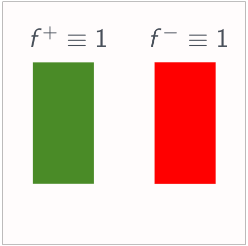
Numerical solution of Optimal Transport Problem on graphs

Enrico Facca, Michele Benzi
8^{th} ECM - June 22^{nd}, 2021

SCUOLA NORMALE SUPERIORE

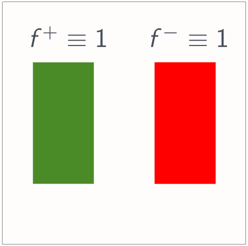
L^1-Optimal Transport Problem (continuous)

$\Omega \subset \mathbb{R}^n$ f^+, f^- two non-negative densities with $\int_\Omega f^+ = \int_\Omega f^-$



L^1-Optimal Transport Problem (continuous)

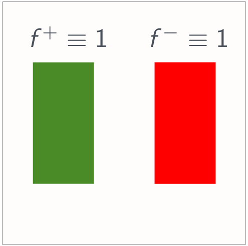
$\Omega \subset \mathbb{R}^n$ f^+, f^- two non-negative densities with $\int_\Omega f^+ = \int_\Omega f^-$

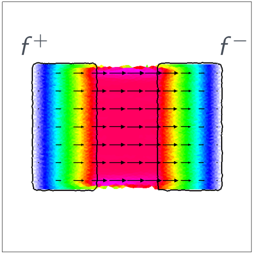


$$\left\{ \begin{array}{c} \min\limits_{\gamma} \displaystyle\int_{\Omega\times\Omega} |x-y| d\gamma(x,y) \\ \gamma \in \mathcal{M}^+(\Omega\times\Omega) \\ (\pi_x)_\# \gamma = f^+\, dx \\ (\pi_y)_\# \gamma = f^-\, dx \end{array} \right\} \text{(Kantorovich P.)}$$

L^1-Optimal Transport Problem (continuous)

$\Omega \subset \mathbb{R}^n$ f^+, f^- two non-negative densities with $\int_\Omega f^+ = \int_\Omega f^-$





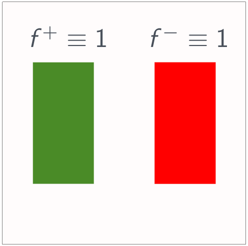
$$\begin{Bmatrix} \min\limits_{\gamma} \int_{\Omega\times\Omega} |x-y| d\gamma(x,y) \\ \gamma \in \mathcal{M}^+(\Omega\times\Omega) \\ (\pi_x)_\# \gamma = f^+\, dx \\ (\pi_y)_\# \gamma = f^-\, dx \end{Bmatrix} \text{(Kantorovich P.)}$$

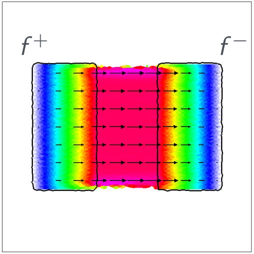
$$= \min_{\mu:\Omega\to\mathbb{R}^+} \left\{ \mathcal{L}(\mu) := \int_\Omega \mu \frac{|\nabla u(\mu)|^2}{2} + \int_\Omega \frac{\mu}{2} \right\}$$

$$u(\mu) := \text{sol: } -\text{div}(\mu \nabla u) = f^+ - f^-$$

L^1-Optimal Transport Problem (continuous)

$\Omega \subset \mathbb{R}^n$ f^+, f^- two non-negative densities with $\int_\Omega f^+ = \int_\Omega f^-$





$$\left\{ \begin{array}{c} \min_{\gamma} \int_{\Omega\times\Omega} |x-y| d\gamma(x,y) \\ \gamma \in \mathcal{M}^+(\Omega\times\Omega) \\ (\pi_x)_{\#}\gamma = f^+\, dx \\ (\pi_y)_{\#}\gamma = f^-\, dx \end{array} \right\} \text{(Kantorovich P.)}$$

$$= \min_{\mu:\Omega\to\mathbb{R}^+} \left\{ \mathcal{L}(\mu) := \int_\Omega \mu \frac{|\nabla u(\mu)|^2}{2} + \int_\Omega \frac{\mu}{2} \right\}$$

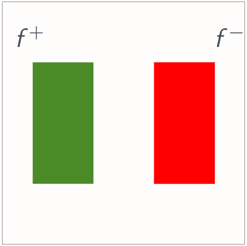
$$u(\mu) := \text{sol: } -\operatorname{div}(\mu \nabla u) = f^+ - f^-$$

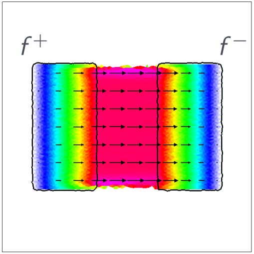
$= W_1(f^+, f^-)$ (Wasserstein 1 distance)

$=$(Distance between non-negative measures on Ω)

L^1-Optimal Transport Problem on graphs

$\Omega \subset \mathbb{R}^n$ $\quad f^+, f^-$ two non-negative densities with $\int_\Omega f^+ = \int_\Omega f^-$





$$\left\{\begin{array}{c} \min_\gamma \int_{\Omega\times\Omega} |x-y| d\gamma(x,y) \\ \gamma \in \mathcal{M}^+(\Omega\times\Omega) \\ (\pi_x)_\#\gamma = f^+\, dx \\ (\pi_y)_\#\gamma = f^-\, dx \end{array}\right\} \text{(Kantorovich P.)}$$

$$= \min_{\mu:\Omega\to\mathbb{R}^+} \left\{ \mathcal{L}(\mu) := \int_\Omega \mu \frac{|\nabla u(\mu)|^2}{2} + \int_\Omega \frac{\mu}{2} \right\}$$

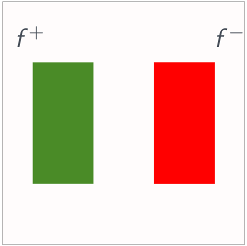
$$u(\mu) := \text{sol: } -\text{div}(\mu \nabla u) = f^+ - f^-$$

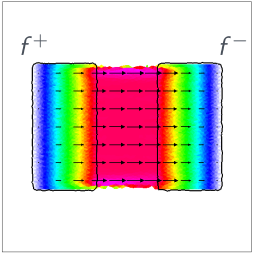
$=W_1(f^+, f^-)$ (Wasserstein 1 distance)

$=$(Distance between non-negative measures on Ω)

L^1-Optimal Transport Problem on graphs

$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{w})$ f^+, f^- two non-negative densities with $\int_\Omega f^+ = \int_\Omega f^-$





$$\left\{ \begin{array}{c} \min_{\gamma} \int_{\Omega\times\Omega} |x-y| d\gamma(x,y) \\ \gamma \in \mathcal{M}^+(\Omega\times\Omega) \\ (\pi_x)_\# \gamma = f^+ \, dx \\ (\pi_y)_\# \gamma = f^- \, dx \end{array} \right\} \text{(Kantorovich P.)}$$

$$= \min_{\mu:\Omega\to\mathbb{R}^+} \left\{ \mathcal{L}(\mu) := \int_\Omega \mu \frac{|\nabla u(\mu)|^2}{2} + \int_\Omega \frac{\mu}{2} \right\}$$

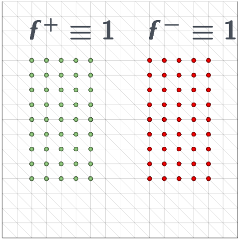
$$u(\mu) := \text{sol: } -\text{div}(\mu \nabla u) = f^+ - f^-$$

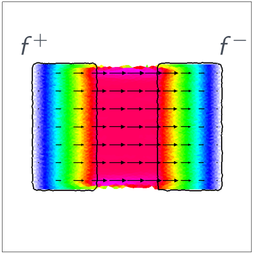
$= W_1(f^+, f^-)$ (Wasserstein 1 distance)

$=$(Distance between non-negative measures on Ω)

L^1-Optimal Transport Problem on graphs

$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{w})$ $\boldsymbol{f}^+, \boldsymbol{f}^- \in \mathbb{R}^n \geq 0$ with $\sum \boldsymbol{f}^+ = \sum \boldsymbol{f}^-$





$$\left\{\begin{array}{c} \min\limits_{\gamma} \displaystyle\int_{\Omega\times\Omega} |x-y| d\gamma(x,y) \\ \gamma \in \mathcal{M}^+(\Omega\times\Omega) \\ (\pi_x)_{\#}\gamma = f^+ \, dx \\ (\pi_y)_{\#}\gamma = f^- \, dx \end{array}\right\} \text{(Kantorovich P.)}$$

$$= \min_{\mu:\Omega\to\mathbb{R}^+} \left\{ \mathcal{L}(\mu) := \int_\Omega \mu \frac{|\nabla u(\mu)|^2}{2} + \int_\Omega \frac{\mu}{2} \right\}$$

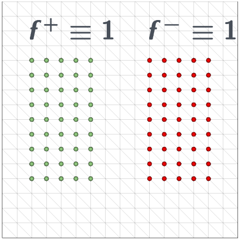
$$u(\mu) := \text{sol: } -\operatorname{div}(\mu \nabla u) = f^+ - f^-$$

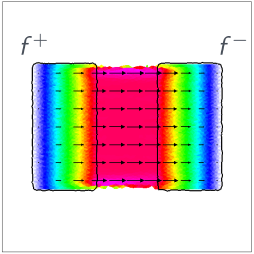
$= W_1(f^+, f^-)$ (Wasserstein 1 distance)

$=$(Distance between non-negative measures on Ω)

L^1-Optimal Transport Problem on graphs

$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{w})$ $\boldsymbol{f}^+, \boldsymbol{f}^- \in \mathbb{R}^n \geq 0$ with $\sum \boldsymbol{f}^+ = \sum \boldsymbol{f}^-$





$\boldsymbol{E}$ = Signed incidence matrix $\boldsymbol{E}\boldsymbol{u} = (\boldsymbol{u}_{n^+} - \boldsymbol{u}_{n^-})$

$-\operatorname{div} \approx \boldsymbol{E}^T$ $\nabla \approx \boldsymbol{G} = \operatorname{diag}(\boldsymbol{w})^{-1}\boldsymbol{E}$

$$\left\{\begin{array}{c} \min_{\gamma} \int_{\Omega\times\Omega} |x-y| d\gamma(x,y) \\ \gamma \in \mathcal{M}^+(\Omega\times\Omega) \\ (\pi_x)_{\#}\gamma = f^+ \, dx \\ (\pi_y)_{\#}\gamma = f^- \, dx \end{array}\right\} \text{(Kantorovich P.)}$$

$$= \min_{\mu:\Omega\to\mathbb{R}^+} \left\{ \mathcal{L}(\mu) := \int_{\Omega} \mu \frac{|\nabla u(\mu)|^2}{2} + \int_{\Omega} \frac{\mu}{2} \right\}$$

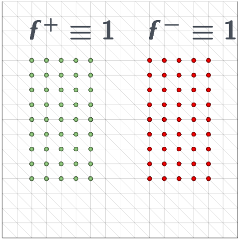
$$u(\mu) := \text{sol: } -\operatorname{div}(\mu \nabla u) = f^+ - f^-$$

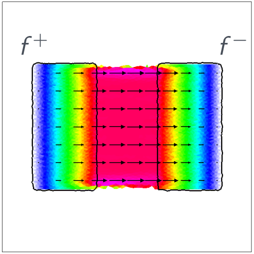
$= W_1(f^+, f^-)$ (Wasserstein 1 distance)

=(Distance between non-negative measures on Ω)

L^1-Optimal Transport Problem on graphs

$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{w})$ $\boldsymbol{f}^+, \boldsymbol{f}^- \in \mathbb{R}^n \geq 0$ with $\sum \boldsymbol{f}^+ = \sum \boldsymbol{f}^-$





$\boldsymbol{E}$ = Signed incidence matrix $\boldsymbol{E}\boldsymbol{u} = (\boldsymbol{u}_{n^+} - \boldsymbol{u}_{n^-})$

$-\operatorname{div} \approx \boldsymbol{E}^T \quad \nabla \approx \boldsymbol{G} = \operatorname{diag}(\boldsymbol{w})^{-1} \boldsymbol{E}$

$$\left\{ \begin{array}{c} \min_{\gamma} \int_{\Omega\times\Omega} |x-y| d\gamma(x,y) \\ \gamma \in \mathcal{M}^+(\Omega\times\Omega) \\ (\pi_x)_{\#}\gamma = f^+ \, dx \\ (\pi_y)_{\#}\gamma = f^- \, dx \end{array} \right\} \text{(Kantorovich P.)}$$

$$= \min_{\mu:\Omega\to\mathbb{R}^+} \left\{ \mathcal{L}(\mu) := \int_\Omega \mu \frac{|\nabla u(\mu)|^2}{2} + \int_\Omega \frac{\mu}{2} \right\}$$

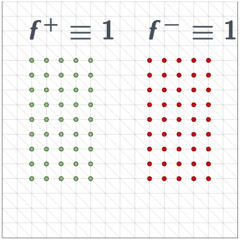
$$u(\mu) := \text{sol: } -\operatorname{div}(\mu \nabla u) = f^+ - f^-$$

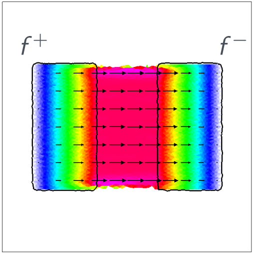
$= W_1(f^+, f^-)$ (Wasserstein 1 distance)

$=$(Distance between non-negative measures on Ω)

L^1-Optimal Transport Problem on graphs

$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{w})$ $\boldsymbol{f}^+, \boldsymbol{f}^- \in \mathbb{R}^n \geq 0$ with $\sum \boldsymbol{f}^+ = \sum \boldsymbol{f}^-$





$\boldsymbol{E}$ = Signed incidence matrix $\boldsymbol{E}\boldsymbol{u} = (\boldsymbol{u}_{n^+} - \boldsymbol{u}_{n^-})$

$-\text{div} \approx \boldsymbol{E}^T \quad \nabla \approx \boldsymbol{G} = \text{diag}(\boldsymbol{w})^{-1}\boldsymbol{E}$

$$\left\{ \begin{array}{c} \min_{\gamma} \sum\sum \text{dist}_{\mathcal{G}}(n_i, n_j)\gamma_{i,j} \\ \gamma \in \mathbb{R}^{n \times n} \geq 0 \\ \mathbf{1}\gamma = \boldsymbol{f}^{+T} \\ \gamma\mathbf{1} = \boldsymbol{f}^- \end{array} \right\} \text{(Kantorovich P.)}$$

$$= \min_{\mu:\Omega \to \mathbb{R}^+} \left\{ \mathcal{L}(\mu) := \int_{\Omega} \mu \frac{|\nabla u(\mu)|^2}{2} + \int_{\Omega} \frac{\mu}{2} \right\}$$

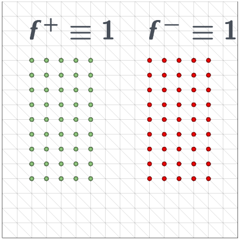
$$u(\mu) := \text{sol:} \ -\text{div}(\mu \nabla u) = f^+ - f^-$$

$= W_1(f^+, f^-)$ (Wasserstein 1 distance)

=(Distance between non-negative measures on Ω)

L^1-Optimal Transport Problem on graphs

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{w}) \quad \boldsymbol{f}^+, \boldsymbol{f}^- \in \mathbb{R}^n \geq 0 \text{ with } \sum \boldsymbol{f}^+ = \sum \boldsymbol{f}^-$$



$\boldsymbol{E}$ = Signed incidence matrix $\quad \boldsymbol{E}\boldsymbol{u} = (\boldsymbol{u}_{n^+} - \boldsymbol{u}_{n^-})$

$-\text{div} \approx \boldsymbol{E}^T \quad \nabla \approx \boldsymbol{G} = \text{diag}(\boldsymbol{w})^{-1}\boldsymbol{E}$

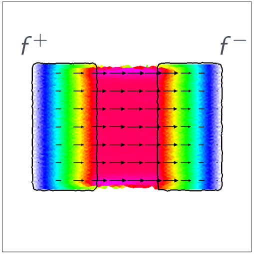
$$\left\{ \begin{array}{c} \min_{\gamma} \sum \sum \text{dist}_{\mathcal{G}}(n_i, n_j)\gamma_{i,j} \\ \boldsymbol{\gamma} \in \mathbb{R}^{n \times n} \geq 0 \\ \mathbf{1}\gamma = \boldsymbol{f}^{+T} \\ \gamma\mathbf{1} = \boldsymbol{f}^- \end{array} \right\} \text{(Kantorovich P.)}$$

$$= \min_{\mu:\Omega \to \mathbb{R}^+} \left\{ \mathcal{L}(\mu) := \int_\Omega \mu \frac{|\nabla u(\mu)|^2}{2} + \int_\Omega \frac{\mu}{2} \right\}$$

$$u(\mu) := \text{sol: } -\text{div}(\mu \nabla u) = f^+ - f^-$$

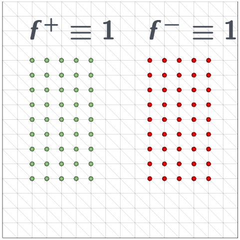
$= W_1(f^+, f^-)$ (Wasserstein 1 distance)

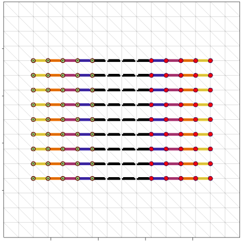
=(Distance between non-negative measures on Ω)



L^1-Optimal Transport Problem on graphs

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{w}) \quad \boldsymbol{f}^+, \boldsymbol{f}^- \in \mathbb{R}^n \geq 0 \text{ with } \sum \boldsymbol{f}^+ = \sum \boldsymbol{f}^-$$





$\boldsymbol{E}$ = Signed incidence matrix $\quad \boldsymbol{E}\boldsymbol{u} = (\boldsymbol{u}_{n^+} - \boldsymbol{u}_{n^-})$

$$-\operatorname{div} \approx \boldsymbol{E}^T \quad \nabla \approx \boldsymbol{G} = \operatorname{diag}(\boldsymbol{w})^{-1} \boldsymbol{E}$$

$$\left\{ \begin{array}{c} \min_{\gamma} \sum \sum \operatorname{dist}_{\mathcal{G}}(n_i, n_j) \gamma_{i,j} \\ \gamma \in \mathbb{R}^{n \times n} \geq 0 \\ \mathbf{1}\gamma = \boldsymbol{f}^{+T} \\ \gamma \mathbf{1} = \boldsymbol{f}^- \end{array} \right\} \text{(Kantorovich P.)}$$

$$= \min_{\boldsymbol{\mu} \in \mathbb{R}^{|\mathcal{E}|} \geq 0} \left\{ \mathcal{L}(\mu) := (\boldsymbol{w} \odot \boldsymbol{\mu})^T \frac{|\boldsymbol{\nabla} u(\boldsymbol{\mu})|^2}{2} + \frac{\boldsymbol{\mu}^T \boldsymbol{w}}{2} \right\}$$

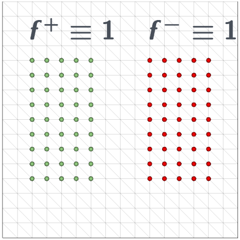
$$u(\boldsymbol{\mu}) := \text{sol: } -\boldsymbol{E} \operatorname{diag}(\boldsymbol{\mu}) \boldsymbol{\nabla} \boldsymbol{u} = \boldsymbol{f}^+ - \boldsymbol{f}^-$$

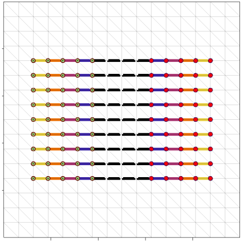
$= W_1(f^+, f^-)$ (Wasserstein 1 distance)

$=$(Distance between non-negative measures on Ω)

L^1-Optimal Transport Problem on graphs

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{w}) \quad \boldsymbol{f}^+, \boldsymbol{f}^- \in \mathbb{R}^n \geq 0 \text{ with } \sum \boldsymbol{f}^+ = \sum \boldsymbol{f}^-$$





$\boldsymbol{E}$ = Signed incidence matrix $\quad \boldsymbol{E}\boldsymbol{u} = (\boldsymbol{u}_{n^+} - \boldsymbol{u}_{n^-})$

$-\operatorname{div} \approx \boldsymbol{E}^T \quad \nabla \approx \boldsymbol{G} = \operatorname{diag}(\boldsymbol{w})^{-1} \boldsymbol{E}$

$$\left\{ \begin{array}{c} \min_{\gamma} \sum \sum \operatorname{dist}_{\mathcal{G}}(n_i, n_j) \gamma_{i,j} \\ \gamma \in \mathbb{R}^{n \times n} \geq 0 \\ \mathbf{1}\gamma = \boldsymbol{f}^{+T} \\ \gamma \mathbf{1} = \boldsymbol{f}^- \end{array} \right\} \text{(Kantorovich P.)}$$

$$= \min_{\boldsymbol{\mu} \in \mathbb{R}^{|\mathcal{E}|}_{\geq 0}} \left\{ \mathcal{L}(\mu) := (\boldsymbol{w} \odot \boldsymbol{\mu})^T \frac{|\boldsymbol{\nabla} u(\boldsymbol{\mu})|^2}{2} + \frac{\boldsymbol{\mu}^T \boldsymbol{w}}{2} \right\}$$

$$u(\boldsymbol{\mu}) := \text{sol: } -\boldsymbol{E} \operatorname{diag}(\boldsymbol{\mu}) \boldsymbol{\nabla} \boldsymbol{u} = \boldsymbol{f}^+ - \boldsymbol{f}^-$$

$= W_1(\boldsymbol{f}^+, \boldsymbol{f}^-)$ (Wasserstein 1 distance)

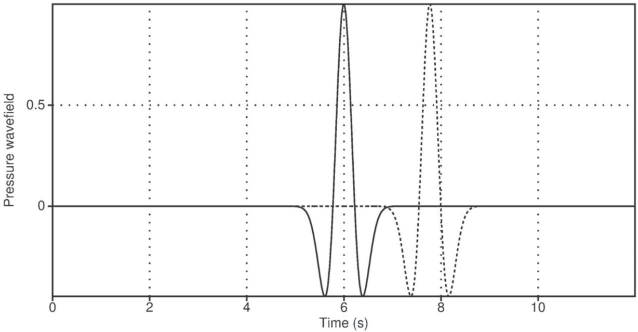
=(Distance between non-negative measures on $\mathcal{G}$)

Wasserstein-1 distance in Inverse Problems

$$\min_{\theta} \|\text{Model}(\theta) - \text{Observ.}\|_{L^2} \qquad \theta = \text{Parameters}$$

Wasserstein-1 distance in Inverse Problems

$$\min_{\theta} \textcolor{red}{W^1}(\text{Model}(\theta), \text{Observ.}) \qquad \theta = \text{Parameters}$$

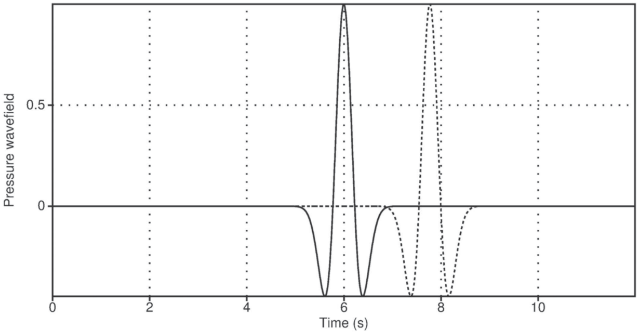


L. Métivier, R. Bros- sier,Q. Mérigot, E. Oudet, J. Virieux, *Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion*, Geophysical Journal International, 2016

Wasserstein-1 distance in Inverse Problems

$$\min_{\theta} W^1(\text{Model}(\theta), \text{Observ.}) \qquad \theta = \text{Parameters}$$

$$\frac{\partial}{\partial f^+} W^1(f^+, f^-) = u(\mu^*)$$

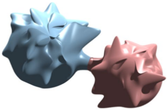


L. Métivier, R. Bros- sier,Q. Mérigot, E. Oudet, J. Virieux, *Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion*, Geophysical Journal International, 2016

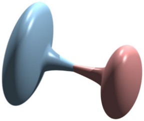
Oliver-Ricci Flow

Reimannian Manifold (M, g)

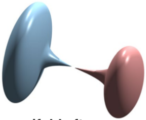
$$\partial_t g = -\text{Ricci}(g)$$



(a) initial manifold



(b) manifold after Ricci flow



(c) manifold after surgery

Ni, CC., Lin, YY., Luo, F. et al. Community Detection on Networks with Ricci Flow. Sci Rep 9, (2019).

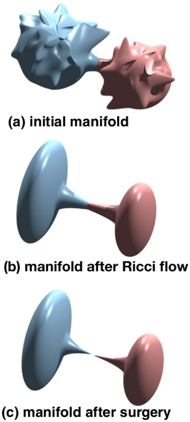
Oliver-Ricci Flow

Reimannian Manifold (M, g)

$$\partial_t g = -\text{Ricci}(g)$$

Weighted Graph$(\mathcal{G}, \boldsymbol{w})$

$$\partial_t w = -????$$



Ni, CC., Lin, YY., Luo, F. et al. Community Detection on Networks with Ricci Flow. Sci Rep 9, (2019).

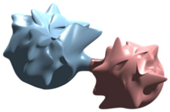
Oliver-Ricci Flow

Reimannian Manifold (M, g)

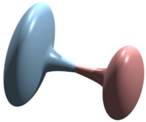
$$\partial_t g = -\text{Ricci}(g) \approx \text{Ricci}(W^1)$$

Weighted Graph$(\mathcal{G}, \boldsymbol{w})$

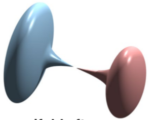
$$\partial_t w = -????$$



(a) initial manifold



(b) manifold after Ricci flow

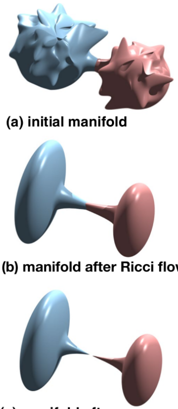


(c) manifold after surgery

Ni, CC., Lin, YY., Luo, F. et al. Community Detection on Networks with Ricci Flow. Sci Rep 9, (2019).

Oliver-Ricci Flow

Reimannian Manifold (M, g)	Weighted Graph$(\mathcal{G}, \boldsymbol{w})$
$\partial_t g = -\text{Ricci}(g) \approx \text{Ricci}(W^1)$	$\partial_t w = -\ \text{Ricci}(W^1)$



(a) initial manifold

(b) manifold after Ricci flow

(c) manifold after surgery

Ni, CC., Lin, YY., Luo, F. et al. Community Detection on Networks with Ricci Flow. Sci Rep 9, (2019).

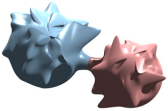
Oliver-Ricci Flow

Reimannian Manifold (M, g)

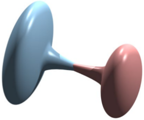
$$\partial_t g = -\text{Ricci}(g) \approx \text{Ricci}(W^1)$$

Weighted Graph$(\mathcal{G}, \boldsymbol{w})$

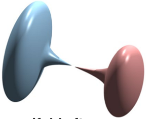
$$\partial_t w = -\ \text{Ricci}(W^1)$$



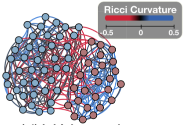
(a) initial manifold



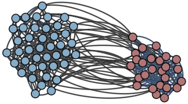
(b) manifold after Ricci flow



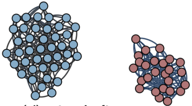
(c) manifold after surgery



(a') initial network



(b') network after Ricci flow



(c') network after surgery

Ni, CC., Lin, YY., Luo, F. et al. Community Detection on Networks with Ricci Flow. Sci Rep 9, (2019).

Take home message: the problem

- The Optimal Transport Problem of moving continuous densities f^+ and f^- with cost equal to Euclidean Distance

$$\min_{\mu:\Omega\to\mathbb{R}^+} \left\{ \mathcal{L}(\mu) := \int_\Omega \mu \frac{|\nabla u(\mu)|^2}{2} + \int_\Omega \frac{\mu}{2} \right\}$$

- The Optimal Transport Problem $\boldsymbol{f}^+$ and $\boldsymbol{f}^-$ defined on graphs with cost equal to the Shortest Path Distance

$$\min_{\boldsymbol{\mu}\in\mathbb{R}^{|\mathcal{E}|}_{\geq 0}} \left\{ \mathcal{L}(\boldsymbol{\mu}) := (\boldsymbol{w}\odot\boldsymbol{\mu})^T \frac{|\nabla u(\boldsymbol{\mu})|^2}{2} + \frac{\boldsymbol{\mu}^T\boldsymbol{w}}{2} \right\}$$

- The resulting minimum defines a useful distance on positive measure, the Wasserstein-1 distance

Take home message: the problem

- The Optimal Transport Problem of moving continuous densities f^+ and f^- with cost equal to Euclidean Distance

$$\min_{\mu:\Omega\to\mathbb{R}^+} \left\{ \mathcal{L}(\mu) := \int_\Omega \mu \frac{|\nabla u(\mu)|^2}{2} + \int_\Omega \frac{\mu}{2} \right\}$$

- The Optimal Transport Problem $\boldsymbol{f}^+$ and $\boldsymbol{f}^-$ defined on graphs with cost equal to the Shortest Path Distance

$$\min_{\boldsymbol{\mu}\in\mathbb{R}^{|\mathcal{E}|}_{\geq 0}} \left\{ \mathcal{L}(\mu) := (\boldsymbol{w} \odot \boldsymbol{\mu})^T \frac{|\boldsymbol{\nabla} u(\boldsymbol{\mu})|^2}{2} + \frac{\boldsymbol{\mu}^T \boldsymbol{w}}{2} \right\}$$

- The resulting minimum defines a useful distance on positive measure, the Wasserstein-1 distance

Take home message: the problem

- The Optimal Transport Problem of moving continuous densities f^+ and f^- with cost equal to Euclidean Distance

$$\min_{\mu:\Omega\to\mathbb{R}^+} \left\{ \mathcal{L}(\mu) := \int_\Omega \mu \frac{|\nabla u(\mu)|^2}{2} + \int_\Omega \frac{\mu}{2} \right\}$$

- The Optimal Transport Problem $\boldsymbol{f}^+$ and $\boldsymbol{f}^-$ defined on graphs with cost equal to the Shortest Path Distance

$$\min_{\boldsymbol{\mu}\in\mathbb{R}^{|\mathcal{E}|}_{\geq 0}} \left\{ \mathcal{L}(\mu) := (\boldsymbol{w} \odot \boldsymbol{\mu})^T \frac{|\boldsymbol{\nabla} u(\boldsymbol{\mu})|^2}{2} + \frac{\boldsymbol{\mu}^T \boldsymbol{w}}{2} \right\}$$

- The resulting minimum defines a useful distance on positive measure, the Wasserstein-1 distance

Gradient Descent Dynamics

$$\inf_{\boldsymbol{\mu}} \left\{ \mathcal{L}(\mu) = (\boldsymbol{w} \odot \boldsymbol{\mu})^T \frac{|\nabla u(\boldsymbol{\mu})|^2}{2} + \frac{\boldsymbol{\mu}^T \boldsymbol{w}}{2} \ : \ \boldsymbol{\mu} \geq 0 \right\}$$

- **Gradient Descent Dynamics**
  $\partial_t \boldsymbol{\mu}(t) = -\nabla \mathcal{L}(\boldsymbol{\mu}(t))$ Projected gradient/Small Time-step
- **Bio-inspired Dynamics (Mirror gradient descent)**
  $\partial_t \boldsymbol{\mu}(t) = -\boldsymbol{\mu}(t) \nabla_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{\mu}(t))$ $\boldsymbol{\mu}(t) \geq \exp^{-t} \min \boldsymbol{\mu}(0)$
  Positiveness preserved
- **Gradient Descent Dynamics**

$$\boldsymbol{\mu} = \Psi(\boldsymbol{\sigma}) = \boldsymbol{\sigma}^2/2 \quad \text{Positiveness preserved}$$

$$\partial_t \boldsymbol{\sigma}(t) = -\nabla_{\boldsymbol{\sigma}} \mathcal{L}(\Psi(\boldsymbol{\sigma}(t))) \quad \text{``Classical GD''}$$

Gradient Descent Dynamics

$$\inf_{\mu} \left\{ \mathcal{L}(\mu) = (\boldsymbol{w} \odot \boldsymbol{\mu})^T \frac{|\nabla u(\boldsymbol{\mu})|^2}{2} + \frac{\boldsymbol{\mu}^T \boldsymbol{w}}{2} \;:\; \boldsymbol{\mu} \geq 0 \right\}$$

- **Gradient Descent Dynamics**
  $\partial_t \boldsymbol{\mu}(t) = -\nabla \mathcal{L}(\boldsymbol{\mu}(t))$ Projected gradient/Small Time-step
- **Bio-inspired Dynamics (Mirror gradient descent)**
  $\partial_t \boldsymbol{\mu}(t) = -\boldsymbol{\mu}(t) \nabla_{\mu} \mathcal{L}(\boldsymbol{\mu}(t))$ $\boldsymbol{\mu}(t) \geq \exp^{-t} \min \boldsymbol{\mu}(0)$
  Positiveness preserved
- **Gradient Descent Dynamics**

$$\boldsymbol{\mu} = \Psi(\boldsymbol{\sigma}) = \boldsymbol{\sigma}^2/2 \quad \text{Positiveness preserved}$$

$$\partial_t \boldsymbol{\sigma}(t) = -\nabla_{\sigma} \mathcal{L}(\Psi(\boldsymbol{\sigma}(t))) \quad \text{“Classical GD”}$$

Gradient Descent Dynamics

$$\inf_{\boldsymbol{\mu}} \left\{ \mathcal{L}(\mu) = (\boldsymbol{w} \odot \boldsymbol{\mu})^T \frac{|\nabla u(\boldsymbol{\mu})|^2}{2} + \frac{\boldsymbol{\mu}^T \boldsymbol{w}}{2} \; : \; \boldsymbol{\mu} \geq 0 \right\}$$

- **Gradient Descent Dynamics**
  $\partial_t \boldsymbol{\mu}(t) = -\nabla \mathcal{L}(\boldsymbol{\mu}(t))$ Projected gradient/Small Time-step
- **Bio-inspired Dynamics (Mirror gradient descent)**
  $\partial_t \boldsymbol{\mu}(t) = -\boldsymbol{\mu}(t) \nabla_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{\mu}(t))$ $\boldsymbol{\mu}(t) \geq \exp^{-t} \min \boldsymbol{\mu}(0)$
  Positiveness preserved
- **Gradient Descent Dynamics**

$$\boldsymbol{\mu} = \Psi(\boldsymbol{\sigma}) = \boldsymbol{\sigma}^2/2 \quad \text{Positiveness preserved}$$

$$\partial_t \boldsymbol{\sigma}(t) = -\nabla_{\boldsymbol{\sigma}} \mathcal{L}(\Psi(\boldsymbol{\sigma}(t))) \quad \text{“Classical GD”}$$

Gradient Descent Dynamics

$$\inf_{\boldsymbol{\mu}} \left\{ \mathcal{L}(\mu) = (\boldsymbol{w} \odot \boldsymbol{\mu})^T \frac{|\nabla u(\boldsymbol{\mu})|^2}{2} + \frac{\boldsymbol{\mu}^T \boldsymbol{w}}{2} \ : \ \boldsymbol{\mu} \geq 0 \right\}$$

- **Gradient Descent Dynamics**
  $\partial_t \boldsymbol{\mu}(t) = -\nabla \mathcal{L}(\boldsymbol{\mu}(t))$ Projected gradient/Small Time-step
- **Bio-inspired Dynamics (Mirror gradient descent)**
  $\partial_t \boldsymbol{\mu}(t) = -\boldsymbol{\mu}(t) \nabla_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{\mu}(t))$ $\mu(t) \geq \exp^{-t} \min \mu(0)$
  Positiveness preserved
- **Gradient Descent Dynamics**

$$\mu = \Psi(\boldsymbol{\sigma}) = \boldsymbol{\sigma}^2/2 \quad \text{Positiveness preserved}$$

$$\partial_t \boldsymbol{\sigma}(t) = -\nabla_{\sigma} \mathcal{L}(\Psi(\boldsymbol{\sigma}(t))) \quad \text{“Classical GD”}$$

Gradient Descent Dynamics

$$\inf_{\boldsymbol{\mu}} \left\{ \mathcal{L}(\mu) = (\boldsymbol{w} \odot \boldsymbol{\mu})^T \frac{|\nabla u(\boldsymbol{\mu})|^2}{2} + \frac{\boldsymbol{\mu}^T \boldsymbol{w}}{2} \;:\; \boldsymbol{\mu} \geq 0 \right\}$$

- **Gradient Descent Dynamics**
  $\partial_t \boldsymbol{\mu}(t) = -\nabla \mathcal{L}(\boldsymbol{\mu}(t))$ Projected gradient/Small Time-step
- **Bio-inspired Dynamics (Mirror gradient descent)**
  $\partial_t \boldsymbol{\mu}(t) = -\boldsymbol{\mu}(t) \nabla_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{\mu}(t))$ $\boldsymbol{\mu}(t) \geq \exp^{-t} \min \boldsymbol{\mu}(0)$
  Positiveness preserved
- **Gradient Descent Dynamics**

$$\mu = \Psi(\boldsymbol{\sigma}) = \boldsymbol{\sigma}^2/2 \quad \text{Positiveness preserved}$$

$$\partial_t \boldsymbol{\sigma}(t) = -\nabla_{\sigma} \mathcal{L}(\Psi(\boldsymbol{\sigma}(t))) \quad \text{“Classical GD”}$$

Gradient Descent Dynamics

$$\inf_{\boldsymbol{\mu}} \left\{ \mathcal{L}(\mu) = (\boldsymbol{w} \odot \boldsymbol{\mu})^T \frac{|\boldsymbol{\nabla} u(\boldsymbol{\mu})|^2}{2} + \frac{\boldsymbol{\mu}^T \boldsymbol{w}}{2} \;:\; \boldsymbol{\mu} \geq 0 \right\}$$

- **Gradient Descent Dynamics**
  $\partial_t \boldsymbol{\mu}(t) = -\nabla \mathcal{L}(\boldsymbol{\mu}(t))$ Projected gradient/Small Time-step
- **Bio-inspired Dynamics (Mirror gradient descent)**
  $\partial_t \boldsymbol{\mu}(t) = -\boldsymbol{\mu}(t) \nabla_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{\mu}(t))$ $\boldsymbol{\mu}(t) \geq \exp^{-t} \min \boldsymbol{\mu}(0)$
  Positiveness preserved
- **Gradient Descent Dynamics**

$$\boldsymbol{\mu} = \Psi(\boldsymbol{\sigma}) = \boldsymbol{\sigma}^2/2 \quad \text{Positiveness preserved}$$

$$\partial_t \boldsymbol{\sigma}(t) = -\nabla_{\boldsymbol{\sigma}} \mathcal{L}(\Psi(\boldsymbol{\sigma}(t))) \quad \text{“Classical GD”}$$

Time Discretization

$$\text{DAE} \begin{cases} \mathbf{div}\, \mathrm{diag}\left(\boldsymbol{\sigma}^2/2\right) \boldsymbol{\nabla} \cdot \boldsymbol{u}(t) = \boldsymbol{f} \\ \boldsymbol{\sigma}'(t) = -\nabla_{\boldsymbol{\sigma}} \mathcal{L}(\Psi(\boldsymbol{\sigma}(t))) \\ \boldsymbol{\sigma}(0) = \boldsymbol{\sigma}^0 > 0 \end{cases} \qquad t^{k+1} = t^k + \Delta t^k$$

- **Explicit Euler (Gradient Descent)** **Accelerated Gradient Descent**
  One weighted Laplacian system $\mathbf{div}\, \mathrm{diag}\left(\tilde{\mu}\right) \boldsymbol{\nabla} \boldsymbol{x} = \boldsymbol{b}$ per iteration
  Time step size limitation
- **Implicit Euler + Inexact Newton**

$$\begin{pmatrix} \mathbf{div}\, \mathrm{diag}\left(\mu\right) \boldsymbol{\nabla} & \mathbf{div}\, \boldsymbol{D}_1 \\ \boldsymbol{D}_1 \boldsymbol{\nabla} & -\frac{1}{\Delta t}\boldsymbol{I} + \boldsymbol{D}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{s}_1^m \\ \boldsymbol{s}_2^m \end{pmatrix} = \begin{pmatrix} -\boldsymbol{F}_1 \\ -\boldsymbol{F}_2 \end{pmatrix}$$

$$\mathbf{div}\, \mathrm{diag}\left(\tilde{\mu}\right) \boldsymbol{\nabla} \boldsymbol{s}_1^m = \boldsymbol{b}$$

More than 1 inexact weighted Laplacian system per iteration
Larger time step size

Time Discretization

$$\text{DAE} \begin{cases} \textbf{div}\,\text{diag}\left(\boldsymbol{\sigma}^2/2\right)\boldsymbol{\nabla}\cdot\boldsymbol{u}(t) = \boldsymbol{f} \\ \boldsymbol{\sigma}'(t) = -\nabla_{\boldsymbol{\sigma}}\mathcal{L}(\Psi(\boldsymbol{\sigma}(t))) \\ \boldsymbol{\sigma}(0) = \boldsymbol{\sigma}^0 > 0 \end{cases} \qquad t^{k+1} = t^k + \Delta t^k$$

- **Explicit Euler (Gradient Descent)** **Accelerated Gradient Descent**
  One weighted Laplacian system $\textbf{div}\,\text{diag}\,(\tilde{\mu})\,\boldsymbol{\nabla}\boldsymbol{x} = \boldsymbol{b}$ per iteration
  Time step size limitation
- **Implicit Euler + Inexact Newton**

$$\begin{pmatrix} \textbf{div}\,\text{diag}\,(\mu)\,\boldsymbol{\nabla} & \textbf{div}\,\boldsymbol{D}_1 \\ \boldsymbol{D}_1\boldsymbol{\nabla} & -\frac{1}{\Delta t}\boldsymbol{I} + \boldsymbol{D}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{s}_1^m \\ \boldsymbol{s}_2^m \end{pmatrix} = \begin{pmatrix} -\boldsymbol{F}_1 \\ -\boldsymbol{F}_2 \end{pmatrix}$$

$$\textbf{div}\,\text{diag}\,(\tilde{\mu})\,\nabla\boldsymbol{s}_1^m = \boldsymbol{b}$$

More than 1 inexact weighted Laplacian system per iteration
Larger time step size

Time Discretization

$$\text{DAE} \begin{cases} \textbf{div}\,\text{diag}\left(\boldsymbol{\sigma}^2/2\right) \boldsymbol{\nabla} \cdot \boldsymbol{u}(t) = \boldsymbol{f} \\ \boldsymbol{\sigma}'(t) = -\nabla_{\boldsymbol{\sigma}} \mathcal{L}(\Psi(\boldsymbol{\sigma}(t))) \\ \boldsymbol{\sigma}(0) = \boldsymbol{\sigma}^0 > 0 \end{cases} \qquad t^{k+1} = t^k + \Delta t^k$$

- **Explicit Euler (Gradient Descent)** **Accelerated Gradient Descent**
  One weighted Laplacian system $\textbf{div}\,\text{diag}\left(\tilde{\mu}\right) \boldsymbol{\nabla} \boldsymbol{x} = \boldsymbol{b}$ per iteration
  Time step size limitation
- **Implicit Euler + Inexact Newton**

$$\begin{pmatrix} \textbf{div}\,\text{diag}\left(\mu\right) \boldsymbol{\nabla} & \textbf{div}\,\boldsymbol{D}_1 \\ \boldsymbol{D}_1 \boldsymbol{\nabla} & -\frac{1}{\Delta t} \boldsymbol{I} + \boldsymbol{D}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{s}_1^m \\ \boldsymbol{s}_2^m \end{pmatrix} = \begin{pmatrix} -\boldsymbol{F}_1 \\ -\boldsymbol{F}_2 \end{pmatrix}$$

$$\textbf{div}\,\text{diag}\left(\tilde{\mu}\right) \boldsymbol{\nabla} \boldsymbol{s}_1^m = \boldsymbol{b}$$

More than 1 inexact weighted Laplacian system per iteration
Larger time step size

Time Discretization

$$\text{DAE}\begin{cases} \mathbf{div}\,\mathrm{diag}\left(\sigma^2/2\right)\boldsymbol{\nabla}\cdot \boldsymbol{u}(t) = \boldsymbol{f} \\ \sigma'(t) = -\nabla_\sigma \mathcal{L}(\Psi(\sigma(t))) \\ \sigma(0) = \sigma^0 > 0 \end{cases} \qquad t^{k+1} = t^k + \Delta t^k$$

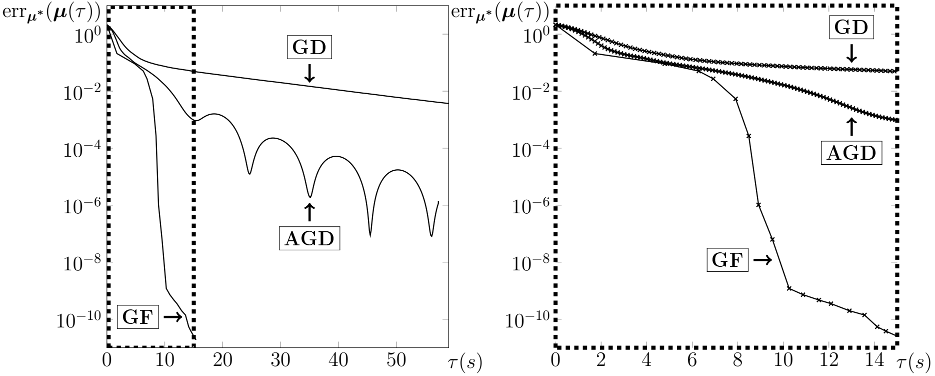
- **Explicit Euler (Gradient Descent)** **Accelerated Gradient Descent**
  One weighted Laplacian system $\mathbf{div}\,\mathrm{diag}\,(\tilde{\mu})\,\boldsymbol{\nabla}\boldsymbol{x} = \boldsymbol{b}$ per iteration
  Time step size limitation
- **Implicit Euler + Inexact Newton**

$$\begin{pmatrix} \mathbf{div}\,\mathrm{diag}\,(\mu)\,\boldsymbol{\nabla} & \mathbf{div}\,D_1 \\ D_1\boldsymbol{\nabla} & -\frac{1}{\Delta t}I + D_2 \end{pmatrix}\begin{pmatrix} s_1^m \\ s_2^m \end{pmatrix} = \begin{pmatrix} -F_1 \\ -F_2 \end{pmatrix}$$

$$\mathbf{div}\,\mathrm{diag}\,(\tilde{\mu})\,\boldsymbol{\nabla} s_1^m = \boldsymbol{b}$$

More than 1 inexact weighted Laplacian system per iteration
Larger time step size

Comparison - Errors vs CPU-time



$\approx 10^5$ nodes/edges

Computational Cost - Scaling with graphs' size

"Grid-like" graphs (only "local" connection)

Graphs $\mathcal{G}$	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{E}\rvert$	Time steps	Newton steps	AGMG Iters.	CPU (s)	Errors $\text{err}_{\boldsymbol{u}^*}(\hat{\boldsymbol{u}})$
$\mathcal{G}_0$	$1.1 \cdot 10^3$	$3.1 \cdot 10^3$	9	29	335	0.0963	3.3e-15
$\mathcal{G}_1$	$4.2 \cdot 10^3$	$1.2 \cdot 10^4$	6	25	359	0.226	2.7e-13
$\mathcal{G}_2$	$1.7 \cdot 10^4$	$4.9 \cdot 10^4$	7	26	399	0.902	9.0e-14
$\mathcal{G}_3$	$6.6 \cdot 10^4$	$2.0 \cdot 10^5$	7	28	456	4.43	3.3e-15
$\mathcal{G}_4$	$2.6 \cdot 10^5$	$7.9 \cdot 10^5$	7	31	545	24.2	6.7e-16
$\mathcal{G}_5$	$1.1 \cdot 10^6$	$3.1 \cdot 10^6$	8	34	646	136	5.0e-14

#Newton-step $\approx \mathcal{O}(1)$ CPU $\approx \mathcal{O}(|\mathcal{E}|^{0.11})$

AGMG=**AG**gregation-based algebraic **M**ulti**G**rid (Notay et al.)

"Watts-Strogatz " graphs

#Newton-step $\approx \mathcal{O}(|\mathcal{E}|^{0.31})/\mathcal{O}(|\mathcal{E}|^{0.36})$ CPU $\approx \mathcal{O}(|\mathcal{E}|^{1.14})/\mathcal{O}(|\mathcal{E}|^{1.51})$

LAMG=**L**ean **A**lgebraic **M**ulti**G**rid (Livne et al.)

Computational Cost - Scaling with graphs' size

"Grid-like" graphs (only "local" connection)

Graphs $\mathcal{G}$	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{E}\rvert$	Time steps	Newton steps	AGMG Iters.	CPU (s)	Errors $\mathrm{err}_{\boldsymbol{u}^*}(\hat{\boldsymbol{u}})$
$\mathcal{G}_0$	$1.1 \cdot 10^3$	$3.1 \cdot 10^3$	9	29	335	0.0963	3.3e-15
$\mathcal{G}_1$	$4.2 \cdot 10^3$	$1.2 \cdot 10^4$	6	25	359	0.226	2.7e-13
$\mathcal{G}_2$	$1.7 \cdot 10^4$	$4.9 \cdot 10^4$	7	26	399	0.902	9.0e-14
$\mathcal{G}_3$	$6.6 \cdot 10^4$	$2.0 \cdot 10^5$	7	28	456	4.43	3.3e-15
$\mathcal{G}_4$	$2.6 \cdot 10^5$	$7.9 \cdot 10^5$	7	31	545	24.2	6.7e-16
$\mathcal{G}_5$	$1.1 \cdot 10^6$	$3.1 \cdot 10^6$	8	34	646	136	5.0e-14

$\#$Newton-step $\approx \mathcal{O}(1)$ CPU $\approx \mathcal{O}(\lvert\mathcal{E}\rvert^{0.11})$

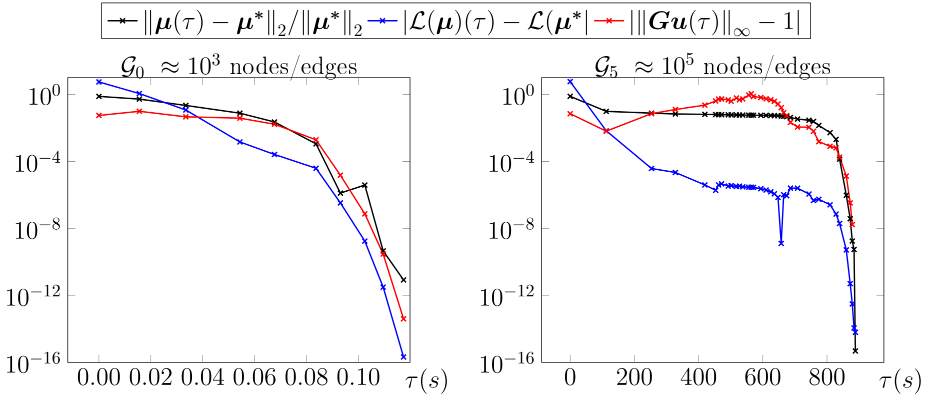
AGMG=**AG**gregation-based algebraic **M**ulti**G**rid (Notay et al.)

"Watts-Strogatz " graphs

$\#$Newton-step $\approx \mathcal{O}(\lvert\mathcal{E}\rvert^{0.31})/\mathcal{O}(\lvert\mathcal{E}\rvert^{0.36})$ CPU $\approx \mathcal{O}(\lvert\mathcal{E}\rvert^{1.14})/\mathcal{O}(\lvert\mathcal{E}\rvert^{1.51})$

LAMG=**L**ean **A**lgebraic **M**ulti**G**rid (Livne et al.)

Computational Cost - Errors evolution



Take home messages on the numerical solution

- Using the $\boldsymbol{\mu} = \boldsymbol{\sigma}^2/2$ positiveness is preserved and purely Gradient Descent approaches can be used
- Explicit (or 1^{st}-order) schemes (GD or AGD) are cheap but requires many iterations to converge
- Implicit Euler + Inexact Newton (2^{nd}-order schemes) more expensive requires less time iterations
- "And the winner is": **Implicit Euler + Inexact Newton**
- The weighted Laplacian systems are solved efficiently by algebraic Multigrid Solvers (AGMG and LAMG)
- Selection of "active" edges can drastically improve efficiency

The paper: E. Facca and M. Benzi, "Fast iterative solution of the optimal transport problem on graphs" SIAM J. Sci. Comput. (2021).

The code : `https://gitlab.com/enrico_facca/dmk_solver`

Take home messages on the numerical solution

- Using the $\mu = \sigma^2/2$ positiveness is preserved and purely Gradient Descent approaches can be used
- Explicit (or 1^{st}-order) schemes (GD or AGD) are cheap but requires many iterations to converge
- Implicit Euler + Inexact Newton (2^{nd}-order schemes) more expensive requires less time iterations
- "And the winner is": **Implicit Euler + Inexact Newton**
- The weighted Laplacian systems are solved efficiently by algebraic Multigrid Solvers (AGMG and LAMG)
- Selection of "active" edges can drastically improve efficiency

The paper: E. Facca and M. Benzi, "Fast iterative solution of the optimal transport problem on graphs" SIAM J. Sci. Comput. (2021).

The code : `https://gitlab.com/enrico_facca/dmk_solver`

Take home messages on the numerical solution

- Using the $\mu = \sigma^2/2$ positiveness is preserved and purely Gradient Descent approaches can be used
- Explicit (or 1^{st}-order) schemes (GD or AGD) are cheap but requires many iterations to converge
- Implicit Euler + Inexact Newton (2^{nd}-order schemes) more expensive requires less time iterations
- "And the winner is": **Implicit Euler + Inexact Newton**
- The weighted Laplacian systems are solved efficiently by algebraic Multigrid Solvers (AGMG and LAMG)
- Selection of "active" edges can drastically improve efficiency

The paper: E. Facca and M. Benzi, "Fast iterative solution of the optimal transport problem on graphs" SIAM J. Sci. Comput. (2021).

The code : `https://gitlab.com/enrico_facca/dmk_solver`

Take home messages on the numerical solution

- Using the $\mu = \sigma^2/2$ positiveness is preserved and purely Gradient Descent approaches can be used
- Explicit (or 1^{st}-order) schemes (GD or AGD) are cheap but requires many iterations to converge
- Implicit Euler + Inexact Newton (2^{nd}-order schemes) more expensive requires less time iterations
- "And the winner is": **Implicit Euler + Inexact Newton**
- The weighted Laplacian systems are solved efficiently by algebraic Multigrid Solvers (AGMG and LAMG)
- Selection of "active" edges can drastically improve efficiency

The paper: E. Facca and M. Benzi, "Fast iterative solution of the optimal transport problem on graphs" SIAM J. Sci. Comput. (2021).

The code : `https://gitlab.com/enrico_facca/dmk_solver`

Take home messages on the numerical solution

- Using the $\boldsymbol{\mu} = \boldsymbol{\sigma}^2/2$ positiveness is preserved and purely Gradient Descent approaches can be used
- Explicit (or 1^{st}-order) schemes (GD or AGD) are cheap but requires many iterations to converge
- Implicit Euler + Inexact Newton (2^{nd}-order schemes) more expensive requires less time iterations
- "And the winner is": **Implicit Euler + Inexact Newton**
- The weighted Laplacian systems are solved efficiently by algebraic Multigrid Solvers (AGMG and LAMG)
- Selection of "active" edges can drastically improve efficiency

The paper: E. Facca and M. Benzi, "Fast iterative solution of the optimal transport problem on graphs" SIAM J. Sci. Comput. (2021).

The code : `https://gitlab.com/enrico_facca/dmk_solver`

Take home messages on the numerical solution

- Using the $\mu = \sigma^2/2$ positiveness is preserved and purely Gradient Descent approaches can be used
- Explicit (or 1^{st}-order) schemes (GD or AGD) are cheap but requires many iterations to converge
- Implicit Euler + Inexact Newton (2^{nd}-order schemes) more expensive requires less time iterations
- "And the winner is": **Implicit Euler + Inexact Newton**
- The weighted Laplacian systems are solved efficiently by algebraic Multigrid Solvers (AGMG and LAMG)
- Selection of "active" edges can drastically improve efficiency

The paper: E. Facca and M. Benzi, "Fast iterative solution of the optimal transport problem on graphs" SIAM J. Sci. Comput. (2021).

The code : `https://gitlab.com/enrico_facca/dmk_solver`

Take home messages on the numerical solution

- Using the $\boldsymbol{\mu} = \boldsymbol{\sigma}^2/2$ positiveness is preserved and purely Gradient Descent approaches can be used
- Explicit (or 1^{st}-order) schemes (GD or AGD) are cheap but requires many iterations to converge
- Implicit Euler + Inexact Newton (2^{nd}-order schemes) more expensive requires less time iterations
- "And the winner is": **Implicit Euler + Inexact Newton**
- The weighted Laplacian systems are solved efficiently by algebraic Multigrid Solvers (AGMG and LAMG)
- Selection of "active" edges can drastically improve efficiency

The paper: E. Facca and M. Benzi, "Fast iterative solution of the optimal transport problem on graphs" SIAM J. Sci. Comput. (2021).

The code : `https://gitlab.com/enrico_facca/dmk_solver`