# ALGORITHMES HYBRIDES POUR LE TRANSPORT OPTIMAL

MAXIME SYLVESTRE

Abstract. Depuis l'algorithme de descente de gradient utilisé en optimisation convexe de nombreux autres types d'algorithmes sont apparus. On citera la descente de gradient proximale ou encore la descente par bloc. Récemment, pour des applications à l'économie notamment, les étapes d'optimisation utilisées étaient de types différents. C'est alors que sont nés les algorithmes hybrides, mêlant descente par bloc et gradient proximal. La preuve de leur convergence est plus complexe mais des cas particuliers ont été explorés en transport optimal notamment.

## 1. Transport optimal

Le problème de transport optimal est formulé par Gaspard Monge en 1781 ([28]). C'est un problème d'optimisation combinatoire qui consiste à aparier de manière optimale deux ensembles contenant le même nombre deux points dans le but de minimiser une fonction de coût. Ce problème peut s'écrire comme un problème d'optimisation matriciel. Étant donnée une matrice $C$ de coût le problème est le suivant

$$\min_{P\in\mathfrak{S}_n} \sum_i \mathrm{Tr} P^\top C$$

Leonid Kantorovich reformule dans les années 40 ce problème. La simplification vient de la reformulation qui permet d'optimiser sur un espace plus grand mais qui est continu. $P$ n'est plus une matrice de permutation mais une matrice bistochastique, ses lignes et ses colonnes somment à 1 et ses coefficients sont positifs. Le problème devient alors la minimisation d'une fonction convexe sur un espace convexe compact, ce qui assure l'exsitence d'une solution. George Birkhoff et John von Neumann ont développé des ([29]) théorèmes permettant de démontrer l'équivalence de ces deux problèmes. La nouvelle formulation permet d'introduire des contraintes aux marges plus variées. Il ne s'agit plus d'apparier des points entre eux mais des poids discrets. On a $n$ points $x_i$ de poids respectifs $a_i$ et $m$ points $y_j$ de poids respectifs $b_j$. On introduit l'espace des matrices d'appariement

$$\mathcal{B}(a,b) = \left\{ P \in M_{n,m}(\mathbb{R}_+) \mid \sum_j P_{i,j} = a_i, \sum_i P_{i,j} = b_j \right\}$$

Le problème devient alors

$$\min_{P\in\mathcal{B}(a,b)} \mathrm{Tr} P^\top D$$

*Date*: March 14, 2022.

Avec $D$ la matrice des distances entre les points. Ce problème introduit un nouveau degré de liberté mais il propose surtout une nouvelle vision du problème initial. En effet $a, b$ sont des mesures de probabilité discrètes sur des espaces. Ainsi il est possible de généraliser une dernière fois ce problème en introduisant des mesures de probabilités quelconques. Soient $\mu \in \mathcal{P}(X)$ et $\nu \in \mathcal{P}(Y)$ deux mesures de probabilité (avec $X, Y$ deux espaces probabilisables). Soit $d : X \times Y \to \mathbb{R}_+$. Le problème est le suivant

$$\inf_{\pi \in \Pi(\mu,\nu)} \int d(x,y) d\pi \tag{1.1}$$

Où $\Pi(\mu, \nu)$ est l'ensemble des mesures de probabilité sur $X \times Y$ telles que les distribution marginales sont respectivement $\mu$ et $\nu$.

**Théorème 1.** *Si $X, Y$ sont des espaces polonais et $d$ est semi-continue inférieurement. Alors 1.1 admet une solution*

La preuve ([35]) repose essentiellement sur la compacité de l'ensemble des plan de transport. Bien que le problème admette des solutions il n'y a pas de garantie sur leur régularité. Ce qui complique leur approximation. C'est pour cela que le problème de transport optimal régularisé par l'entropie est introduit.

$$\inf_{\pi \in \Pi(\mu,\nu)} \int d(x,y) d\pi + \epsilon \int \pi \log \pi d\mu \otimes \nu$$

Dans le cas où les marginales ont une entropie finie ce problème admet une unique solution ([11]). On y retrouve le problème discret dans le cas où les ensembles $X, Y$ le sont et c'est d'ailleurs pour ce type de problème discret qu'ont émergé les premières applications en informatique. En informatique on citera l'algorithme de Sinkhorn permettant de calculer le plan de transport optimal dont la convergence dans le cas discret fut prouvée en ([13]) et pour le cas continu ([32]). Ce problème trouve des applications en informatique avec son utilisation en machine learning et plus particulièrement dans les modèles d'attention ([26]). On note aussi des applications à la résolution du pont de Schrödinger ([16]). Il y a enfin des applications multiples en économie notamment pour l'étude des appariemments ([18]).

Pour rentrer plus en détail dans les fondements techniques de ces problèmes le livre de Villani fait office de référence ([35]).

## 2. Optimisation convexe

Dans le cadre du problème de transport optimal l'objectif est semblable à un plus grand ensemble de problèmes; celui de la minimisation d'une fonctionnelle sur un espace. Le problème général est le suivant: étant donnée une fonction $f$ sur un espace $\mathcal{H}$ à valeur dans $\mathbb{R}$ construire une suite $(x_n)_n$ convergeante dont la limite est un minimiseur de $f$. Dans le cas où $f$ est convexe la question équivalente est celle de la construction d'une suite convergeant vers un zéro de $\partial f$ l'opérateur sous-différentiel de $f$. En toute généralité il n'est pas possible de minimiser de manière déterministe une fonction quelconque. En effet les algorithmes déterministes permettent seulement d'obtenir des minima locaux. Cependant si l'hypothèse de convexité est faite alors les minima locaux sont finalement globaux. On

remarque que les fonctionnelles à minimiser dans le cadre du transport optimal sont bien convexes.

Dans son livre sur l'analyse convexe ([30]) Rockafellar introduit tous les outils de base pour étudier les problèmes d'optimisation convexe.

## 3. Algorithmes d'optimisation

Soit $H$ un espace de Hilbert et $f$ une fonction de $H$ dans $\mathbb{R}$. L'objectif est alors de trouver un opérateur $T$ tel que la suite $x_{n+1} = Tx_n$ converge vers un point fixe de $T$. Avec la condition supplémentaire que l'ensemble des points fixes de $T$ est l'ensemble des minimiseurs de la fonction $f$. Les différents types d'algorithmes reposent donc sur des opérateurs différents dont les propriétés spécifiques (lipschitz, fermement non expansif, ...) assurent la convergence du procédé. Deux grandes classes de propriétés sont à distinguer.

Les propriétés de régularité comme la convexité, le caractère lipschitz du gradient ou la forte convexité. Ces propriétés reposent en partie sur le caractère différentiable de la fonction à minimiser et permet de localiser le minimum en assurant que la fonction ne varie pas trop fortement.

Les propriétés géométriques comme la sous-modularité ([3], [27]) ou plus récemment l'échangeabilité ([19]). Ces propriétés ne nécessitent pas que la fonction soit différentiable et traduise l'interdépendance des coordonnées. Ces propriétés sont notamment utile dans le cadre des algorithmes de descentes de coordonnée ou hybrides.

### 3.1. Propriétés fondamentales.

Les propriétés suivantes sont souvent faites pour démontrer la convergence des algorithmes.

**Définition 1.** *$f$ est dite convexe si*

$$\forall x, y \in H, \forall t \in [0,1], f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

**Définition 2.** *$f$ est dite semi-continue inférieurement (sci) si*

$$\forall x, x_n \in H, x_n \to x \implies \liminf_n f(x_n) \geq f(x)$$

**Définition 3.** *$f$ est dite fermée si sont graphe est fermé.*

**Définition 4.** *$f$ est dite propre si elle ne prend pas la valeur $-\infty$.*

Sous ces conditions il est possible d'introduire une généralisation de la notion de gradient. Celle de sous-gradient et de sous-différentiel en un point $\partial f(x)$. Il est défini comme suit

$$\partial f(x) = \{s \in H \mid \forall y \in H, \langle s, y - x\rangle + f(x) \leq f(y)\}$$

Le premier fait essentiel peut alors être énoncé.

**Proposition 1.** *L'opérateur sous-différentiel $\partial f$ est maximal monotone. De plus les minimiseurs de $f$ sont les points tels que $0 \in \partial f$*

On peut alors introduire les propriétés de régularités énoncées précédemment.

**Définition 5.** *$f$ convexe est dite $\gamma$-fortement convexe si*

$$\forall x, y \in H, \forall s \in \partial f(x)$$
$$\frac{\gamma}{2}\|x-y\|^2 + \langle s, y-x\rangle + f(x) \leq f(y)$$

**Définition 6.** *$f$ a un sous-gradient $L$-Lipschitz dès lors que*

$$\forall x, y \in H, \forall \alpha \in \partial f(x), \forall \beta \in \partial f(y)$$
$$\|\alpha - \beta\|^2 \leq L\langle x-y, \alpha-\beta\rangle$$

3.2. **Approche différentielle.** Dans le cas où la fonction est différentiable, la formule de Taylor donne l'approximation suivante:

$$f(y) = f(x) + \langle \nabla f(x), y-x\rangle + o(y)$$

Ainsi pour $y = x - \alpha\nabla f(x)$ et $\alpha$ petit on a bien $f(y) < f(x)$ dès lors que $\nabla f(x) \neq 0$. Le théorème de convergence est le suivant.

**Théorème 2.** *(9.3 [6]) La suite $(x_n)$ définit comme suit*

$$x_{n+1} = x_n - \alpha\nabla f(x_n) = Tx_n$$

*converge vers un minimiseur de $f$ si les trois affirmations suivantes sont satisfaites:*

- *$f$ est $\gamma$-fortement convexe*
- *$\nabla f$ est $L$-lipschitz*
- $0 < \alpha < \frac{2}{L}$

La preuve repose sur une utilisation fine de la formule de Taylor au deuxième ordre qui garantit la décroissance stricte des évaluation. On observe alors que l'opérateur est contractant. Ce qui assure la convergence des itérés par théorème de point fixe de Banach. Bien que contraignantes, les hypothèses faites dans ce théorème sont classiques.

Le principe de cet algorithme remonte au moins à Augustin Louis Cauchy en 1847. Depuis, de nombreuses variantes ont été développées. Parmi les plus récentes on notera l'algorithme accéléré ([33]) ou encore l'algorithme BFGS ([7]).

3.3. **Approche opérateur.** On pose $R = (I + \lambda\partial f)^{-1}$ la résolvante de paramètre $\lambda$ de l'opérateur maximal monotone $\partial f$. Observons dans un premier temps que les points fixes de $R$ sont les zéros de $\partial f$. D'après ce qui précède les points fixes de $R$ sont bien les minimiseurs de $f$. Dans ce cas $R$ est appelé opérateur proximal et noté $\text{prox}_{\lambda f}$. Nous allons voir que cet opérateur satisfait les conditions pour générer un algorithme d'optimisation.

**Définition 7.** *Un opérateur $T$ est dit fermement non expansif si*

$$\|Tx - Ty\|^2 \leq \langle x-y, Tx-Ty\rangle$$

L'opérateur proximal est fermement non-expansif. Ainsi par le théorème de Krasnoselskii-Mann la suite d'itération converge.

**Théorème 3.** *[4] La suite des itérés d'un opérateur fermement non-expansif converge vers un point fixe de l'opérateur.*

La preuve repose sur le fait que pour $y$ point fixe de $T$ la suite $(\|x_n - y\|)_n$ converge. Ainsi pour la topologie faible la suite converge vers un point fixe de $T$ du fait de son caractère fermement non expansif.

L'intérêt de cette méthode est qu'elle est plus stable que la précédente, il y a moins que contrainte sur le paramètre de l'opérateur proximal. Cela tient au fait que l'opérateur proximal est fermement non-expansif de manière générale. Enfin c'est un algorithme d'optimisation qui requiert moins d'hypothèses sur le caractère différentiable de la fonction. Cependant le calcul de l'opérateur n'est pas évident, il est équivalent à la résolution du problème de minimisation suivant.

$$\mathrm{prox}_f(x) = \arg\min_y f(y) + \frac{1}{2}\|x - y\|^2$$

Ce calcul est simple pour des fonctions classiques comme la norme 1. On retrouve alors l'opérateur softmax utilisé en machine learning pour la régularisation. Dans le cas contraire le problème d'optimisation est aussi complexe que celui pour lequel l'opérateur est utilisé.

3.4. **Approche de descente par bloc.** Dans certains cas il est possible de minimiser de manière exacte la fonction sur une partie de ses coordonnées. Ainsi une méthode consiste à itérer ces minimisations successives coordonnées par coordonnées. C'est la première approche qui ne concerne pas l'ensemble des coordonnées et qui nécessite de s'assurer que la minimisation selon une coordonnée n'est pas défavorable à l'objectif final. Un problème qui sera d'autant plus complexe dans le cas hybride. L'étape centrale de l'algorithme est la suivante.

$$x_i^{k+1} \in \arg\min_x f(x, x_{-i}^k) \tag{3.1}$$

Deux approches ont été exploitées jusqu'alors pour démontrer la convergence de cet algorithme.

La première reprend l'esprit de la méthode avec opérateur. Elle requiert cependant les hypothèses classiques de $\gamma$-forte convexité et de lipschtzianité de l'opérateur sous-différentiel.

**Théorème 4.** *[24] L'algorithme converge dès lors que $f$ est $\gamma$-fortement convexe et de sous-différentiel $L$-Lipschitz. Avec $\frac{L}{\gamma} < 2$*

La preuve de ce théorème repose sur le fait que l'application qui à $x_{-i}$ associe le $x_i$ minimisant la fonction toutes choses égales par ailleurs est contractante sous ces hypothèses.

La seconde approche repose sur une propriété géométrique de la fonction. La sous-modularité.

**Définition 8.** *$f$ est dite sous-modulaire si*

$$f(x \vee y) + f(x \wedge y) \leq f(x) + f(y)$$

Cette propriété est équivalente dans le cas différentiable à la négativité des dérivées croisées : $\frac{\partial f}{\partial x \partial y} \leq 0$. Ces deux articles [36], [34] assurent la convergence de l'algorithme de manière locale lorsque la fonction est deux fois continuement dérivable. La preuve repose

sur le fait que localement la hessienne a un rayon spectral plus petit que 1 ce qui assure que le processus de Jacobi associé est convergeant.

Ces méthodes sont réputées par leur vitesse de convergence en pratique. En particulier le célèbre algorithme de Sinkhorn ([13]) utilisé en transport optimal est un algorithme de descente par bloc.

Les grandes démonstrations et outils de base pour prouver les convergences d'algorithmes sont présenté dans le livre de Bauschke sur les opérateurs monotones dans les espaces Hilbertiens ([4]).

## 4. Optimisation sous contrainte et problème dual

L'algorithme de Sinkhorn cité précédemment permet d'obtenir un plan de transport optimal. Cependant il porte sur l'optimisation de variables duales et non du plan de transport lui même. D'une manière générale tout problème d'optimisation convexe sous contrainte admet une formulation dite duale. Soit $f$ une fonction convexe sur un hilbert $H$. Soit $C$ un sous ensemble convexe de $H$, le problème primal est le suivant

$$\min_C f$$

Il existe une fonction $\mathcal{L} : H \times H$ telle que $\min_C f = \min_H \max_H \mathcal{L}$ et de plus sous certaines conditions

$$\min_H \max_H \mathcal{L} = \max_H \min_H \mathcal{L}$$

On dit alors que le saut de dualité est nul. C'est le cas lorsque $C$ est un compact. Le problème dit dual consiste à maximiser la fonction

$$v \to \min_{u \in H} \mathcal{L}(u, v)$$

### 4.1. Cas du transport optimal.

Rappelons dans le cas du transport optimal que l'optimisation est réalisée selon des contraintes sur les marginales. Et on a :

$$\inf_{\pi \in \Pi(\mu,\nu)} \int d(x,y)d\pi + \int \pi \log \pi d\mu \otimes \nu = \sup_{\phi,\psi} \int \psi d\mu + \int \phi d\nu - \int \exp(\phi + \psi - d)d\mu \otimes \nu$$

Il se trouve que dans le cas où les marginales sont d'entropie finie ([11]) il est possible d'obtenir la solution du problème primal à l'aide du problème dual. Ce qui justifie l'utilisation de l'algorithme de Sinkhorn. L'expression est la suivante

$$\exp(\phi^* + \psi^* - d) = \pi^*$$

## 5. Problèmes hybrides en transport optimal

Parmi les applications les plus récentes du transport optimal on compte le transport martingale ([14]), le transport avec contrainte linéaire ([37]), le transport causal ([1]) ou encore le calcul de métrique ([23]). Ces problèmes émergeants conservent la base d'optimisation du transport optimal mais y ajoutent une contrainte sur le plan de transport. Hors dans le cas continu ces contraintes peuvent-être en nombre infini non dénombrable, c'est ce qui

est observé en transport martingale notamment. L'enjeu est d'assurer l'existence d'une solution au problème puis de développer une méthode permettant de l'approcher.

5.1. **De nouveaux problèmes de minimisation.** Afin de gérer de nouvelles contraintes sur le plan de transport un terme linéaire est ajouté au problème de transport optimal. Introduisons la fonctionnelle suivante

$$F(\phi, \psi, \beta) = \int \exp(\phi + \psi - d^\beta) d\lambda + \int \hat{\pi}(d^\beta - \psi - \phi) d\lambda$$

À $\beta$ fixé la condition de minimisation assure que les $\phi, \psi$ optimaux sont solutions du problème de transport optimal dual associé à la fonction de coût $d^\beta$. Le plan de transport optimal est $\pi^\beta = \exp(\phi^* + \psi^* - d^\beta)$. Ainsi la minimisation selon $\beta$ permet, si la fonction $\beta \to d^\beta$ est bien choisie, d'obtenir des conditions sur le plan de transport optimal obtenu. En effet dans le cas où $d^\beta = d_0 + \beta d$ la condition nécessaire d'optimalité est la suivante:

$$\int d d\pi^\beta = \int d d\hat{\pi}$$

Dans le cas du problème de transport optimal discret cette approche porte ses fruits pour l'apprentissage de métrique ([10]). Cette dernière égalité est du type introduite dans le cadre du transport otpimal avec contrainte linéaire.

$$\inf \left\{ \int d_0 d\pi + \int \pi \log \pi d\lambda \mid \pi \in \Pi, \forall w \in W, \int w d\pi = 0 \right\} \tag{5.1}$$

Ce problème admet une solution dès lors qu'il existe un plan de transport satisfaisant les contraintes linéaires ([38]). La preuve utilise encore une fois le caractère compact de ce sous ensemble de plan de transport. La fonctionnelle introduite précédemment permet de gérer un nombre fini de contraintes linéaires. Ainsi une approximation consiste à remplacer $W$ par un espace linéaire de dimension finie.

Le transport optimal martingale quant à lui impose dans le cas continu un nombre infini de contraintes non linéaires.

$$\inf \left\{ \int d_0 d\pi + \int \pi \log \pi d\lambda \mid \pi \in \Pi, \mathbb{E}_\pi[Y|X] = X \right\}$$

Les tentatives de résolution reposent sur une discretisation du problème. Mais pour éviter une situation de fonctionnelle discontinue un autre type de contrainte est introduit ([22]). D'autres tentatives reposent sur l'introduction du coût dit de Marton([21]).

5.2. **De nouveaux algorithmes de minimisation.** La fonction présentée précédemment est intéressante pour la résolution de problèmes plus complexes. Cependant l'ajout de la nouvelle variable rend impossible l'utilisation d'un algorithme type Sinkhorn, car il n'y a pas de formule de minimisation exacte pour ce dernier terme. Ainsi il est question d'hybrider les méthodes d'optimisation. Cette approche rejoint l'idée de la descente de coordonnée par bloc. Le principe est simple, certaines méthodes sont plus efficaces sur certains types de problèmes. N'est-il donc pas plus efficace d'utiliser sur chaque ensemble de coordonnées les méthodes les plus adaptées?

5.3. **Descente de gradient proximale.** Une première méthode d'hybridation s'intéresse à la minimisation de la somme de deux fonctions ayant des régularités différentes. Soient $f, g$ deux fonctions convexes sur $H$ à valeur dans $\mathbb{R}$ telles que $f$ est continuement différentiable. Aucune hypothèse supplémentaire n'est faite sur $g$. Comment résoudre le problème

$$\min_H f(x) + g(x)$$

L'algorithme proposé est le suivant

$$x_{n+1} = \text{prox}_{\lambda g}(x_n - \lambda \nabla f(x_n))$$

Il peut aussi être perçu comme une première étape de descente de gradient puis une seconde proximale. Bien qu'individuellement ces deux méthodes convergent qu'en est-il de la composition des deux?

**Théorème 5.** *([12]) On suppose que le gradient de $f$ est L-lipschitz. Alors pour $\lambda \in ]0, \frac{1}{L}]$ l'algorithme ci-dessus converge vers une solution du problème.*

La preuve de convergence repose encore une fois sur le caractère fermement non expansif de l'opérateur proximal.

Notons que cette méthode englobe d'autres méthodes comme celle de la descente de gradient projeté. En effet dans le cas où $g$ est la fonction indicatrice d'un ensemble convexe alors l'opérateur proximal est simplement la projection sur cet ensemble.

5.4. **Descente par bloc proximale.** La seconde méthode d'hybridation présentée ici est utile dans le problème de transport optimal de la section précédente. En effet nous avons vu que $\phi, \psi$ peuvent être obtenu à l'aide de l'algorithme de Sinkhorn. Afin de profiter de cette vitesse de convergence l'algorithme proposé utilise la descente par bloc sur certaines coordonnées et une descente plus classique par gradient ou opérateur proximal sur les autres. Soit $F : H_1 \times H_2 \times H_3 \to \mathbb{R}$ convexe, continument différentiable. Le problème est

$$\min_{u \in H_1} \min_{v \in H_2} \min_{\beta \in H_3}$$

On propose l'algorithme de minimisation suivant

$$u_{n+1} \in \arg\min_u F(u, v_n, \beta_n)$$
$$v_{n+1} \in \arg\min_v F(u_{n+1}, v, \beta_n)$$
$$\beta_{n+1} = \beta_n - \lambda \nabla_\beta F(u_{n+1}, v_{n+1}, \beta_n)$$

La convergence de cet algorithme a été prouvée dans le cas de la fonctionnelle ([10]) lorsque le problème de transport est discret. Des pistes d'améliorations sont possibles pour accélérer la convergence de l'algorithme dont la vitesse est affectée par la dernière étape plus lente. [5] présente une descente proximale accélérée.

## 6. Applications à d'autres domaines

Bien que les différents problèmes soient présentés comme des problèmes d'optimisation il est aussi possible de les interpréter comme la résolution d'un système non linéaire si l'on prend le point de vue sous-différentiel. De plus les problèmes de transport optimal avec contrainte sont en partie nés dans d'autres domaines. Il y a donc de multiples applications de ces résultats mathématiques.

6.1. **Physique.** Le problème dual de transport optimal fourni des solutions au problème du pont de Schrödinger. Une des preuves de la convergence de l'algorithme de Sinkhorn dans le cas continu est justement dérivée de ce problème. La preuve est détaillée dans [16].

6.2. **Économie.** L'implémentation des méthodes itératives hybrides permet l'étude de problème de matching ([19],[20]). L'algorithme hybride dans [10] est utilisé pour l'apprentissage de métrique dans le cadre de l'étude des migrations. En finance le transport martingale est utilisé pour le "model-free superhedging" ([17]). Cependant les tentatives actuelles de calcul effectif sont imparfaites. Enfin en économétrie ces algorithmes hybrides trouvent une utilité pour la régression de vecteurs de quantile ([8]).

## 7. Grandes questions ouvertes

7.1. **Pistes explorées et à explorer.** La sous-modularité est utilisée dans certains travaux d'optimisation ([3]). Cependant les approches restent discrètes. Hors comme nous l'avons vu cette propriété géométrique assure la convergence de certains types de méthodes itératives. Dans ([19]) nous avons étendu des résultats classique sur des fonctions sous-modulaire discrètes à des fonctions sur $\mathbb{R}^n$. Nous avons pu en déduire des informations sur le comportement du sous-différentiel d'une fonction convexe sous-modulaire. L'expression de ce résultat est plus simple dans le cas différentiable

**Théorème 6.** *Soit $f$ une fonction convexe différentiable. $f$ est sous-modulaire si et seulement si*

$$\forall x \leq x', \forall i \in \{x = x'\}, (\nabla f(x))_i \geq (\nabla f(x'))_i$$

Ce théorème dans le cas différentiable est une application directe de la sous-modularité dans le calcul du gradient. Le point intéressant est l'introduction d'un ordre sur les sous-différentiels, afin de traiter le cas général, qui est cohérent avec le cas différentiable. Cette propriété permet de développer un algorithme d'optimisation pour les fonctions convexes sous-modulaires. Il est intéressant de noter que la fonctionnelle du problème dual du transport optimal est convexe sous-modulaire. Ainsi l'enjeu est d'exploiter cette propriété géométrique pour dériver des vitesses de convergence de l'algorithme hybride et d'une manière générale réfléchir autour d'un cadre général de manipulation de ces méthodes hybrides. L'objectif final étant de pouvoir les appliquer aux problèmes de transport optimal avec des contraintes supplémentaires.

7.2. **Grands axes de recherche.** De nombreuses questions restent ouvertes.

- L'algorithme de Sinkhorn converge effectivement dans le cas discret et la vitesse de convergence est du bon ordre. Cependant dans le cas continu la question reste ouverte. On notera l'analogie avec l'algorithme de descente de gradient qui permet de dériver une vitesse de convergence sous-linéaire, qui reste toutefois plus lente que dans le cas discret ([25])
- Bien que le problème régularisé a favorisé l'apparition d'un grand nombre d'applications la question de la convergence de la solution de ce problème lorsque le terme d'entropie est multiplié par un terme tendant vers 0 reste ouverte. ([11],[9])
- Chaque algorithme hybride demande une attention particulière. Toutefois certaines approches plus générales ont tenté de démontrer la convergence d'algorithme mettant en jeu des opérateurs différents.

- Pour chacun des problèmes d'optimisation il est possible de relâcher la condition de convexité. Les algorithmes déterministes ne permettent pas d'obtenir un minimum global ([31]). En revanche des algorithmes stochastiques convergent vers des minima globaux, mais à une vitesse lente. ([2])
- D'une manière générale beaucoup de méthodes n'ont pas encore de bonnes vitesses de convergence. Les constantes étant particulièrement grande. ([25],[15])

## References


[1] Beatrice Acciaio, Julio Backhoff Veraguas, and Anastasiia Zalashko. *Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization*. 2016. DOI: 10.48550/ARXIV.1611.02610. URL: https://arxiv.org/abs/1611.02610.

[2] Hilal Asi and John C. Duchi. "The importance of better models in stochastic optimization". In: *Proceedings of the National Academy of Sciences* 116.46 (2019), pp. 22924–22930. DOI: 10.1073/pnas.1908018116. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1908018116. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1908018116.

[3] Francis Bach. *Learning with Submodular Functions: A Convex Optimization Perspective*. 2013. arXiv: 1111.6453 [cs.LG].

[4] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer New York, 2011. DOI: 10.1007/978-1-4419-9467-7. URL: https://doi.org/10.1007/978-1-4419-9467-7.

[5] Amir Beck and Marc Teboulle. "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems". In: *SIAM Journal on Imaging Sciences* 2.1 (Jan. 2009), pp. 183–202. DOI: 10.1137/080716542. URL: https://doi.org/10.1137/080716542.

[6] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[7] C. G. BROYDEN. "The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations". In: *IMA Journal of Applied Mathematics* 6.1 (Mar. 1970), pp. 76–90. ISSN: 0272-4960. DOI: 10.1093/imamat/6.1.76. eprint: https://academic.oup.com/imamat/article-pdf/6/1/76/2233756/6-1-76.pdf. URL: https://doi.org/10.1093/imamat/6.1.76.

[8] Guillaume Carlier, Victor Chernozhukov, and Alfred Galichon. "Vector quantile regression: An optimal transport approach". In: *The Annals of Statistics* 44.3 (2016), pp. 1165 –1192. DOI: 10.1214/15-AOS1401. URL: https://doi.org/10.1214/15-AOS1401.

[9] Guillaume Carlier et al. "Convergence of Entropic Schemes for Optimal Transport and Gradient Flows". In: *SIAM Journal on Mathematical Analysis* 49.2 (Apr. 2017). DOI: 10.1137/15M1050264. URL: https://hal.archives-ouvertes.fr/hal-01246086.

[10] Guillaume Carlier et al. *SISTA: learning optimal transport costs under sparsity constraints*. 2020. arXiv: 2009.08564 [math.OC].

[11] Christian Clason et al. "Entropic regularization of continuous optimal transport problems". In: *Journal of Mathematical Analysis and Applications* 494.1 (2021), p. 124432. ISSN: 0022-247X. DOI: 10.1016/j.jmaa.2020.124432. URL: http://dx.doi.org/10.1016/j.jmaa.2020.124432.

[12] Patrick Combettes and Jean-Christophe Pesquet. "Proximal Splitting Methods in Signal Processing". In: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering* 49 (Dec. 2009). DOI: 10.1007/978-1-4419-9569-8_10.

[13] Marco Cuturi. "Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances". In: (2013). DOI: 10.48550/ARXIV.1306.0895. URL: https://arxiv.org/abs/1306.0895.

[14] Yan Dolinsky and H. Mete Soner. *Martingale Optimal Transport and Robust Hedging in Continuous Time*. 2013. arXiv: 1208.4922 [math.PR].

[15] Stephan Eckstein and Marcel Nutz. *Quantitative Stability of Regularized Optimal Transport and Convergence of Sinkhorn's Algorithm*. 2021. DOI: 10.48550/ARXIV.2110.06798. URL: https://arxiv.org/abs/2110.06798.

[16] Montacer Essid and Michele Pavon. "Traversing the SchröDinger Bridge Strait: Robert Fortet's Marvelous Proof Redux". In: *J. Optim. Theory Appl.* 181.1 (2019), 23–60. ISSN: 0022-3239. DOI: 10.1007/s10957-018-1436-9. URL: https://doi.org/10.1007/s10957-018-1436-9.

[17] A. Galichon, P. Henry-Labordère, and N. Touzi. "A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options". In: *The Annals of Applied Probability* 24.1 (2014), pp. 312 –336. DOI: 10.1214/13-AAP925. URL: https://doi.org/10.1214/13-AAP925.

[18] Alfred Galichon. *Optimal Transport Methods in Economics*. Economics Books 10870. Princeton University Press, 2016. URL: https://ideas.repec.org/b/pup/pbooks/10870.html.

[19] Alfred Galichon, Yu-Wei Hsieh, and Maxime Sylvestre. *Characterization of Gross substituability: a deferred acceptance algoritm*. 2022.

[20] Alfred Galichon and Bernard Salanie. "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models". In: *SSRN Electronic Journal* (2011). DOI: 10.2139/ssrn.1804623. URL: https://doi.org/10.2139/ssrn.1804623.

[21] Nathael Gozlan et al. "Kantorovich duality for general transport costs and applications". In: *Journal of Functional Analysis* 273.11 (2017), pp. 3327–3405. ISSN: 0022-1236. DOI: https://doi.org/10.1016/j.jfa.2017.08.015. URL: https://www.sciencedirect.com/science/article/pii/S0022123617303294.

[22] Gaoyue Guo and Jan Obłój. "Computational methods for martingale optimal transport problems". In: *The Annals of Applied Probability* 29.6 (2019), pp. 3311 –3347. DOI: 10.1214/19-AAP1481. URL: https://doi.org/10.1214/19-AAP1481.

[23] Matthieu Heitz et al. "Ground Metric Learning on Graphs". In: *Journal of Mathematical Imaging and Vision* 63.1 (2020), pp. 89–107. DOI: 10.1007/s10851-020-00996-z. URL: https://doi.org/10.1007%2Fs10851-020-00996-z.

[24] Hanbaek Lyu. *Convergence and complexity of block coordinate descent with diminishing radius for nonconvex optimization*. 2020. DOI: 10.48550/ARXIV.2012.03503. URL: https://arxiv.org/abs/2012.03503.

[25] Flavien Léger. *A gradient descent perspective on Sinkhorn*. 2020. DOI: 10.48550/ARXIV.2002.03758. URL: https://arxiv.org/abs/2002.03758.

[26] Grégoire Mialon et al. *A Trainable Optimal Transport Embedding for Feature Aggregation and its Relationship to Attention*. 2020. DOI: 10.48550/ARXIV.2006.12065. URL: https://arxiv.org/abs/2006.12065.

[27] Paul Milgrom and Chris Shannon. "Monotone Comparative Statics". In: *Econometrica* 62.1 (1994), pp. 157–180. URL: https://ideas.repec.org/a/ecm/emetrp/v62y1994i1p157-80.html.

[28] Gaspard Monge. *Mémoire sur la théorie des déblais et des remblais*. De l'Imprimerie Royale, 1781.

[29] John von Neumann. "1. A Certain Zero-sum Two-person Game Equivalent to the Optimal Assignment Problem". In: *Contributions to the Theory of Games (AM-28), Volume II*. Ed. by Harold William Kuhn and Albert William Tucker. Princeton University Press, 2016, pp. 5–12. DOI: doi:10.1515/9781400881970-002. URL: https://doi.org/10.1515/9781400881970-002.

[30] R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series. Princeton, N. J.: Princeton University Press, 1970.

[31] Alessandro Rudi, Ulysse Marteau-Ferey, and Francis Bach. *Finding Global Minima via Kernel Approximations*. 2020. DOI: 10.48550/ARXIV.2012.11978. URL: https://arxiv.org/abs/2012.11978.

[32] Ludger Ruschendorf. "Convergence of the iterative proportional fitting procedure". In: *The Annals of Statistics* (1995), pp. 1160–1174.

[33] Mark Schmidt, Nicolas Le Roux, and Francis Bach. *Convergence Rates of Inexact Proximal-Gradient Methods for Convex Optimization*. 2011. DOI: 10.48550/ARXIV.1109.2415. URL: https://arxiv.org/abs/1109.2415.

[34] R.S. Varga. *Matrix Iterative Analysis*. Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2009. ISBN: 9783642051548. URL: https://books.google.fr/books?id=ix__1MNMHfIC.

[35] C. Villani and American Mathematical Society. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003. ISBN: 9781470418045. URL: https://books.google.fr/books?id=MyPjjgEACAAJ.

[36] M.N. Vrahatis, G.D. Magoulas, and V.P. Plagianakos. "From linear to nonlinear iterative methods". In: *Applied Numerical Mathematics* 45.1 (2003). 5th IMACS Conference on Iterative Methods in Scientific Computing 28-31 May, 2001, Heraklion, Crete (Greece), pp. 59–77. ISSN: 0168-9274. DOI: https://doi.org/10.1016/S0168-9274(02)00235-0. URL: https://www.sciencedirect.com/science/article/pii/S0168927402002350.

[37] Danila Zaev. *On the Monge-Kantorovich problem with additional linear constraints*. 2014. DOI: 10.48550/ARXIV.1404.4962. URL: https://arxiv.org/abs/1404.4962.

[38] Danila Zaev. *On the Monge-Kantorovich problem with additional linear constraints*. 2014. arXiv: 1404.4962 [math.FA].