \end{bmatrix} \\
F &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & \vdots \\ \vdots & & \ddots & \\ 1 & \cdots & & 1 \end{bmatrix} \otimes \begin{bmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & \vdots \\ \vdots & & \ddots & \\ 0 & & \cdots & \delta_L \end{bmatrix}.
\end{aligned}$$

Ejemplo $N = L = 2$

$$A = \begin{bmatrix} a_{11} + \epsilon_1 + \delta_1 & \epsilon_1 & \delta_1 & 0 \\ \epsilon_1 & a_{12} + \epsilon_1 + \delta_2 & 0 & \delta_2 \\ \delta_1 & 0 & a_{21} + \epsilon_2 + \delta_1 & \epsilon_2 \\ 0 & \delta_2 & \epsilon_2 & a_{22} + \epsilon_2 + \delta_2 \end{bmatrix}$$

Lema

El determinante de A es estrictamente positivo si todos los parámetros son estrictamente positivos.

- $D \succ 0$: $x^\top Dx = \sum_{i,j} a_{ij} x_{ij}^2 > 0$ para todo $x \neq 0$ (pues $a_{ij} > 0$).
- $E \succeq 0$: $x^\top Ex = \sum_{i=1}^N \epsilon_i \left(\sum_{j=1}^L x_{ij} \right)^2 \geq 0$.
- $F \succeq 0$: $x^\top Fx = \sum_{j=1}^L \delta_j \left(\sum_{i=1}^N x_{ij} \right)^2 \geq 0$.
- Entonces, para todo $x \neq 0$,

$$x^\top Ax = x^\top Dx + x^\top Ex + x^\top Fx \geq x^\top Dx > 0,$$

de modo que $A \succ 0$. En particular $\det(A) > 0$ y A es invertible.

Resultados

- Se puede.
- Fórmula cerrada: complicado (Wolfram 3×3).
- Aproximación de Neumann.
- Sherman-Morrison: Algoritmo para invertir.

Supuesto

Supóngase que $a_{ij} > 0$ y que:

$$\max_i \{\epsilon_i\} \cdot L + \max_j \{\delta_j\} \cdot N < \min_{(i,j)} \{a_{ij}\}.$$

- Esta condición garantiza que $\|D^{-1}X\|_\sigma < 1$ (norma espectral).
- Bajo esta condición:

$$A^{-1} = \left(\sum_{k=0}^{\infty} (-1)^k (D^{-1}X)^k \right) D^{-1}.$$

- Se obtiene convergencia del método iterativo:

$$\pi_n \rightarrow \pi^* = A^{-1}b.$$

- $A = D + X$ con $X = E + F$; como D es invertible,

$$A = D(I + D^{-1}X) \Rightarrow A^{-1} = (I + D^{-1}X)^{-1}D^{-1}.$$

- Basta probar $\|D^{-1}X\|_{\sigma} < 1$. Con $a_{\min} = \min_{(i,j)} a_{ij} > 0$,

$$\|D^{-1}\|_{\sigma} = \frac{1}{a_{\min}}, \quad \|E\|_{\sigma} = (\max_i \epsilon_i)L, \quad \|F\|_{\sigma} = N(\max_j \delta_j).$$

- Por desigualdad triangular y submultiplicatividad,

$$\|D^{-1}X\|_{\sigma} \leq \|D^{-1}\|_{\sigma}(\|E\|_{\sigma} + \|F\|_{\sigma}) = \frac{L \max_i \epsilon_i + N \max_j \delta_j}{a_{\min}} < 1.$$

- Luego la serie de Neumann converge:

$$(I + D^{-1}X)^{-1} = \sum_{k=0}^{\infty} (-1)^k (D^{-1}X)^k.$$

Claramente $\|\pi_n - \pi^*\|_\infty = \|\mathcal{E}_n b\|_\infty \leq \|\mathcal{E}_n b\|_2$

$$\begin{aligned}\|\mathcal{E}_n b\|_2 &\leq \sqrt{NL} \left\| \sum_{k=n+1}^{\infty} (-1)^k (D^{-1}X)^k \right\|_\sigma \|D^{-1}b\|_\infty \\ &\leq \frac{\sqrt{NL} \|D^{-1}X\|_\sigma^{n+1} \|D^{-1}b\|_\infty}{1 - \|D^{-1}X\|_\sigma}.\end{aligned}$$

Entonces, dado $\eta > 0$,

$$N_\eta = \max \left\{ 1, \left\lceil \left\lceil \log_{\|D^{-1}X\|_\sigma} \left(\frac{\eta (1 - \|D^{-1}X\|_\sigma)}{\sqrt{NL} \|D^{-1}b\|_\infty} \right) \right\rceil \right\rceil \right\},$$

y para $n \geq N_\eta$ se sigue que $\|\pi_n - \pi^*\|_\infty < \eta$.

Supuesto

Supóngase que $\delta_j = 0$ para todo j , que $D = \beta I$ con $\beta > 0$, y que $L\epsilon_i < \beta$ para todo i .

- Fórmula explícita para A^{-1} :

$$A^{-1} = \frac{I}{\beta} + \frac{1}{\beta} \text{Diag} \left(-\frac{\epsilon_1}{\beta + L\epsilon_1}, \dots, -\frac{\epsilon_N}{\beta + L\epsilon_N} \right) \otimes \mathbf{1}_{L \times L}.$$

- De ahí, fórmula explícita para π_{ij}^* (siempre que el lado derecho sea positivo):

$$\pi_{ij}^* = \frac{b_{ij}}{\beta} - \sum_{\ell=1}^L \frac{b_{i\ell}\epsilon_i}{\beta^2 + L\epsilon_i\beta}.$$

$$\begin{aligned}
 A^{-1} &= \frac{I}{\beta} + \left[\sum_{k=1}^{\infty} (-1)^k \left(\frac{1}{\beta} \right)^k (\text{Diag}(\epsilon_1, \dots, \epsilon_N) \otimes \mathbf{1}_{L \times L})^k \right] D^{-1} \\
 &= \frac{I}{\beta} + \frac{1}{\beta L} \sum_{k=1}^{\infty} (-1)^k \left(\frac{L}{\beta} \right)^k (\text{Diag}(\epsilon_1^k, \dots, \epsilon_N^k) \otimes \mathbf{1}_{L \times L}) \\
 &= \frac{I}{\beta} + \frac{1}{\beta L} \text{Diag} \left(\sum_{k=1}^{\infty} (-1)^k \left(\frac{L\epsilon_1}{\beta} \right)^k, \dots, \sum_{k=1}^{\infty} (-1)^k \left(\frac{L\epsilon_N}{\beta} \right)^k \right) \otimes \mathbf{1}_{L \times L} \\
 &= \frac{I}{\beta} + \frac{1}{\beta} \text{Diag} \left(-\frac{\epsilon_1}{\beta + L\epsilon_1}, \dots, -\frac{\epsilon_N}{\beta + L\epsilon_N} \right) \otimes \mathbf{1}_{L \times L}.
 \end{aligned}$$

Usamos la propiedades del producto Kronecker:

- $(A \otimes B)^k = A^k \otimes B^k$.
- $\mathbf{1}_{L \times L}^k = L^{k-1} \mathbf{1}_{L \times L}$.

Supuesto

Sean $a_{ij} = \rho$ y $\epsilon_i = \delta_j = \zeta$ para todo (i, j) , con $\rho > 2NL\zeta > 0$.

- En este caso, $A = \rho I + \zeta Y$ donde Y es una matriz estructurada.
- Entonces:

$$A^{-1} = \frac{1}{\rho} \sum_{k=0}^{\infty} \left(-\frac{\zeta}{\rho}\right)^k Y^k.$$

- Se pueden acotar los términos $(A^{-1})_{ij}$ por constantes explícitas C_1 , C_2 .
- Concretamente:

$$\pi_{ij}^* \leq NL \cdot \max\{|C_1|, C_2\} \cdot \max_{i,j} \left| \epsilon_i \mu_i + \delta_j \nu_j - \frac{c_{ij}}{2} \right|.$$

Donde:

$$C_1 = \frac{1}{\rho} \left(-2 \left(\frac{\zeta}{\rho} \right) + \frac{\zeta^2}{\rho^2 - \zeta^2 NL} - \frac{8\zeta^3 N^2 L^2}{\rho(\rho^2 - 4\zeta^2 N^2 L^2)} \right),$$

$$C_2 = \frac{1}{\rho} \left(1 + \frac{4\zeta^2 NL}{\rho^2 - 4\zeta^2 N^2 L^2} - \frac{\zeta^3}{\rho(\rho^2 - \zeta^2 NL)} \right).$$

Algoritmo

Algoritmo

Primero, escribamos

$$\begin{aligned} A = & \text{Diag}(a_{11}, \dots, a_{NL}) + \sum_{i=1}^N \left(\epsilon_i^{1/2} e_i \otimes \mathbf{1}_{L \times 1} \right) \left(\epsilon_i^{1/2} e_i^T \otimes \mathbf{1}_{1 \times L} \right) \\ & + \sum_{j=1}^L \left(\delta_j^{1/2} \mathbf{1}_{N \times 1} \otimes e_j \right) \left(\delta_j^{1/2} \mathbf{1}_{1 \times N} \otimes e_j^T \right). \end{aligned} \quad (7)$$

Usamos la propiedad del producto de Kronecker:

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD),$$

siempre que las dimensiones sean compatibles. Entonces, para cada i se tiene:

$$\begin{aligned} \left(\epsilon_i^{1/2} e_i \otimes \mathbf{1}_{L \times 1} \right) \left(\epsilon_i^{1/2} e_i^\top \otimes \mathbf{1}_{1 \times L} \right) &= \epsilon_i \left(e_i e_i^\top \otimes \mathbf{1}_{L \times 1} \cdot \mathbf{1}_{1 \times L} \right) \\ &= \epsilon_i \left(e_i e_i^\top \otimes \mathbf{1}_{L \times L} \right). \end{aligned}$$

Sumando sobre $i = 1, \dots, N$:

$$\begin{aligned} \sum_{i=1}^N \epsilon_i \left(e_i e_i^\top \otimes \mathbf{1}_{L \times L} \right) &= \left(\sum_{i=1}^N \epsilon_i e_i e_i^\top \right) \otimes \mathbf{1}_{L \times L} \\ &= \text{Diag}(\epsilon_1, \dots, \epsilon_N) \otimes \mathbf{1}_{L \times L}. \end{aligned}$$

- Supongamos que ya conocemos la inversa de una matriz A^{-1} .
- Recibe un cambio pequeño: una actualización de rango 1 de la forma $A + uv^{\top}$.

- Notar que

$$uv^{\top} = \begin{bmatrix} v_1 u & \cdots & v_n u \end{bmatrix}.$$

- ¿Debemos volver a invertir toda la matriz?
- **¡No!** Usemos Sherman–Morrison.

Sean:

- $A \in \mathbb{R}^{n \times n}$: matriz invertible,
- $u, v \in \mathbb{R}^n$: vectores columna,
- $uv^\top$: producto exterior (rango 1).

Si $1 + v^\top A^{-1}u \neq 0$, entonces:

$$\left(A + uv^\top\right)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u}$$

Los siguientes términos son **actualizaciones de rango 1**:

$$\epsilon_i^{1/2} e_i \otimes \mathbf{1}_{L \times 1} \cdot \epsilon_i^{1/2} e_i^\top \otimes \mathbf{1}_{1 \times L} \quad \text{y} \quad \delta_j^{1/2} \mathbf{1}_{N \times 1} \otimes e_j \cdot \delta_j^{1/2} \mathbf{1}_{1 \times N} \otimes e_j^\top.$$

Usando las propiedades del producto Kronecker $((A \otimes B)^T = A^T \otimes B^T)$:

$$\epsilon_i \left(e_i e_i^\top \otimes \mathbf{1}_{L \times L} \right) \quad \text{y} \quad \delta_j \left(\mathbf{1}_{N \times N} \otimes e_j e_j^\top \right).$$

Primero:

$$e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^n \Rightarrow e_2 e_2^\top = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Luego:

$$\epsilon_2 \left(e_2 e_2^\top \otimes \mathbf{1}_{2 \times 2} \right) = \epsilon_2 \cdot \begin{bmatrix} \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{1}_{2 \times 2} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \end{bmatrix} = \epsilon_2 \cdot \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

¿Quién es u en Sherman-Morrison?

En cada paso, aplicamos Sherman-Morrison para una matriz de la forma $A + uu^\top$.

- En el bucle sobre $i = 1, \dots, N$:

$$u^{(i)} := \epsilon_i^{1/2} e_i \otimes \mathbf{1}_{L \times 1}$$

- En el bucle sobre $j = 1, \dots, L$:

$$v^{(j)} := \delta_j^{1/2} \mathbf{1}_{N \times 1} \otimes e_j$$

Ambos vectores definen productos exteriores del tipo $uv^\top$, que son actualizaciones de rango 1.

Aplicamos:

$$A^{-1} \leftarrow A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u}$$

Condición clave: $1 + u^\top A^{-1}u \neq 0$

Para que la fórmula de Sherman–Morrison sea válida, es necesario que:

$$1 + u^\top A^{-1}u \neq 0$$

- Esto se cumple automáticamente si:

- $A^{-1} \succ 0$ (matriz definida positiva),
- $u \neq 0$,

- En tu caso:

- Se inicializa $A^{-1} = \text{Diag}(1/a_{ii})$, con $a_{ii} > 0$,
- Los vectores $u^{(i)}$ y $v^{(j)}$ son no nulos porque $\epsilon_i, \delta_j > 0$.

Entonces: todas las actualizaciones cumplen la condición.

Algorithm: Optimize($a, b, \epsilon_1, \dots, \epsilon_N, \delta_1, \dots, \delta_L$)

- 1: **Input:** Vector $a \in \mathbb{R}_{++}^{NL}$, vector $b \in \mathbb{R}^{NL}$ and parameters $\epsilon_1, \dots, \epsilon_N$, $\delta_1, \dots, \delta_L$ in $\mathbb{R}_{++}$
- 2: **Output:** $\pi^* \in \mathbb{R}^{NL}$
- 3: Initialize $A^{-1} \leftarrow \text{Diag}(1/a_{11}, \dots, 1/a_{NL})$
- 4: **for** $i = 1$ to N **do**
- 5: Define $u^{(i)} \leftarrow \epsilon_i^{1/2} e_i \otimes 1_{L \times 1}$
- 6:
$$A^{-1} \leftarrow A^{-1} - \frac{A^{-1} u^{(i)} u^{(i)T} A^{-1}}{1 + u^{(i)T} A^{-1} u^{(i)}}$$
- 7: **end for**
- 8: **for** $j = 1$ to L **do**
- 9: Define $v^{(j)} \leftarrow \delta_j^{1/2} 1_{N \times 1} \otimes e_j$
- 10:
$$A^{-1} \leftarrow A^{-1} - \frac{A^{-1} v^{(j)} v^{(j)T} A^{-1}}{1 + v^{(j)T} A^{-1} v^{(j)}}$$
- 11: **end for**
- 12: **return** $A^{-1} b$

Teorema: Complejidad del Algoritmo Optimize

Teorema

Sea π^* una solución interior del sistema $A\pi = b$. Entonces, el algoritmo Optimize computa π^* en tiempo

$$O((N + L)(NL)^2).$$

- La matriz A tiene la forma:

$$A = \text{Diag}(a) + \sum_{i=1}^N u^{(i)} u^{(i)\top} + \sum_{j=1}^L v^{(j)} v^{(j)\top}$$

donde $u^{(i)}$ y $v^{(j)}$ son vectores definidos por productos de Kronecker.

- Se inicializa $A^{-1} = \text{Diag}(1/a)$, y se aplican sucesivas actualizaciones de rango 1.
- Cada actualización usa la fórmula de Sherman-Morrison:

$$A^{-1} \leftarrow A^{-1} - \frac{A^{-1} w w^{\top} A^{-1}}{1 + w^{\top} A^{-1} w},$$

donde $w = u^{(i)}$ o $w = v^{(j)}$.

- Hay $N + L$ actualizaciones.
- Cada actualización requiere computar:
 - $w^\top A^{-1}w$: $O((NL)^2)$,
 - $A^{-1}w$: $O((NL)^2)$, $NL \times NL$ (se recorre toda la matriz)
 - $A^{-1}ww^\top A^{-1}$: $O((NL)^2)$ con una buena parentización (agrupar $ww^\top$ perjudica).
- Así, en total: $O((N + L)(NL)^2)$.

Comparación de algoritmos para resolver $A\pi = b$

Tiempo	A dispersa	Galáctico	Referencia
$O((NL)^3)$	No	No	Folklore
$O((NL)^{2,81})$	No	No	Strassen (1969)
$O((N + L)(NL)^2)$	No	No	Este trabajo
$O((NL)^{2,371339})$	No	Sí	Alman (2025)
$O((NL)^{2,331645})$	Sí	Sí	PV (2024)

Cuadro: Comparación de algoritmos para resolver sistemas lineales.

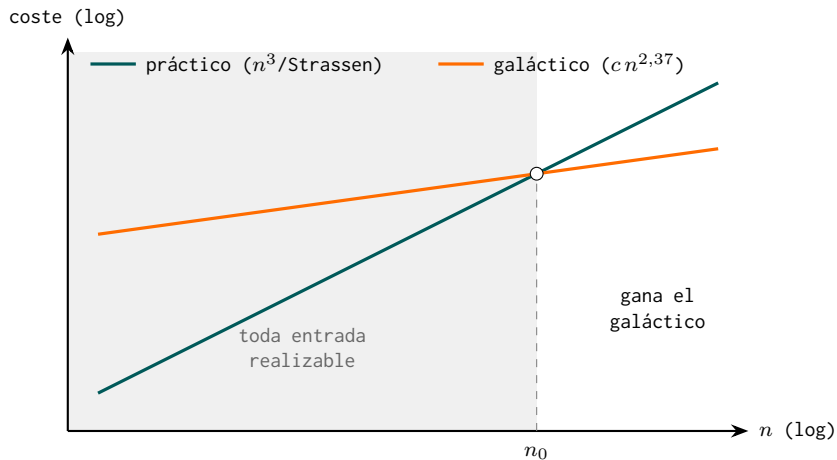
¿Qué es un algoritmo galáctico?

Definición (informal)

Un **algoritmo galáctico** tiene mejor complejidad asintótica, pero con constantes ocultas tan grandes que el umbral n_0 a partir del cual supera a los métodos usuales excede cualquier entrada físicamente representable: es inservible en la práctica (término de R. Lipton).

- Alman et al. [Alman et al., 2025] dan los mejores exponentes conocidos para multiplicación de matrices, vía el laser method [Strassen, 1986, Coppersmith and Winograd, 1990, Davie and Stothers, 2013, Williams, 2012].
- Bajan el exponente ω (de 3 hacia $\approx 2,37$), **no** el tiempo real: el n_0 de cruce es astronómico.
- En la tabla, la columna Galáctico marca justamente estos métodos.

La intuición: el punto de cruce n_0



Lectura para el modelo

El galáctico solo gana en $n > n_0$, con n_0 , por ejemplo, mayor que los átomos del universo. Armar e invertir A , truncar Neumann para A^{-1}) vive a la izquierda de n_0 : la cota relevante es la **práctica**, no el 2,37 teórico.

Estática comparativa

Estática comparativa

- Aunque podemos computar π^* con la serie de Neumann o el Algoritmo 1, no es fácil obtener una expresión cerrada de π_{ij}^* .
- Como primera aproximación, usamos $A^{-1} \approx D^{-1}$, lo que da la forma cerrada aproximada

$$\pi_{ij}^* \simeq \frac{2(\epsilon_i \mu_i + \delta_j \nu_j) - c_{ij}}{2a_{ij}}.$$

- Signos (aproximados): $\partial \pi_{ij}^* / \partial a_{ij}, \partial \pi_{ij}^* / \partial c_{ij} < 0$;
 $\partial \pi_{ij}^* / \partial \epsilon_i, \partial \pi_{ij}^* / \partial \delta_j, \partial \pi_{ij}^* / \partial \mu_i, \partial \pi_{ij}^* / \partial \nu_j > 0$.
- Limitación: $\partial \pi_{ij}^* / \partial \theta_{kl} \simeq 0$ si $(k, l) \neq (i, j) \Rightarrow$ no hay efecto sustitución (poco realista).

- Si $\pi^* \in \mathbb{R}_{++}^{NL}$ es una solución interior de $\mathcal{P}_{CP}$, la estática comparativa suave da la igualdad exacta

$$\left[\frac{\partial \pi_{ij}^*}{\partial \theta_{k\ell}} \right] = -A^{-1} \left[\frac{1}{2} I_{NL \times NL} \mid \text{Diag}(\pi_{11}^*, \dots, \pi_{NL}^*) \right].$$

- El problema es que A^{-1} no tiene forma cerrada manejable.
- Idea: reemplazar A^{-1} por una **aproximación** de Neumann truncada para obtener expresiones interpretables.

- Aproximamos ahora $A^{-1} \sim D^{-1} - D^{-1}XD^{-1}$ (Neumann a segundo orden).
- Sustituyendo en la fórmula exacta:

$$\left[\frac{\partial \pi_{ij}^*}{\partial \theta_{k\ell}} \right] \simeq - \left[\frac{1}{2} \left(D^{-1} - D^{-1}XD^{-1} \right) \mid A_{\Pi,2}^{-1} \right],$$

donde $A_{\Pi,2}^{-1}$ multiplica la columna ij de $D^{-1} - D^{-1}XD^{-1}$ por π_{ij}^* .

- Si $\max_{i,j} \{\epsilon_i + \delta_j\} < \min_{(i,j) \in I \times J} a_{ij}$, los signos (aproximados) son:
 - $\partial \pi_{ij}^* / \partial \theta_{ij} < 0$
 - $\partial \pi_{ij}^* / \partial \theta_{i\ell}, \partial \pi_{ij}^* / \partial \theta_{kj} > 0$
 - $\partial \pi_{ij}^* / \partial \theta_{k\ell} \simeq 0$ si $i \neq k$ y $j \neq \ell$

Todas las expresiones siguientes son **aproximaciones** (Neumann truncada):

$$\partial \pi_{ij}^* / \partial c_{ij} \simeq - \frac{a_{ij} - (\epsilon_i + \delta_j)}{2a_{ij}^2} < 0,$$

$$\partial \pi_{ij}^* / \partial c_{il} \simeq \frac{\epsilon_i}{2a_{ij}a_{il}} > 0, \quad \partial \pi_{ij}^* / \partial c_{kj} \simeq \frac{\delta_j}{2a_{ij}a_{kj}} > 0,$$

$$\partial \pi_{ij}^* / \partial c_{k\ell} \simeq 0 \text{ si } i \neq k, j \neq \ell.$$

$$\partial \pi_{ij}^* / \partial a_{ij} \simeq - \frac{\pi_{ij}^* (a_{ij} - (\epsilon_i + \delta_j))}{a_{ij}^2} < 0,$$

$$\partial \pi_{ij}^* / \partial a_{il} \simeq \frac{\pi_{il}^* \epsilon_i}{a_{ij}a_{il}} > 0, \quad \partial \pi_{ij}^* / \partial a_{kj} \simeq \frac{\pi_{kj}^* \delta_j}{a_{ij}a_{kj}} > 0,$$

$$\partial \pi_{ij}^* / \partial a_{k\ell} \simeq 0 \text{ si } i \neq k, j \neq \ell.$$

- Ahora sí aparece el **efecto sustitución**: si aumenta el costo de unir a i con j (ceteris paribus), π_{ij}^* baja y π_{il}^* , π_{kj}^* suben.
- Énfasis: todo lo anterior son **aproximaciones** (Neumann truncada), no igualdades exactas; deben interpretarse como tales, aunque capturan la intuición económica esperada.
- En el caso especial con $\delta_j = 0$, $D = \beta I$ ($\beta > 0$) y $L\epsilon_i < \beta$, los efectos sí pueden computarse de forma exacta con la fórmula cerrada de A^{-1} .

Aplicaciones

Aplicaciones del modelo

Motivación: Aplicamos el modelo en sectores críticos del Perú donde existen problemas estructurales y costos de acceso significativos.

- El modelo es **normativo** (prescriptivo), no descriptivo.
- Permite capturar congestión, ineficiencias y demandas no satisfechas.
- Da flexibilidad al SP.

- Fragmentado en SIS (Seguro Integral de Salud), EsSalud, y EPS (Entidades Prestadoras de Salud) [Anaya-Montes and Gravelle, 2024].
 - ① EPS: privado.
 - ② EsSalud: trabajadores formales, tanto privado como público.
 - ③ SIS: universal, aplica a informales.
- SIS y EsSalud: $\simeq 80\%$. $\simeq 10\%$ EPS.
- Cobertura: sin intersecciones, salvo EPS y EsSalud [Velasquez, 2020].

Problema identificado	Indicador cuantificable
Escasez de personal médico en atención primaria	12 médicos por cada 10,000 habitantes, muy por debajo del estándar de 43 recomendado por la OMS [Bendezu-Quispe et al., 2020].
Falta de camas hospitalarias	1.6 camas por cada 1,000 habitantes, debajo del promedio regional [World Bank, 2020].
Congestión en UCI neonatales	50% de las unidades presentan ineficiencia por sobrecarga de pacientes [Arrieta and Guillén, 2017].
Ineficiencias en el sistema de referencias	Alta proporción de pacientes tratados en centros no equipados [Soto, 2019].
Baja cobertura y largas esperas	Cobertura de hasta 86% para ciertos tratamientos complejos [EsSalud, 2025a].
Altas tasas de postergación en ciudades clave	Más del 23% de citas postergadas (enero-marzo 2025) [EsSalud, 2025b].

Nota: En 2016, el MINSA reportó un déficit de más de 47,000 profesionales de salud. El 36% de los

- 3 grupos
- Fricciones en los grupos: sectores económicos, geografía.
- Tabla (slide anterior): congestión, excesos de demanda.

Ejemplo (continúa)

■ Resolvemos el problema $\mathcal{P}_{CP}$ con $\mathbf{d} = 5I_{3 \times 3}$ y $\alpha = 0,5$.

■ Parámetros utilizados:

$$\mathbf{c} = \begin{bmatrix} 1 & 50 & 20 \\ 50 & 1 & 20 \\ 20 & 10 & 1 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} 1 & 5 & 10 \\ 5 & 1 & 2 \\ 10 & 5 & 1 \end{bmatrix}, \quad \boldsymbol{\epsilon} = \boldsymbol{\delta} = \begin{bmatrix} 0,3 \\ 0,3 \\ 0,3 \end{bmatrix}$$

$$\boldsymbol{\mu} = \begin{bmatrix} 100 \\ 50 \\ 20 \end{bmatrix}, \quad \boldsymbol{\nu} = \begin{bmatrix} 90 \\ 40 \\ 40 \end{bmatrix}.$$

■ Solución óptima $\boldsymbol{\pi}^*$ obtenida en Mathematica 14.1 usando el Algoritmo 1¹:

$$\boldsymbol{\pi}^* = \begin{bmatrix} 34,7802 & 0,19412 & 1,65935 \\ 0,10148 & 15,6978 & 3,41038 \\ 0,883807 & 0,905689 & 9,65139 \end{bmatrix}.$$

¹También verificamos el resultado con QuadraticOptimization.

- Se resuelve $\mathcal{P}_Q$ y $\mathcal{P}_O$ con las mismas entradas que en $\mathcal{P}_{CP}$, pero imponiendo $\Pi(\mu, \nu)$ y eliminando penalizaciones.
- Soluciones óptimas:

$$\pi_{\mathcal{P}_Q}^* = \begin{bmatrix} 84,275 & 8,84062 & 6,88442 \\ 4,2985 & 30,4206 & 15,2809 \\ 1,42655 & 0,73873 & 17,8347 \end{bmatrix}, \quad \pi_{\mathcal{P}_O}^* = \begin{bmatrix} 90 & 0 & 10 \\ 0 & 40 & 10 \\ 0 & 0 & 20 \end{bmatrix}.$$

- Tráfico vehicular.
- Preferencias: Afinidad académica-cultural, entorno social.
- Capacidad (instalación).
- Álvaro Mauricio Morán (usando el API de Google Maps).



Figura: Lunes 7pm



Figura: Lunes 7pm

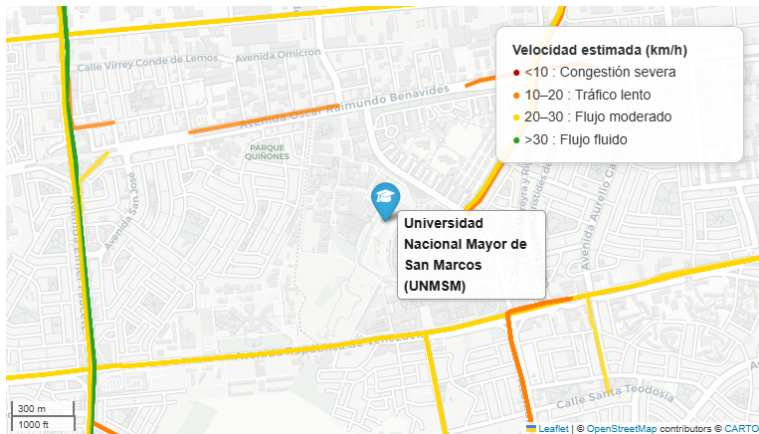


Figura: Lunes 7pm



Figura: Lunes 7pm

Extensiones

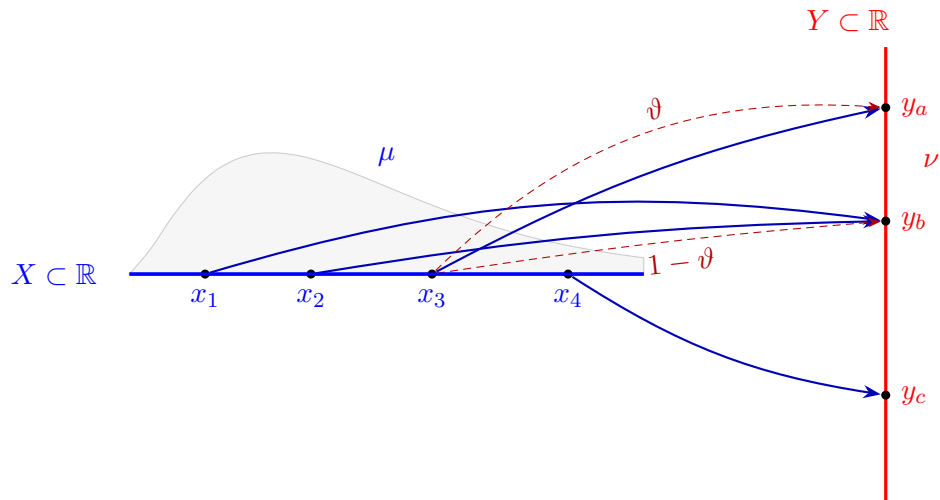
Extensiones

- Infinitos tipos (caso continuo).
- Dinámica y aleatoriedad.
- Estimación de los parámetros.
- Determinar solución interior.

Infinitos tipos

- Aparece de forma natural: p. ej. mercado laboral, donde emparejamos tipos de trabajador (productividad en un compacto de $\mathbb{R}$) con firmas (nivel tecnológico en un compacto de $\mathbb{R}$).
- Necesitamos un marco medida-teórico: pasamos de vectores μ, ν a **medidas** μ, ν y de matrices π_{ij} a un **plan** (medida) π sobre $X \times Y$.
- El ambiente natural son los **espacios Polacos**.

Matching between bounded subsets of $\mathbb{R}$



Modelos en el continuo

Extensión continua: modelo 2 (metas penalizadas)

Pesos $\epsilon \in L^\infty(X)$, $\delta \in L^\infty(Y)$:

$$\inf_{f \in L^2_+(X \times Y)} \left\{ \int_{X \times Y} (a(x, y) f^2 + c(x, y) f) dx dy \right. \\ \left. + \int_X \epsilon(x) (F(x) - f_\mu(x))^2 dx + \int_Y \delta(y) (G(y) - f_\nu(y))^2 dy \right\}.$$

- $F = \int_Y f dy$, $G = \int_X f dx$ (densidades de las marginales).
- Las metas f_μ, f_ν se penalizan (no se imponen exactamente).

- Caso particular $\epsilon \equiv \delta \equiv 1$:

$$\inf_{f \in L^2_+} \int_{X \times Y} (af^2 + cf) \, dx \, dy + \|F - f_\mu\|_{L^2(X)}^2 + \|G - f_\nu\|_{L^2(Y)}^2.$$

- f = derivada de R-N de π respecto a Lebesgue; $a, c: X \times Y \rightarrow \mathbb{R}_+$.
- Preguntas: ¿existencia? ¿unicidad? ¿qué hipótesis sobre espacios y funciones?

Supuesto

X, Y compactos métricos, cada uno con una medida de probabilidad Boreliana (Lebesgue normalizada en subconjuntos de $\mathbb{R}^d$). Además,

$$a(x, y) \geq a_0 > 0 \text{ c.t.p.}, \quad c \in L_+^2(X \times Y),$$

$$\epsilon \in L^\infty(X), \quad \epsilon \geq 0 \text{ c.t.p.}, \quad \delta \in L^\infty(Y), \quad \delta \geq 0 \text{ c.t.p.}$$

- Compacidad de $X, Y \Rightarrow$ heterogeneidad económicamente acotada (habilidades, capacidad, calidad en rangos factibles); masa y agregados finitos.
- Normalizar no cambia las políticas óptimas.

L_+^2 es convexo y cerrado

Convexo: trivial. Para **cerrado**, sea $\{f_n\} \subset L_+^2$ con $f_n \rightarrow f$ en L^2 . Supongamos, por contradicción, $f \notin L_+^2$: existen S medible con $\mathcal{L}(S) > 0$ y $k \in \mathbb{N}$ tales que

$$f(x) < -\frac{1}{k} \quad \forall x \in S.$$

Entonces, para todo n ,

$$\|f_n - f\|_{L^2}^2 \geq \int_S |f_n - f|^2 dx \geq \int_S \frac{1}{k^2} dx = \frac{1}{k^2} \mathcal{L}(S) > 0,$$

lo que contradice $f_n \rightarrow f$. Por tanto $f \geq 0$ c.t.p., i.e. $f \in L_+^2$. $\square$

- $\|f\sqrt{a}\|_{L^2}^2 = \int_{X \times Y} a f^2$ es (estrictamente²) convexa:

$$\int a (\theta f + (1 - \theta)g)^2 < \int a (\theta f^2 + (1 - \theta)g^2).$$

- $I : f \mapsto F$ y $J : f \mapsto G$ son lineales; como $\|\cdot\|_{L^2}^2$ es convexa, los términos de penalización $\|I(f) - f_\mu\|^2 + \|J(f) - f_\nu\|^2$ son convexos.
- $\langle f, c \rangle = \int c f$ es lineal.
- Suma de convexa estricta + convexas $\Rightarrow \Phi$ estrictamente convexa.

²Si $a > 0$ c.t.p.

- $\langle f, c \rangle$ es lineal y continua: $|\int c(f_n - f)| \leq \|c\|_{L^2} \|f_n - f\|_{L^2} \rightarrow 0$.
- I, J son contracciones (espacios de probabilidad, $\mathcal{L}(X) = \mathcal{L}(Y) = 1$):

$$\|I(f)\|_{L^2(X)}^2 = \int_X \left| \int_Y f \, dy \right|^2 dx \leq \mathcal{L}(Y) \|f\|_{L^2(X \times Y)}^2 = \|f\|_{L^2}^2,$$

luego I, J continuas $\Rightarrow$ términos de penalización continuos (s.c.i.).

- El término $\|f\sqrt{a}\|_{L^2}^2$ es s.c.i. por el **Lema de Fatou** ($a \geq 0$).

Sea $f_n \rightarrow f$ en L^2 . Pasando a una subsucesión (no renombrada), $f_n \rightarrow f$ c.t.p. Con $g_n := a f_n^2 \geq 0$ y $g := a f^2$, por Fatou:

$$\begin{aligned} \int_{X \times Y} a f^2 dx dy &= \int_{X \times Y} g dx dy \\ &\leq \int_{X \times Y} \liminf_{n \rightarrow \infty} g_n dx dy \\ &\leq \liminf_{n \rightarrow \infty} \int_{X \times Y} a f_n^2 dx dy. \end{aligned}$$

Es decir,

$$\|f\sqrt{a}\|_{L^2}^2 \leq \liminf_{n \rightarrow \infty} \|f_n\sqrt{a}\|_{L^2}^2.$$

Sumando las partes: $\Phi(f) \leq \liminf_n \Phi(f_n)$, y como L_+^2 es cerrado, la restricción también es s.c.i.

- **Reflexivo:** E es reflexivo si la inyección canónica $E \hookrightarrow E^{**}$ es sobreyectiva. Aquí $E = L^2(X \times Y)$ es de Hilbert $\Rightarrow$ reflexivo.
- **Convexo y cerrado:** $L^2_+ = \{f \in L^2 : f \geq 0 \text{ c.t.p.}\}$ es convexo y cerrado en norma.
- **Lineal continuo $\iff$ acotado:** $T : E \rightarrow F$ lineal es continuo si y solo si $\exists C \geq 0$ con $\|Tx\|_F \leq C\|x\|_E$; la menor C es $\|T\|$.
- **En el modelo:** los operadores marginales $(P_X f)(x) = \int_Y f(x, y) dy$ son lineales y acotados ($\|P_X f\| \leq \sqrt{|Y|} \|f\|$), luego continuos $\Rightarrow$ las penalizaciones son continuas.

- **Propia:** $\Phi : A \rightarrow (-\infty, +\infty]$ con $\text{dom } \Phi = \{f : \Phi(f) < +\infty\} \neq \emptyset$ y $\Phi > -\infty$. Como el codominio es $(-\infty, +\infty]$, basta que Φ no sea idénticamente $+\infty$.
- **s.c.i. (débil):** $f_n \rightharpoonup f \Rightarrow \Phi(f) \leq \liminf_n \Phi(f_n)$. Clave: para Φ convexa, s.c.i. fuerte $\iff$ s.c.i. débil (epígrafo convexo y cerrado en norma $\Rightarrow$ débilmente cerrado).
- **Coerciva:** $\Phi(f) \rightarrow +\infty$ cuando $\|f\| \rightarrow \infty$.
- **Estrictamente convexa:** $\Phi(tf + (1-t)g) < t\Phi(f) + (1-t)\Phi(g)$ para $f \neq g$, $t \in (0,1) \Rightarrow$ minimizador único.

Teorema

Bajo el Supuesto anterior, el modelo 2 admite un **único** minimizador en $L^2_+(X \times Y)$.

Idea (método directo). Con las definiciones anteriores, Φ cumple sobre el convexo cerrado L^2_+ (en L^2 reflexivo):

- **coerciva**: como $a \geq a_0 > 0$ y las penalizaciones son ≥ 0 ,

$$\Phi(f) \geq a_0 \|f\|_{L^2}^2 \xrightarrow{\|f\| \rightarrow \infty} +\infty;$$

- **s.c.i. débil**: por ser convexa (parte cuadrática continua + penalizaciones convexas continuas);
- **estrictamente convexa**: por $a \geq a_0 > 0$.

Brezis da la existencia; la unicidad se sigue de la convexidad estricta.

□

Teorema (Brezis, 2010)

Sea E un Banach reflexivo, $A \subseteq E$ cerrado y convexo, y $\varphi: A \rightarrow (-\infty, \infty]$ convexa, s.c.i. y propia. Si A es acotado o φ es coerciva, entonces φ alcanza su mínimo en A .

- Si se quita $a \geq a_0 > 0$, se restringe a

$$\mathcal{A}_M = \{ f \in L_+^2(X \times Y) : 0 \leq f \leq M \text{ c.t.p.} \}.$$

- $\mathcal{A}_M$ es no vacío, cerrado, convexo y acotado en L^2 (reflexivo) $\Rightarrow$ débilmente compacto.
- Φ convexa y débilmente s.c.i. $\Rightarrow$ alcanza el mínimo por Brezis (caso acotado).
- Útil también para compactar con fines computacionales.

Estimación de los parámetros

- Nelder-Mead para

$$\min_{\mathbf{c}, \mathbf{a}} \left\{ \sum_{i,j} \left(\pi_{ij}^{\text{obs}} - \underbrace{\pi_{ij}^*(\mathbf{c}, \mathbf{a})}_{A^{-1}b} \right)^2 \right\},$$

dados $\epsilon, \delta, \mu, \nu, \alpha$.

- Especificación: $c_{ij} = \beta_0 + \beta_1 X_{1ij} + \dots + \beta_k X_{kij} + \eta_{ij}$.
- Regímenes límite: $\iota a_{ij} \gg \tilde{a}_{ij}$? $\mu, \nu \rightarrow \infty$, $[c_{ij}] \simeq c \cdot I_{NL \times NL}$.
- Ejemplo:

$$c(i, j) = \Psi(\text{dificultad admisión}_j, \text{rendimiento}_i, \\ \text{gusto deporte}_i, \text{infraestructura}_j).$$

Dinámica y aleatoriedad

- Secuencia de planes $\{\boldsymbol{\pi}_t\}_{t=1}^T$; parámetros $c_{ij}^t, a_{ij}^t, \mu_i^t, \nu_j^t$ estocásticos.
- Penalizamos cambios abruptos con una norma cuadrática intertemporal:

$$\begin{aligned} \min_{\{\boldsymbol{\pi}_t\}} \mathbb{E} \bigg[& \sum_{t=1}^T \left(\sum_{i,j} c_{ij}^t \pi_{ij}^t + a_{ij}^t (\pi_{ij}^t)^2 \right. \\ & + \sum_i \epsilon_i^t \left(\sum_j \pi_{ij}^t - \mu_i^t \right)^2 + \sum_j \delta_j^t \left(\sum_i \pi_{ij}^t - \nu_j^t \right)^2 \bigg) \\ & \left. + \lambda \sum_{t=2}^T \|\boldsymbol{\pi}_t - \boldsymbol{\pi}_{t-1}\|_F^2 \right]. \end{aligned}$$

- Evolución autorregresiva:

$$\begin{aligned}c_{ij}^{t+1} &= \rho_c c_{ij}^t + \eta_{ij}^{t+1}, & a_{ij}^{t+1} &= \rho_a a_{ij}^t + \zeta_{ij}^{t+1}, \\ \mu_i^{t+1} &= \rho_\mu \mu_i^t + \xi_i^{t+1}, & \nu_j^{t+1} &= \rho_\nu \nu_j^t + \vartheta_j^{t+1}.\end{aligned}$$

- Innovaciones $\eta, \zeta, \xi, \vartheta$: media cero, varianza constante.

- El planificador afecta costos vía inversión en infraestructura:

$$\begin{aligned}c_{ij}^{t+1} &= \rho_c c_{ij}^t + \beta u_{ij}^t + \eta_{ij}^{t+1}, \\a_{ij}^{t+1} &= \rho_a a_{ij}^t + \beta' u_{ij}^t + \zeta_{ij}^{t+1}.\end{aligned}$$

- u_{ij}^t : control (inversión que mejora conectividad entre i y j).
- Objetivo: maximizar la utilidad social agregada descontada,

$$\max_{\{u_t\}} \mathbb{E} \left[\sum_{t=1}^T \rho^t \left(F(\boldsymbol{\pi}_t) - \gamma \|u_t\|_F^2 \right) \right], \quad \rho \in (0, 1).$$

Gracias

Anexo matemático

Definición

Un espacio topológico (X, τ) es **Polaco** si existe una métrica d sobre X , completa y separable, que induce τ .

- **Completo:** toda sucesión de Cauchy converge en X .
- **Separable:** existe un subconjunto numerable denso.
- Ejemplos: $\mathbb{R}^n$ (euclídeo), espacios de Banach separables.
- Combinan las dos propiedades necesarias para hacer transporte óptimo sin patologías.

Definición (Push-forward)

Para $f : X \rightarrow Y$ Borel medible, el operador $f_{\#} : \mathcal{M}(X) \rightarrow \mathcal{M}(Y)$ se define por

$$(f_{\#}\mu)(B) = \mu(f^{-1}(B)), \quad \forall B \text{ Boreliano.}$$

Es la distribución de masa obtenida al transportar cada punto de $\text{supp } \mu$ a lo largo de f .

Definición (Plan de transporte)

Con p_X, p_Y las proyecciones canónicas, $\pi \in \Delta(X \times Y)$ es un plan si

$$(p_X)_{\#}\pi = \mu, \quad (p_Y)_{\#}\pi = \nu,$$

es decir, π tiene marginales μ y ν .

$$\Pi(\mu, \nu) = \left\{ \pi \in \Delta(X \times Y) : \begin{aligned} \pi(A \times Y) &= \mu(A), \\ \pi(X \times B) &= \nu(B), \quad \forall A, B \text{ Borelianos} \end{aligned} \right\}.$$

- Nunca es vacío: la medida producto $\mu \otimes \nu$ siempre pertenece a $\Pi(\mu, \nu)$ (acoplamiento independiente).

Problema de Monge (transporte determinista)

Dadas $\mu \in \Delta(X)$, $\nu \in \Delta(Y)$ y $c: X \times Y \rightarrow \mathbb{R}_+ \cup \{\infty\}$:

$$\inf \left\{ \int_X c(x, T(x)) d\mu(x) : T: X \rightarrow Y \text{ Borel}, T_{\#}\mu = \nu \right\}.$$

- T es un **mapa determinista**: toda la masa en x va íntegramente a $T(x)$ (no se divide).
- La restricción $T_{\#}\mu = \nu$ obliga a que T empuje μ exactamente sobre ν .
- **Limitación**: tal T puede no existir. Ej.: $\mu = \delta_a$ y $\nu = \frac{1}{2}\delta_b + \frac{1}{2}\delta_c$; un mapa no puede partir una masa puntual.

Con X, Y Polacos y $c: X \times Y \rightarrow \overline{\mathbb{R}}_+$:

$$\inf \left\{ \int_{X \times Y} c(x, y) d\pi(x, y) : \pi \in \Pi(\mu, \nu) \right\}.$$

- Generaliza el problema de Monge (transporte determinista T con $T_{\#}\mu = \nu$): $\inf(M) \geq \inf(K)$.
- **Pratelli**: si μ es no atómica y c continua, hay igualdad y el ínfimo se alcanza.

Definición

$(\mu_n) \subset \Delta(Z)$ converge débilmente a μ ($\mu_n \rightharpoonup \mu$) si

$$\int_Z f d\mu_n \rightarrow \int_Z f d\mu, \quad \forall f \in \underbrace{C_b(Z)}_{\text{continuas y acotadas}}.$$

La topología débil es la más gruesa que hace continuas las evaluaciones $\mu \mapsto \int_Z f d\mu$.

- Si Z es Polaco, $\Delta(Z)$ con la topología débil es también Polaco.
- Es la topología que vuelve s.c.i. al funcional de costo.

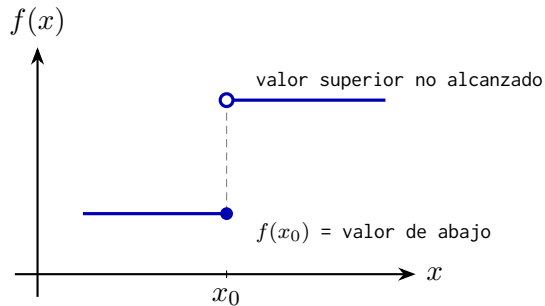
Definición

$f : X \rightarrow [-\infty, \infty]$ es **semicontinua inferior** si $\{x : f(x) > \alpha\}$ es abierto para todo $\alpha \in \mathbb{R}$. En espacios métricos, equivale a

$$f(x) \leq \liminf_{n \rightarrow \infty} f(x_n) \quad \text{siempre que } x_n \rightarrow x.$$

- Interpretación: cerca de x , los valores de f no caen abruptamente por debajo de $f(x)$; no hay saltos hacia abajo.
- Es la regularidad mínima para que el ínfimo se alcance (método directo del cálculo de variaciones).

Semicontinuidad inferior: intuición gráfica



En el salto, f toma el valor **inferior**: $f(x_0) \leq \liminf_{x \rightarrow x_0} f(x)$. No hay “caída abrupta” en el propio punto.

- La “variable” a optimizar es una **medida** π ; el objeto de estudio es el funcional de costo

$$J(\pi) := \int_{X \times Y} c(x, y) d\pi(x, y).$$

- **Propiedad clave:** si c es s.c.i., entonces J es s.c.i. respecto a la convergencia débil:

$$\pi_n \rightharpoonup \pi \implies J(\pi) \leq \liminf_{n \rightarrow \infty} J(\pi_n).$$

- Interpretación: el costo no salta hacia abajo en el límite. Es justo lo que impide que el ínfimo “se escape” y garantiza que se alcance.

- **Compacto** (en topología débil): de toda sucesión (π_n) en el conjunto se puede extraer una subsucesión que converge débilmente a un elemento del mismo conjunto. Es lo que permite “extraer límites”.
- **Relativamente compacto**: toda sucesión tiene una subsucesión débilmente convergente, aunque el límite quizá no esté en el conjunto (su clausura sí es compacta).
- **Uniformemente ajustada (tight)**: para cada $\varepsilon > 0$ existe un compacto K que concentra casi toda la masa de todas las medidas de la familia a la vez:

$$\mu(Z \setminus K) < \varepsilon \quad \forall \mu \text{ en la familia.}$$

Evita que la masa “se escape al infinito”.

- T1** Si c es s.c.i., entonces el funcional $J(\pi) = \int_{X \times Y} c d\pi$ es s.c.i. en la topología débil: $\pi_n \rightharpoonup \pi \Rightarrow J(\pi) \leq \liminf_n J(\pi_n)$.
- T2 (Prokhorov)** En un espacio Polaco, una familia acotada de medidas es relativamente compacta (débil) $\iff$ es uniformemente ajustada (tight).
- L1** $\Pi(\mu, \nu)$ es **compacto** en la topología débil (sus marginales fijas μ, ν garantizan tight; se aplica **T2**).
- L2** Toda función s.c.i. sobre un conjunto compacto alcanza su mínimo.

Teorema

Si $c : X \times Y \rightarrow [0, \infty]$ es s.c.i., el mínimo de Kantorovich se alcanza.

Idea (método directo del cálculo de variaciones):

- 1 El conjunto factible $\Pi(\mu, \nu)$ es **compacto** en la topología débil (**L1**, vía Prokhorov **T2**).
- 2 El funcional de costo J es **s.c.i.** en esa topología (**T1**).
- 3 Una función s.c.i. sobre un compacto alcanza su mínimo (**L2**).

Combinando 1-3: existe $\pi^* \in \Pi(\mu, \nu)$ con $J(\pi^*) = \text{mín.}$ $\square$

Definición

Sean μ, ν medidas σ -finitas sobre $(X, \mathcal{G})$ con $\nu \ll \mu$ (es decir, $\mu(A) = 0 \Rightarrow \nu(A) = 0$). Entonces existe $f : X \rightarrow [0, \infty)$ medible, **única salvo conjuntos μ -nulos**, tal que

$$\nu(A) = \int_A f \, d\mu, \quad \forall A \in \mathcal{G}.$$

Se denota $f = \frac{d\nu}{d\mu}$ y es la **densidad** de ν respecto a μ .

- La densidad depende de la medida de referencia μ : cambiar μ cambia f .

Definición

Sobre $X = \{x_1, \dots, x_N\}$, la **medida de conteo** C_X asigna $C_X(A) = \#(A)$ a cada $A \subseteq X$.

- El único conjunto C_X -nulo es $\emptyset$. Luego toda medida finita μ cumple $\mu \ll C_X$ automáticamente.

- Por Radon-Nikodym existe $f = \frac{d\mu}{dC_X}$ con

$$\mu(A) = \int_A f dC_X = \sum_{x \in A} f(x).$$

- La densidad respecto a C_X es simplemente la función de masa.

■ μ sobre $X = \{x_1, \dots, x_N\}$: $\mu = \sum_{i=1}^N \mu_i \delta_{x_i}$, con $\mu_i = \left(\frac{d\mu}{dC_X} \right) (x_i)$.

■ ν sobre $Y = \{y_1, \dots, y_L\}$: $\nu = \sum_{j=1}^L \nu_j \delta_{y_j}$, con $\nu_j = \left(\frac{d\nu}{dC_Y} \right) (y_j)$.

■ Toda medida π sobre $X \times Y$ (referencia $C_X \otimes C_Y$):

$$\pi = \sum_{i=1}^N \sum_{j=1}^L \pi_{ij} \delta_{(x_i, y_j)}, \quad \pi_{ij} = \left(\frac{d\pi}{d(C_X \otimes C_Y)} \right) (x_i, y_j).$$

■ Puente formal entre la formulación de medidas y la matricial.

Discreto:

$$\sum_{i=1}^N \sum_{j=1}^L \left(c_{ij} \pi_{ij} + a_{ij} \pi_{ij}^2 \right).$$

Integral (respecto a $C_X \otimes C_Y$, escribiendo $\psi = \frac{d\pi}{d(C_X \otimes C_Y)}$):

$$\int_{X \times Y} \left[c(x, y) \psi + a(x, y) \psi^2 \right] d[C_X \otimes C_Y].$$

- Contra $C_X \otimes C_Y$ la “integral” es la suma; el término ψ^2 está bien definido puntualmente.

- Reemplazamos C_X, C_Y por Lebesgue $\mathcal{L}_X, \mathcal{L}_Y$. Suponemos $\pi \ll \mathcal{L}_X \otimes \mathcal{L}_Y$ con densidad

$$f = \frac{d\pi}{d(\mathcal{L}_X \otimes \mathcal{L}_Y)} \in L^2(X \times Y).$$

- **Punto clave (R-N):** las marginales del plan también tienen densidad. La marginal $(p_X)_\# \pi$ tiene densidad respecto a $\mathcal{L}_X$

$$F(x) := \int_Y f(x, y) dy = \frac{d(p_X)_\# \pi}{d\mathcal{L}_X}(x), \quad G(y) := \int_X f(x, y) dx.$$

- Por eso penalizar $\|F - f_\mu\|$ compara densidades R-N: $f_\mu = \frac{d\mu}{d\mathcal{L}_X}$,
 $f_\nu = \frac{d\nu}{d\mathcal{L}_Y}$.

Definiciones: el marco funcional

- **Banach:** espacio vectorial normado completo (toda sucesión de Cauchy converge en la norma).
- **Hilbert:** espacio con producto interno, completo bajo la norma inducida $\|x\| = \sqrt{\langle x, x \rangle}$. Todo Hilbert es Banach.
- **Reflexivo:** un Banach E es reflexivo si la inyección canónica $E \hookrightarrow E^{**}$ es sobreyectiva, i.e. $E \cong E^{**}$.
- **Todo Hilbert es reflexivo:** el isomorfismo $E \cong E^*$ (Riesz, inducido por el producto interno) es isométrico y biyectivo $\Rightarrow E \cong E^{**}$.
- Aquí $E = L^2(X \times Y)$ es de Hilbert $\Rightarrow$ reflexivo; y $A = L^2_+$ es convexo y cerrado.

Teorema (Brezis, 2010)

Sea E un Banach reflexivo, $A \subset E$ cerrado y convexo, y $\varphi: A \rightarrow (-\infty, \infty]$ convexa, s.c.i. y no idénticamente $+\infty$. Si además

- A es acotado, o
- $\lim_{\substack{x \in A \\ \|x\| \rightarrow \infty}} \varphi(x) = \infty$ (coercividad),

entonces φ alcanza su mínimo en A : $\exists x^* \in A$ con $\varphi(x^*) = \inf_{x \in A} \varphi(x)$.

Estrategia: verificar convexidad, s.c.i. y coercividad de Φ sobre L_+^2 .

Referencias

Referencias I



Abdulkadiroğlu, A. and Sönmez, T. (2003).
School Choice: A Mechanism Design Approach.
The American Economic Review, 93(3):729–747.



Alman, J., Duan, R., Williams, V. V., Xu, Y., Xu, Z., and Zhou, R. (2025).
More asymmetry yields faster matrix multiplication.
In Proceedings of the 2025 ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 2005–2039, Philadelphia. SIAM.



Anaya-Montes, M. and Gravelle, H. (2024).
Health insurance system fragmentation and COVID-19 mortality: Evidence from Peru.
PLOS ONE, 19(8):e0309531.



Arrieta, A. and Guillén, J. (2017).
Output congestion leads to compromised care in peruvian public hospital neonatal units.
Health Care Management Science, 20(2):209–221.



Bendezu-Quispe, G., Mari-Huarache, L. F., Taype-Rondan, A., Mejia, C. R., and Inga-Berrosapi, F. (2020).
Effect of rural and marginal urban health service on the physicians' perception of primary health care in peru.
Revista Peruana de Medicina Experimental y Salud Pública, 37(4):636–644.



Carlier, G., Dupuy, A., Galichon, A., and Sun, Y. (2020).
SISTA: Learning Optimal Transport Costs under Sparsity Constraints.
arXiv preprint arXiv:2009.08564.
Submitted on 18 Sep 2020, last revised 21 Oct 2020.



Coppersmith, D. and Winograd, S. (1990).
Matrix multiplication via arithmetic progressions.
Journal of Symbolic Computation, 9(3):251–280.



Davie, A. M. and Stothers, A. J. (2013).
Improved bound for complexity of matrix multiplication.
Proceedings of the Royal Society of Edinburgh, Section A: Mathematics, 143(2):351–369.



Dupuy, A. and Galichon, A. (2022).
A Note on the Estimation of Job Amenities and Labor Productivity.
Quantitative Economics, 13:153–177.



Echenique, F., Immorlica, N., and Vazirani, V. V. (2023).
Online and Matching-Based Market Design.
Cambridge University Press.



Echenique, F. and Yenmez, M. B. (2015).
How to Control Controlled School Choice.
The American Economic Review, 105(8):2679–2694.



EsSalud (2025a).
Dashboard de indicadores fonafe y tablero estratégico.



EsSalud (2025b).
Tablero de diferimento de citas.



Gale, D. and Shapley, L. S. (1962).
College Admissions and the Stability of Marriage.
The American Mathematical Monthly, 69(1):9–15.

Referencias III



Galichon, A. (2016).
Optimal Transport Methods in Economics.
Princeton University Press.



Galichon, A. (2021).
The Unreasonable Effectiveness of Optimal Transport in Economics.
Preprint submitted on 12 Jan 2023.



Hochbaum, D. S. and Shanthikumar, J. G. (1990).
Convex separable optimization is not much harder than linear optimization.
Journal of the ACM, 37(4):843–862.



Huang, J. (2018).
Practical fast matrix multiplication algorithms.
Ph.d. dissertation, University of Texas at Austin, Austin, TX.



Izmailov, A. F. and Solodov, M. V. (2023).
Convergence rate estimates for penalty methods revisited.
Computational Optimization and Applications, 85(3):973–992.



Karstadt, E. and Schwartz, O. (2017).
Matrix multiplication, a little faster.
In Proceedings of the 29th ACM Symposium on Parallelism in Algorithms and Architectures (SPAA), pages 101–110, New York. ACM.



Koutis, I., Miller, G. L., and Peng, R. (2012).
A fast solver for a class of linear systems.
Communications of the ACM, 55(10):99–107.



Pan, V. Y. (1982).

Trilinear aggregating with implicit canceling for a new acceleration of matrix multiplication. *Computers & Mathematics with Applications*, 8(1):23–34.



Park, J. and Boyd, S. (2017).

A semidefinite programming method for integer convex quadratic minimization. *Optimization Letters*, 11(4):725–739.



Peng, R. and Vempala, S. S. (2024).

Solving sparse linear systems faster than matrix multiplication. *Communications of the ACM*, 67(7):79–86.



Roth, A. E. (1982).

The Economics of Matching: Stability and Incentives. *Mathematics of Operations Research*, 7(4):617–628.



Roth, A. E. and Sotomayor, M. A. O. (1990).

Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis, volume 18 of *Econometric Society Monographs*. Cambridge University Press.



Soto, A. (2019).

Barreras para una atención eficaz en los hospitales de referencia del ministerio de salud del Perú: atendiendo pacientes en el siglo xxi con recursos del siglo xx. *Revista Peruana de Medicina Experimental y Salud Pública*, 36(2):304.



Strassen, V. (1969).
Gaussian elimination is not optimal.
Numerische Mathematik, 13(4):354–356.



Strassen, V. (1986).
The asymptotic spectrum of tensors and the exponent of matrix multiplication.
In 27th Annual Symposium on Foundations of Computer Science (SFCS 1986), pages 49–54. IEEE.



Velasquez, A. (2020).
Consideraciones éticas del aseguramiento universal de salud en el Perú.
Technical report, Universidad Antonio Ruiz de Montoya, Lima, Perú.



Villani, C. (2009).
Optimal Transport: Old and New, volume 338 of Grundlehren der mathematischen Wissenschaften.
Springer.



Williams, V. V. (2012).
Multiplying matrices faster than coppersmith-winograd.
In Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC), pages 887–898, New York. ACM.



Winograd, S. (1971).
On multiplication of 2×2 matrices.
Linear Algebra and its Applications, 4(4):381–388.



World Bank (2020).

Health at a glance: Latin america and the caribbean 2020.

[https:](https://documents1.worldbank.org/curated/en/383471608633276440/pdf/Health-at-a-Glance-Latin-America-and-the-Caribbean-2020.pdf)

[//documents1.worldbank.org/curated/en/383471608633276440/pdf/Health-at-a-Glance-Latin-America-and-the-Caribbean-2020.pdf](https://documents1.worldbank.org/curated/en/383471608633276440/pdf/Health-at-a-Glance-Latin-America-and-the-Caribbean-2020.pdf).

Accessed: 18 March 2025.