Constrained Optimal Transport Maps

Eloi Tanguy, Agnès Desolneux and Julie Delon

MAP5, Université Paris-Cité

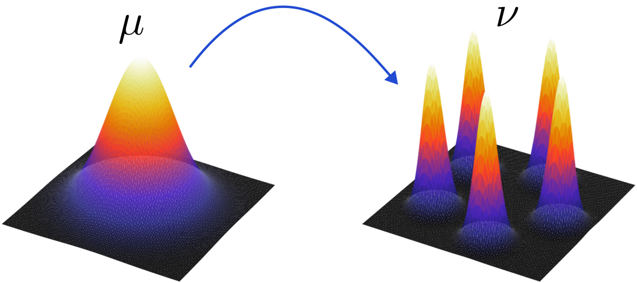
17th February 2025

1. Transporting Measures

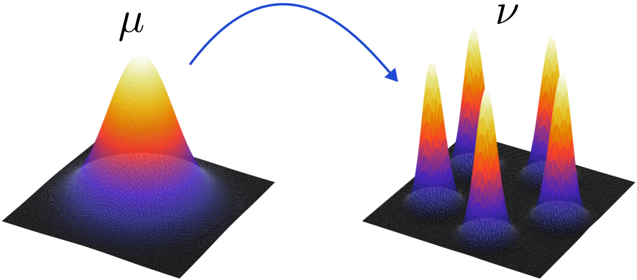
2. Constrained Approximate Transport Maps

3. Zoom on the L^2 Case on on $\mathbb{R}^d$

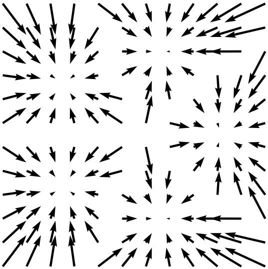
Push-Forward Measures



Push-Forward Measures

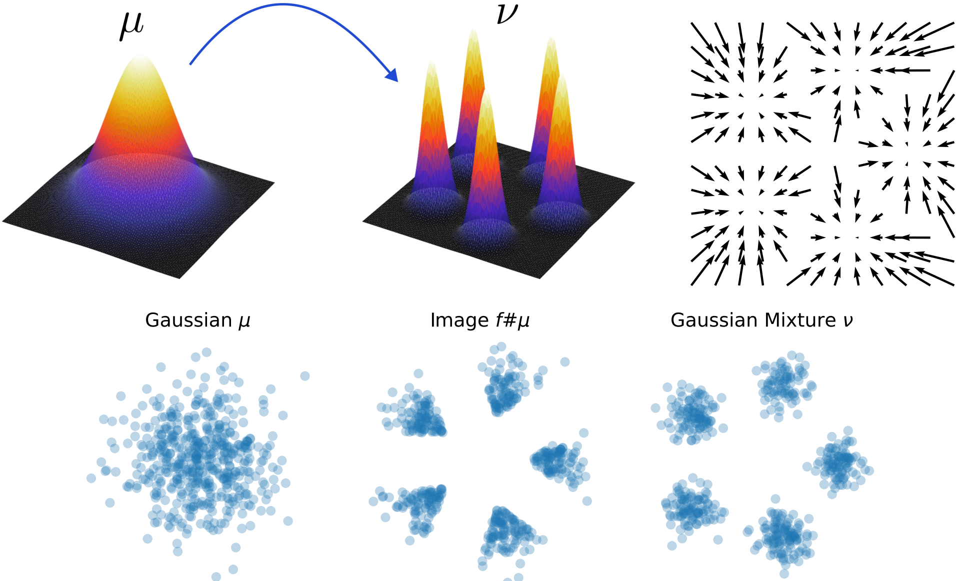


$$f\#\mu := \operatorname*{Law}_{X\sim\mu}[f(X)].$$



Push-Forward Measures

$$f\#\mu := \underset{X\sim\mu}{\mathrm{Law}}[f(X)].$$



The Monge Problem

$$\inf_{T\colon T\#\mu=\nu} \int c(x, T(x))\mathrm{d}\mu(x).$$

The Monge Problem

$$\inf_{T:\, T\#\mu=\nu} \int c(x, T(x))\mathrm{d}\mu(x).$$

Brenier's Theorem

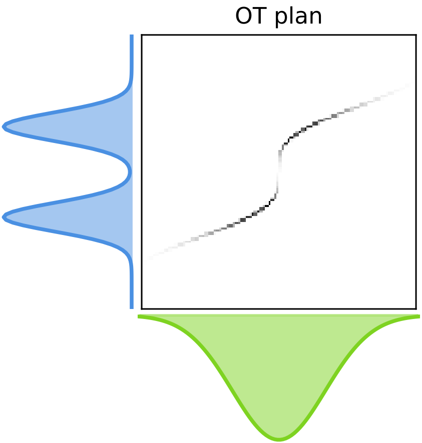
If $c(x,y) = \|x-y\|_2^2$, and $\mu \ll \mathscr{L}^d$, then there is a unique solution, and $T = \nabla\varphi$ with φ convex.

The Monge Problem

$$\inf_{T\colon T\#\mu=\nu} \int c(x, T(x))\mathrm{d}\mu(x).$$

Brenier's Theorem

If $c(x,y) = \|x-y\|_2^2$, and $\mu \ll \mathscr{L}^d$, then there is a unique solution, and $T = \nabla\varphi$ with φ convex.

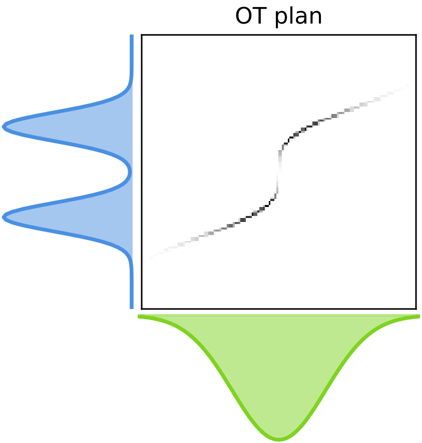


The Monge Problem

$$\inf_{T:\, T\#\mu=\nu} \int c(x, T(x))\mathrm{d}\mu(x).$$

Brenier's Theorem

If $c(x,y) = \|x-y\|_2^2$, and $\mu \ll \mathscr{L}^d$, then there is a unique solution, and $T = \nabla\varphi$ with φ convex.

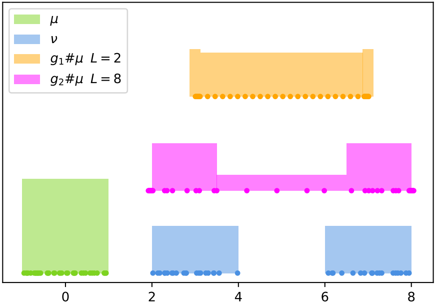


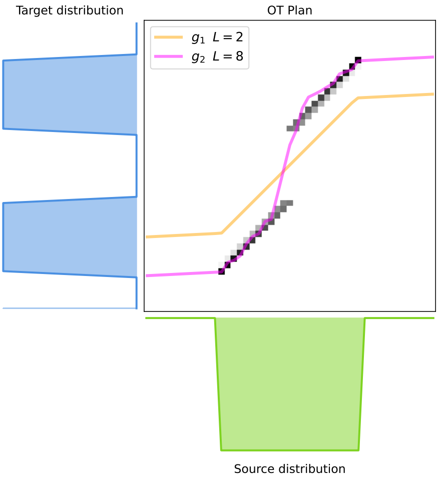
Kantorovich relaxation: $\mathcal{T}_c(\mu,\nu) := \min\limits_{\pi\in\Pi(\mu,\nu)} \int c(x,y)\mathrm{d}\pi(x,y).$

1. Transporting Measures
2. Constrained Approximate Transport Maps
3. Zoom on the L^2 Case on on $\mathbb{R}^d$

Problem statement

$$\mathcal{P}: \underset{g \in G}{\operatorname{argmin}}\ \mathcal{T}_c(g\#\mu, \nu)$$





Smooth Strongly Convex Nearest Brenier Potentials (Paty 2020 [2])

Case $G = G_{\ell,L} := \{\nabla\varphi : \ell I \preceq D^2\varphi \preceq LI\}$ and $c(x,y) = \|x-y\|_2^2$.

Smooth Strongly Convex Nearest Brenier Potentials (Paty 2020 [2])

Case $G = G_{\ell,L} := \{\nabla\varphi : \ell I \preceq D^2\varphi \preceq LI\}$ and $c(x,y) = \|x-y\|_2^2$.

$$\mu := \sum_{i=1}^n a_i \delta_{x_i},\ \nu := \sum_{j=1}^m b_j \delta_{y_j}.$$

Smooth Strongly Convex Nearest Brenier Potentials (Paty 2020 [2])

Case $G = G_{\ell,L} := \{\nabla\varphi : \ell I \preceq D^2\varphi \preceq LI\}$ and $c(x,y) = \|x-y\|_2^2$.

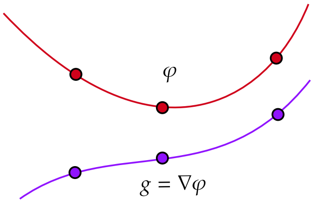
$$\mu := \sum_{i=1}^{n} a_i \delta_{x_i},\ \nu := \sum_{j=1}^{m} b_j \delta_{y_j}.$$

Interpolation (Taylor 2017 [3])

$$\exists g = \nabla\varphi \in G_{\ell,L} :$$

$$\forall i,\ g(x_i) = g_i,\ \varphi(x_i) = \varphi_i$$

iif $\forall i, j,\ Q_{\ell,L}(x_i, x_j, \varphi_i, \varphi_j, g_i, g_j) \geq 0.$



Smooth Strongly Convex Nearest Brenier Potentials (Paty 2020 [2])

Case $G = G_{\ell,L} := \{\nabla\varphi : \ell I \preceq D^2\varphi \preceq LI\}$ and $c(x,y) = \|x-y\|_2^2$.

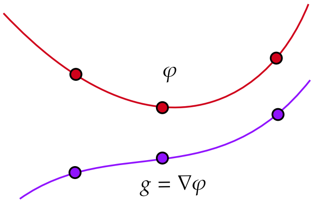
$$\mu := \sum_{i=1}^{n} a_i \delta_{x_i},\ \nu := \sum_{j=1}^{m} b_j \delta_{y_j}.$$

Interpolation (Taylor 2017 [3])

$$\exists g = \nabla\varphi \in G_{\ell,L} :$$

$$\forall i,\ g(x_i) = g_i,\ \varphi(x_i) = \varphi_i$$

iif $\forall i,j,\ Q_{\ell,L}(x_i, x_j, \varphi_i, \varphi_j, g_i, g_j) \geq 0.$



$$\underset{g\in G_{\ell,L}}{\operatorname{argmin}}\ \mathrm{W}_2^2(g\#\mu,\nu) \longleftrightarrow \underset{\substack{\pi\in\mathbb{R}^{n\times m},\ \varphi\in\mathbb{R}^n,\ g\in\mathbb{R}^{n\times d} \\ \pi\geq 0,\ \pi\mathbf{1}=a,\ \pi^T\mathbf{1}=b \\ Q_{\ell,L}(x_i,x_j,\varphi_i,\varphi_j,g_i,g_j)\geq 0}}{\operatorname{argmin}} \sum_{i,j} \|g_i - y_j\|_2^2 \pi_{i,j}.$$

Sufficient Conditions for Existence

$$\mathcal{P} : \underset{g \in G}{\operatorname{argmin}} \ \mathcal{T}_c(g\#\mu, \nu)$$

Existence if:

- Finite problem value,
- $c : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}_+$ lower semi-continuous,
- $c(y_1, y_2) \geq \alpha + \eta(\|y_1 - y_2\|)$ with η non-decreasing and coercive.
- G is a subclass of L-Lipschitz functions stable by local uniform limit.

Example classes: Neural Networks, $G_{\ell,L}$.

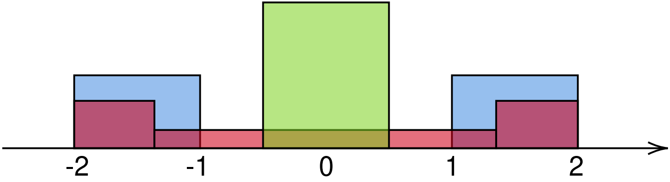
Sufficient Conditions for Existence

$$\mathcal{P} : \underset{g \in G}{\text{argmin}}\ \mathcal{T}_c(g\#\mu, \nu)$$

Existence if:

- Finite problem value,
- $c : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}_+$ lower semi-continuous,
- $c(y_1, y_2) \geq \alpha + \eta(\|y_1 - y_2\|)$ with η non-decreasing and coercive.
- G is a subclass of L-Lipschitz functions stable by local uniform limit.

Example classes: Neural Networks, $G_{\ell,L}$. Counter-example:



SGD for Neural Networks

Objective: $\min_\theta \mathcal{T}_c(g_\theta \# \mu, \nu)$.

Minibatch version: $\min_\theta \int \mathcal{T}_c(\delta_{g_\theta(X^{(n)})}, \delta_{Y^{(m)}}) \mathrm{d}\mu^{\otimes n}(X^{(n)}) \mathrm{d}\nu^{\otimes m}(Y^{(m)})$.

$$\theta_{t+1} = \theta_t - \alpha_t \left[\frac{\partial}{\partial \theta} \mathcal{T}_c(\delta_{g_\theta(X^{(n)})}, \delta_{Y^{(m)}}) \right]_{\theta = \theta_t}, X^{(n)} \sim \mu^{\otimes n},\ Y^{(n)} \sim \nu^{\otimes m}.$$

SGD for Neural Networks

Objective: $\min_\theta \mathcal{T}_c(g_\theta \# \mu, \nu)$.

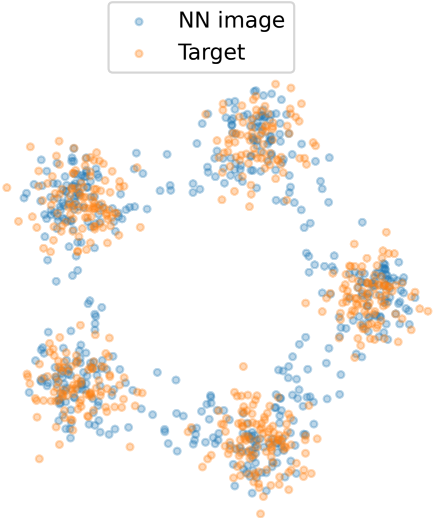
Minibatch version: $\min_\theta \int \mathcal{T}_c(\delta_{g_\theta(X^{(n)})}, \delta_{Y^{(m)}}) \mathrm{d}\mu^{\otimes n}(X^{(n)}) \mathrm{d}\nu^{\otimes m}(Y^{(m)})$.

$$\theta_{t+1} = \theta_t - \alpha_t \left[\frac{\partial}{\partial \theta} \mathcal{T}_c(\delta_{g_\theta(X^{(n)})}, \delta_{Y^{(m)}}) \right]_{\theta=\theta_t}, X^{(n)} \sim \mu^{\otimes n},\ Y^{(n)} \sim \nu^{\otimes m}.$$

SGD Convergence, using Bolte-Le-Pauwels [1]

For c and g semi-algebraic and μ, ν discrete or AC with semi-algebraic density, almost-surely accumulation points of (θ_t) are Clarke critical points of $\theta \longmapsto \mathcal{T}_c(g_\theta \# \mu, \nu)$.

Illustration: Neural Network Vector Fields



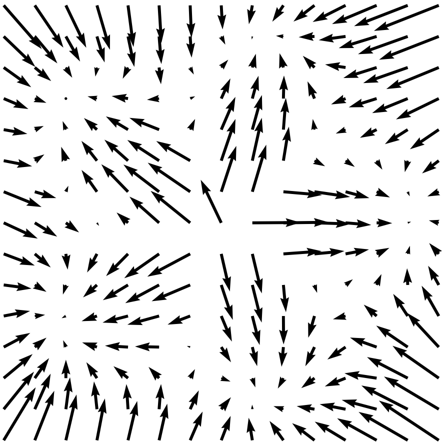
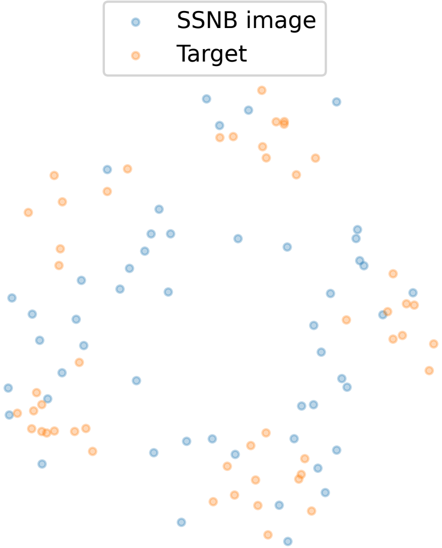
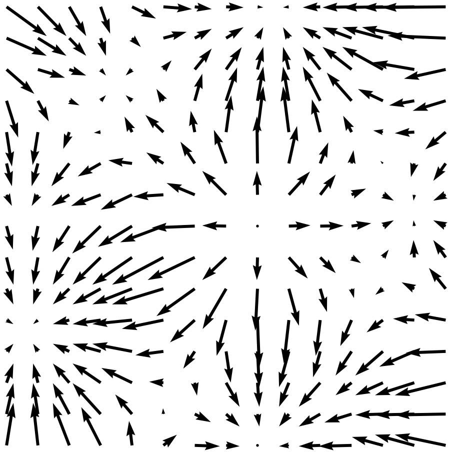


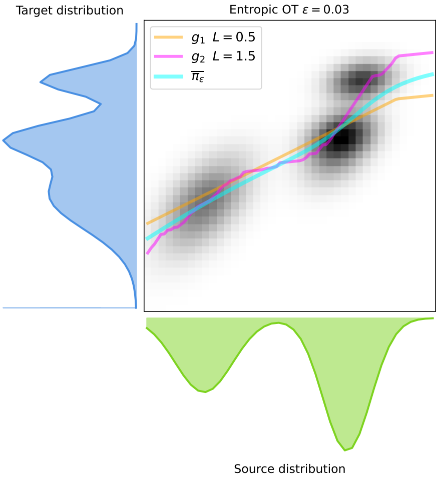
Illustration: Gradients of Strongly Convex Functions

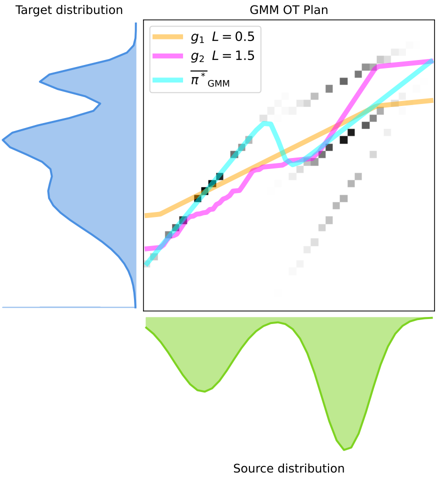




Plan Variant 1/2

$$\mathcal{P}_{\text{plan}} : \underset{g \in G}{\text{argmin}} \; \mathcal{T}_C((I, g)\#\mu, \gamma)$$





Plan Variant 2/2

$$\mathcal{P}_{\text{plan}} : \underset{g \in G}{\text{argmin}}\ \mathcal{T}_C((I, g)\#\mu, \gamma)$$

Problem Equivalence

$$\mathcal{T}_C((I, g)\#\mu, \gamma) = \mathcal{T}_{c_2}(g\#\mu, \nu)$$

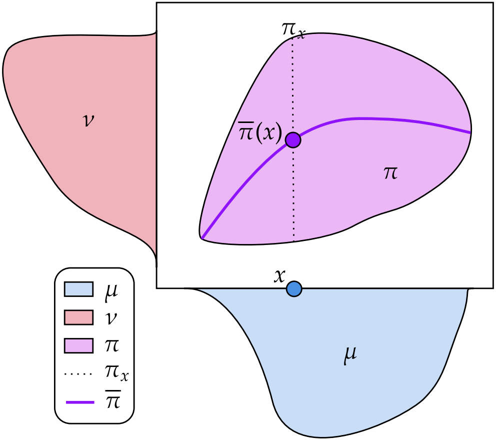
for $C((x_1, x_2), (y_1, y_2)) = h(c_1(x_1, y_1), c_2(x_2, y_2))$ if:

- $c_1(x, x) = 0$,
- $h(u, v) \geq v$,
- $h(0, v) = v$.

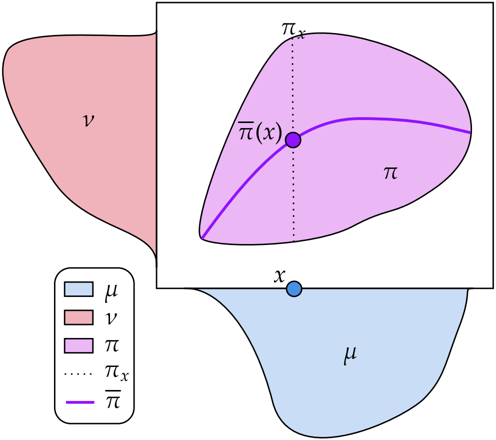
Ex: $C(\cdot, \cdot) = \| \cdot - \cdot \|_p^{qp},\ p \in [1, +\infty],\ q \geq 1.$

1. Transporting Measures
2. Constrained Approximate Transport Maps
3. Zoom on the L^2 Case on on $\mathbb{R}^d$

Barycentric Projections



Barycentric Projections



$$\overline{\pi}(x) = \int y \mathrm{d}\pi_x(y).$$

$$\overline{\pi}(x) = \mathbb{E}_{(X,Y)\sim\pi}[Y|X = x].$$

$$\overline{\pi} = \underset{f \in L^2(\mu)}{\operatorname{argmin}} \int \|f(x) - y\|_2^2 \mathrm{d}\pi(x, y).$$

Alternate Formulation for the L^2 cost on $\mathbb{R}^d$

$$\mathcal{P}: \min_{g\in G} \min_{\pi\in\Pi(\mu,\nu)} \int \|g(x) - y\|_2^2 \mathrm{d}\pi(x,y)$$

With π fixed:

$$\int \|g(x) - y\|_2^2 \mathrm{d}\pi(x,y) = \int \|g(x) - \overline{\pi}(x)\|_2^2 \mathrm{d}\mu(x) + K(\overline{\pi}).$$

Alternate Formulation for the L^2 cost on $\mathbb{R}^d$

$$\mathcal{P} : \min_{g \in G} \min_{\pi \in \Pi(\mu,\nu)} \int \|g(x) - y\|_2^2 \mathrm{d}\pi(x, y)$$

With π fixed:

$$\int \|g(x) - y\|_2^2 \mathrm{d}\pi(x, y) = \int \|g(x) - \overline{\pi}(x)\|_2^2 \mathrm{d}\mu(x) + K(\overline{\pi}).$$

Question: for $\pi \in \Pi^*(\mu, \nu)$, do we have

$$\operatorname*{argmin}_{g \in G} \int \|g(x) - \overline{\pi^*}(x)\|_2^2 \mathrm{d}\mu(x) \overset{?}{=} \operatorname*{argmin}_{g \in G} \min_{\pi \in \Pi(\mu,\nu)} \int \|g(x) - y\|_2^2 \mathrm{d}\pi(x, y).$$

Positive Answer in 1D

$$\underset{g\in G}{\operatorname{argmin}} \int \|g(x)-\overline{\pi^*}(x)\|_2^2 \mathrm{d}\mu(x) \overset{?}{=} \underset{g\in G}{\operatorname{argmin}} \min_{\pi\in\Pi(\mu,\nu)} \int \|g(x)-y\|_2^2 \mathrm{d}\pi(x,y).$$

Equivalence to L^2 projection in 1D for the L^2 cost

If all $g : \mathbb{R} \longrightarrow \mathbb{R}$ are non-decreasing and $\pi^* \in \Pi^*(\mu,\nu)$, then

$$\mathcal{P} : \underset{g\in G}{\operatorname{argmin}} \mathrm{W}_2^2(g\#\mu,\nu) = \underset{g\in G}{\operatorname{argmin}} \|g-\overline{\pi^*}\|_{L^2(\mu)}^2.$$

Counter-examples exist in higher dimensions. Generalises Paty 2020 [2].

Thanks

[1] Jérôme Bolte, Tam Le, and Edouard Pauwels.
Subgradient sampling for nonsmooth nonconvex minimization.
SIAM Journal on Optimization, 33(4):2542–2569, 2023.

[2] François-Pierre Paty, Alexandre d'Aspremont, and Marco Cuturi.
Regularity as regularization: Smooth and strongly convex brenier potentials in optimal transport.
In *International Conference on Artificial Intelligence and Statistics*, pages 1222–1232. PMLR, 2020.

[3] Adrien B Taylor.
Convex interpolation and performance estimation of first-order methods for convex optimization.
PhD thesis, Catholic University of Louvain, Louvain-la-Neuve, Belgium, 2017.