

Convex Order and Arbitrage

Erica Zhang

Supervisors: Professor Marcel Nutz,
Professor Johannes Wiesel

A thesis presented for the Honors degree of
B.A. in Mathematics-Statistics



Department of Mathematics
Department of Statistics
Columbia University
March 31st, 2023

Acknowledgements

Writing this thesis would not have been possible without the mentorship and support from my supervisors. I would like to thank Professor Marcel Nutz for kindly agreeing to supervise my thesis, despite having a heavily packed schedule. I sincerely appreciate his feedback and patience. I would like to also thank my supervisor Professor Johannes Wiesel for giving me an incredible amount of guidance and support throughout my time at Columbia. I would like to thank him for mentoring me with great patience and kindness, introducing me to the topic of optimal transport, and inspiring me to develop a keen interest in stochastic processes and mathematical finance. His immense knowledge and passion for his field of study have greatly encouraged me in many aspects of my academic research. My growth today would have been very difficult to achieve without him. For that, I am eternally grateful.

I would also like to express my gratitude for Professor Ioannis Karatzas, who kindly gave me guidance and enlightened me with key insights on functional portfolio theory during his free time, despite my not being enrolled in his class. I would like to thank him for the feedback he gave me on Section 6 of the thesis as well as key ideas that helped formulate Conjecture 6.3.1. I would like to thank him for sharing his passion in functional portfolio theory with me as well as introducing me to the brand-new field of stochastic portfolio theory.

Contents

1	Introduction and Main Results	3
1.1	Convex Order through the 2-Wasserstein Distance	4
1.2	Breeden-Litzenberger formula	5
1.3	Model-free version of arbitrage strategy	6
2	Setup and Notation	7
3	Optimizing ρ	8
3.1	Indirect Dirichlet Sampling	8
3.2	Direct Randomized Dirichlet Sampling	9
3.3	Optimization with <code>cvxpy</code>	10
3.4	Complexity and Performance	13
3.4.1	Analysis of Complexity	13
3.4.2	Experiments and Results	17
4	Recovering Convex Function \hat{f}	19
4.1	Recovering \hat{f} in 1D	19
4.1.1	Algorithm for 1D	20
4.1.2	Experiments and Results for 1D	20
4.2	Recovering \hat{f} in 2D	21
4.2.1	Algorithm for 2D	22
4.2.2	Experiments and Results for 2D	22
5	Model-free arbitrage strategy via \hat{f}	23
6	Arbitrage between Portfolios	25
6.1	Functionally generated portfolio	25
6.1.1	Market model	26
6.1.2	Trading strategies	26
6.1.3	Proportional Investment	27
6.1.4	Additively generated portfolios	27
6.2	Functionally generated arbitrage	28
6.2.1	Relative arbitrage	29
6.2.2	Relative arbitrage between additively generated portfolios	29
6.3	Convex order and the gamma process	31
7	Conclusion	33
8	Appendix	33
8.1	A. Derivation of the TPE Acquisition Function	33
8.2	B. Derivation of the Poisson Equation with Neumann Boundary Conditions	34

Convex Order and Arbitrage

Erica Zhang

March 2023

Abstract

Wiesel and Zhang [2023] [1] established that two probability measures μ, ν on \mathbb{R}^d with finite second moments are in convex order (i.e. $\mu \preceq_c \nu$) if and only if $W_2(\nu, \rho)^2 - W_2(\mu, \rho)^2 \leq \int |y|^2 \nu(dy) - \int |x|^2 \mu(dx)$. Let us call a measure ρ maximizing $W_2(\nu, \rho)^2 - W_2(\mu, \rho)^2$ the optimal ρ . This paper summarizes key findings by Wiesel and Zhang, develops new algorithms enhancing the search of optimal ρ , and builds on the paper through constructing a model-independent arbitrage strategy and developing associated numerical methods via the convex function recovered from the optimal ρ through Brenier's theorem. In addition to examining the link between convex order and arbitrage through the lens of optimal transport, the paper also gives a brief survey of functionally generated portfolio in stochastic portfolio theory and offers a conjecture of the link between convex order and arbitrage between two functionally generated portfolios.

1 Introduction and Main Results

Fix two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ with

$$\int |x| \mu(dx) < \infty, \quad \int |y| \nu(dy) < \infty.$$

Then μ and ν are defined to be in convex order (denoted by $\mu \preceq_c \nu$) iff

$$\int f d\mu \leq \int f d\nu \quad \text{for all convex functions } f : \mathbb{R}^d \rightarrow \mathbb{R}.$$

The above integrals take values in $(-\infty, \infty]$ as any convex function is bounded from below by an affine function. Convex order is a form of stochastic ordering, and it plays an essential role in mathematical finance. One of the most important theories in this domain is Strassen's theorem [1965] [2], which states that $\mu \preceq_c \nu$ if and only if $\mathcal{M}(\mu, \nu)$ - the set of martingale laws on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν - is non-empty. The well-established first fundamental theorem of asset pricing (FTAP) states that a market model has no arbitrage if and only if there exists a martingale deflator. If we can somehow make sense of μ and ν as certain financial instruments in the market, we can combine Strassen's theorem with the first FTAP and develop a test on arbitrage through

convex order. This is our original motivation for the paper. A natural link of probability measures μ and ν with assets in the market is through Breeden-Litzenberger formula [1979] [3], where μ and ν represent laws of stock prices at maturities. The details of the formula are discussed in Section 1.2.

There are a number of applications of convex order in the financial market as well as in financial modeling. To name a few, Guyon [2019] [4] studied the inversion of convex ordering in the VIX market and established that a continuous model on the S&P 500 index (SPX) jointly calibrating to a full surface of SPX implied volatilities and to the VIX smiles exists if and only if for each time t , the local variance is smaller than the instantaneous variance in convex order. Carr et al. [2018] [5] focused on convex order in risk-neutral densities for call options. In examining realized variance option and options on quadratic variation normalized to unit expectation, they found that if monotonicity in maturity for call options at a fixed strike is satisfied, the risk-neutral densities are increasing in the convex order. In addition, they also developed modeling strategies guaranteeing an increase in convex order for the normalized quadratic variation. Last but not least, the role of convex order in dependence modelling and risk aggregation is discussed extensively in paper such as Tchen [1980] [6] and Bernard et al. [2016] [7].

In this paper, we offer a novel perspective by connecting convex order with the existence of arbitrage opportunities in the market and by developing trading strategies that leverage convex order in financial market. Specifically, the paper considers arbitrage in the trading of financial assets in a market and relative arbitrage between two functionally generated portfolios, with a focus on the former. Numerical implementation in this paper can be found in our [Github repository](#). To make any of the convex order based arbitrage strategies practically realizable, it is critical to be able to identify convex order in general dimensions. To do this, we leverage the new optimal transport-based characterization of convex order method developed by Wiesel and Zhang, introduced in the section below.

1.1 Convex Order through the 2-Wasserstein Distance

In paper “[An optimal transport based characterization of convex order](#)”, we, Wiesel and Zhang [2023] [1] developed a new method of characterizing convex order in general dimensions using tools from optimal transport. Specifically, two probability measures μ, ν with finite second moments are in convex order (denoted $\mu \preceq_c \nu$) if and only if

$$W_2(\nu, \rho)^2 - W_2(\mu, \rho)^2 \leq \int |y|^2 \nu(dy) - \int |x|^2 \mu(dx) \quad (1)$$

This result is obtained through combining classical optimal transport duality with Brenier’s theorem, see Brenier [1991] [8]. Optimal transport goes back to the seminal works of Monge [1981] [9] and Kantorovich [1958] [10]. It is concerned with the problem of transporting probability distributions in a cost-optimal way. We refer to Rachev and Rüschendorf [1998] [11] and Villani [2003, 2008] [12] for an overview. In this paper, Wiesel

and Zhang attempt to develop a method to characterize convex order in general dimensions beyond the real line. To the best of our knowledge, this link between convex order and optimal transport can first be found in Carlier [2008] [13]. Contrary to Carlier's proof, however, we make no assumptions on the probability measures μ, ν except for that they have finite second moments. In particular, we prove the theorem in a general case where there is no need to assume that the measures are absolutely continuous wrt. the Lebesgue measure (see proof of Theorem 1.1 in Wiesel and Zhang [1]).

To develop model-independent strategy based on this optimal transport-based characterization, we introduce the Breeden-Litzenberger formula.

1.2 Breeden-Litzenberger formula

The Breeden-Litzenberger (B-L) formula is a fundamental tool in option pricing theory that has been widely used in finance and economics since its introduction in 1978, see Breeden and Litzenberger [1978] [3]. The formula builds on the concept of state-contingent financial claims, which are financial instruments whose payoff depends on the occurrence or non-occurrence of certain events. Options are a common example of state-contingent claims, as their payoffs depend on the price of the underlying asset at expiration. The B-L formula relates the prices of options on the same underlying asset with different strike prices and maturities to the prices of the underlying asset and other financial instruments, such as bonds. The set-up is the following.

Consider a market with a risk-free asset with interest rate $r = 0$ and d risky assets. Assume that the risky asset $S = (S^1, \dots, S^d)$ over a time horizon T is a Markov process. Now consider a European call option portfolio with options at the same maturity T and traded at all strikes. Then the B-L formula enables us to recover the joint probability distribution of the underlying risky assets by observing the call option prices in the market's European call option portfolios.

We give a brief overview of the derivation of the B-L results. Fix a maturity T . Consider a martingale measure μ and denote the law of S under μ by \mathbb{P} . Then the price of a general European contingent call option of maturity T and strike K , denoted $C(T, K)$, is given by the following:

$$\begin{aligned} C(T, K) &= \mathbb{E}_\mu[(S_T - K)^+] \\ &= \int_K^\infty (x - K)\mu(dx) \end{aligned}$$

Thus,

$$\begin{aligned} \frac{\partial}{\partial K} C(T, K) &= \int_K^\infty -\mu(dx) \\ &= -\mu([S_T \geq K]) \end{aligned}$$

and

$$\frac{\partial^2}{\partial K^2} C(T, K) = \mu(dK)$$

implying that observing all call prices yields the marginal density of S_T . The above results is exactly the B-L formula.

Indeed, one of the key insights of the B-L formula is that the risk-neutral probability of the underlying asset's price changing by a certain amount can be inferred from the prices of options on the same asset. This allows one to price options using the Black-Scholes formula or other option pricing models that assume a constant volatility, even if the actual volatility of the underlying asset is unknown or time-varying. However, some practical issues remain. For example, interpolating observed call prices with a smooth function is very difficult, and the interpolation scheme might not be arbitrage-free. Additionally, the formula relies on the assumption of no-arbitrage, which may not hold in certain market conditions or during periods of market stress. Several papers have extended or modified the B-L formula to address some of its limitations. For instance, Rubinstein [1994] [14] proposed a modification to the B-L formula that allows for non-log-normal underlying asset distributions. Dumas, Fleming, and Whaley [1998] [15] proposed a modified B-L formula for valuing interest rate options that accounts for term structure effects. Wiggins [1987] [16] proposed a method for estimating the implied risk-neutral probability density function using the B-L formula.

In our paper, we restrict ourselves to the classical setting of the B-L formula and leave the discussion of relaxed assumptions to another paper. We then introduce the final element for the construction of our model-free trading strategy.

1.3 Model-free version of arbitrage strategy

We consider in this paper the so-called *model-independent approach* to the study of arbitrage and trading strategy, i.e. the consideration of the set of all models contingent with the prices observed in the market. The search of robust methods to mathematical finance problems such as asset-pricing and super-replications has been gaining traction especially recently. Hobson [1998] [17] pioneered the study of *model-independent arbitrage* with respect to *semi-trading strategies* and robust hedging. Since then, numerous literature considering the necessary and sufficient conditions for the existence of calibrated arbitrage-free models under different settings given an observation of prices of finitely (or infinitely) European call options is developed, see Cousot [2007] [18], Buehler [2006] [19], and Carr and Madan [2005] [20]. The model-free version of the Fundamental Theorem of Asset Pricing is studied by Riedel [2011] [21] and Acciaio et al. [2018] assumes the prices of finitely many options and considers the robust version of the FTAP with respect to a weak notion of arbitrage under the assumption of a compact state-space in a one-period setting. Moreover, given a path-dependent derivative Φ , Cox and Obloj [2011] [22] and others made a series of important advancements in the development of a robust version of the *Super-Replication Theorem*, where they determined explicitly and demonstrated the equivalence in values of the values of upper martingale price $p^M(\Phi)$, i.e. the supremum of the expected payoff over admissible martingale measures and the minimal endowment $p^R(\Phi)$ required for super-replication. Following their work, Nutz [2013] [23] ex-

tends the setup and studies optimal super-replication strategies where super-replication is understood not in a point-wise sense, but rather w.r.t a family of probability measures.

In addition to robust trading strategies developed in a more “traditional”, stochastic integral-based mathematical finance literature, stochastic portfolio theory, such as the idea of the *functionally generated portfolio*, also enables the construction of a robust trading strategy that does not make any assumptions on the price dynamics of the underlying assets, see Fernholtz, Karatzas, Ruf [2018] [24] for an example. Recently, there is also an emergence of literature on combining tools from machine learning with traditional mathematical models that describe the dynamics of underlying financial assets to develop robust statistical tools for financial markets (in option pricing or calibration) and arbitrage. Gierjatowicz et al. [2020] [25], for example, studied the performance of *neural-SDE* where they combine feed-forward neural networks with a stochastic volatility model to construct robust hedging and pricing techniques.

Up to Section 6, we restrict ourselves to the concept of *model-independent arbitrage* via semi-static trading strategies, as introduced by Davis and Hobson [2007] [26]. The resulting trading strategy for model-independent arbitrage is commonly known as the *calendar spread*. We refer to a more elaborate discussion in Section 5.

2 Setup and Notation

For the majority of the paper (up to Section 6), we follow the same market setup as in Acciaio [2013] [27]. We consider a finite, discrete time setting with some time horizon T and a market consisting of d Markovian risky assets with price process $(S_t)_{t=0}^T = (S_t^1, \dots, S_t^d)_{t=0}^T$ and a normalized risk free asset $B = (B_t)_{t=0}^T$.

For the purpose of discussing the optimal transport-based characterization of convex order, we consider probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ with finite first moments. We define cost functionals

$$C(\mu, \rho) := \sup_{\pi \in \Pi(\mu, \rho)} \int \langle x, y \rangle \pi(dx, dy), \quad C(\nu, \rho) := \sup_{\pi \in \Pi(\nu, \rho)} \int \langle x, y \rangle \pi(dx, dy),$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product and $\Pi(\cdot, \cdot)$ denotes the set of couplings. As in the paper by Wiesel and Zhang, we define

$$\mathcal{P}^1(\mathbb{R}^d) := \{\rho \in \mathcal{P}(\mathbb{R}^d) : \text{supp}(\rho) \subseteq B_1(0)\},$$

where $B_1(0)$ denotes the Euclidean ball around 0 with a radius of 1. In addition, for two matrices \mathbf{M}, \mathbf{N} , we denote matrix multiplication simply via \mathbf{MN} . We use \mathbf{M}^T to denote the transpose of a matrix \mathbf{M} . We use bold font of mathematical symbols to denote its vector form, and we denote the discretized samples from distributions μ, ν, ρ as $\mathbf{a}, \mathbf{b}, \mathbf{r}$ respectively. Denote the discretized matrix of joint probability distribution π as γ .

3 Optimizing ρ

In this section, we discuss numerical algorithms we used in search of the optimal $\rho^* \in \mathcal{P}^1(\mathbb{R}^d)$ with domain $B_1(0)$ such that

$$\rho^* = \inf_{\rho \in \mathcal{P}^1(\mathbb{R}^d)} (C(\nu, \rho) - C(\mu, \rho))$$

Recall that we denote the discretized samples from distributions μ, ν, ρ as $\mathbf{a}, \mathbf{b}, \mathbf{r}$ respectively. If measure μ, ν are finitely supported, then we can solve the problem numerically by solving the earth mover’s distance (EMD) problem:

$$\min_{\mathbf{r} \in B_1(0)} \{-\min_{\gamma_\mu} \{M_\mu \gamma_\mu\} + \min_{\gamma_\nu} \{M_\nu \gamma_\nu\}\}$$

where M_μ, M_ν denote respectively the squared-Euclidean distance matrix between μ and ρ and ρ and ν , such that $\gamma_\mu \mathbf{1} = \mathbf{a}, \gamma_\mu^T \mathbf{1} = \mathbf{r}, \gamma_\nu \mathbf{1} = \mathbf{b}, \gamma_\nu^T \mathbf{1} = \mathbf{r}$ and $\gamma_\mu, \gamma_\nu \geq 0$, where $\mathbf{1}$ denote a column matrix of 1s.

It is clear from the definition that the optimization of measure ρ hinges on the numerical exploration of the convex set $\mathcal{P}^1(\mathbb{R}^d)$. We (Wiesel and Zhang) developed two Bayesian optimization-based methods for this. Implementation can be found in the [Github repository](#). In the implementation, we use the [POT package](#) to compute optimal transport distances and [HyperOpt](#) for Bayesian optimization.

3.1 Indirect Dirichlet Sampling

We consider finitely supported d -dimensional measures ρ on an p -evenly partitioned grid G on $B_1(0)$, which are dense in $\mathcal{P}^1(\mathbb{R}^d)$ in the Wasserstein topology. The measure ρ is sampled from the Dirichlet distribution on the space \mathbb{R}^{g-1} , $g \in \mathbb{N}$, with density

$$f(x_1, \dots, x_g; \alpha_1, \dots, \alpha_g) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^g x_i^{\alpha_i-1}$$

for $x_1, \dots, x_g \in [0, 1]$ satisfying $\sum_{i=1}^g x_i = 1$. Here $\alpha_1, \dots, \alpha_g > 0$, $\boldsymbol{\alpha} := (\alpha_1, \dots, \alpha_g)$ and $B(\boldsymbol{\alpha})$ denotes the Beta function. Fixing g grid points $\{k_1, \dots, k_g\}$ in $B_1(0)$, we are able to indirectly sample ρ by considering any realization of a Dirichlet random variable (X_1, \dots, X_g) as a probability distribution assigning probability mass X_i to the grid point k_i , $i \in \{1, \dots, g\}$, on which ρ is supported upon. We name this the “Indirect Dirichlet” method. Let us define

$$\rho(\boldsymbol{\alpha}) = \sum_{i=1}^g X_i \delta_{k_i}.$$

To sample the optimal probability measure $\rho^*(\boldsymbol{\alpha})$ parametrized by $\boldsymbol{\alpha}$, we exploit Bayesian hyperparameter optimization of parameter $\boldsymbol{\alpha}$ via a Tree-structured Parzen Estimator Approach (TPE) (Bergstra, Bardenet et al.) to optimize $\boldsymbol{\alpha}$. This leads to the following algorithm:

Algorithm 1 Basic algorithm for Indirect Dirichlet method

Input: samples (or histograms) \mathbf{a}, \mathbf{b} drawn respectively from probability measures μ, ν , maximal number of Bayesian optimization evaluations N , optimizing parameter search space lower bound l and upper bound u , number of discretized grid points g .

Initialization: $\boldsymbol{\alpha} := (\alpha_1, \dots, \alpha_g)$.

Generate a grid $B_1(0)$ of g equidistant points.

for Trial 1: N **do**

Sample a Dirichlet random variable $\rho(\boldsymbol{\alpha})$ parametrized by $\boldsymbol{\alpha}$ supported on M . Use TPE to update $\boldsymbol{\alpha} := (\alpha_1, \dots, \alpha_g)$, where

$$\boldsymbol{\alpha}^* = \arg \min_{\boldsymbol{\alpha} \in [l, u]^g} (C(\rho(\boldsymbol{\alpha}), \nu) - C(\rho(\boldsymbol{\alpha}), \mu))$$

end for

return $C(\rho(\boldsymbol{\alpha}^*), \nu) - C(\rho(\boldsymbol{\alpha}^*), \mu), \boldsymbol{\alpha}^*$.

A major computational challenge in Algorithm 1 is the efficient evaluation of $C(\rho(\boldsymbol{\alpha}), \nu)$ and $C(\rho(\boldsymbol{\alpha}), \mu)$. We offer two variants:

- *Indirect Dirichlet method with histograms:* If we have access to finitely supported approximations \mathbf{a} and \mathbf{b} of μ and ν respectively and the measure ρ is supported on G , then we can solve the linear programs $C(\mathbf{a}, \rho)$ and $C(\mathbf{b}, \rho)$ as is standard in optimal transport theory. Due to its relative lower complexity, this method enjoys a higher computational efficiency. Note that the histogram method is generally available to 1D transportation problems. We refer to `histograms.py` for a sampling from commonly seen distributions.
- *Indirect Dirichlet method with samples:* we draw from a number of samples from μ and ν respectively and denote these empirical samples as \mathbf{a}, \mathbf{b} respectively. As before, we assume that the probability measure ρ is finitely supported on a grid G . We then solve the linear programs $C(\mathbf{a}, \rho)$ and $C(\mathbf{b}, \rho)$. Because the transportation costs need to be evaluated at each sample point, this method has a higher complexity than the histograms method above, although it is useful when a histogram is computationally expensive to obtain. A more detailed discussion of analysis of runtime is discussed in section 2.4.

All in all, the efficient evaluation of global optimizer $\boldsymbol{\alpha} = (\alpha, \dots, \alpha_g)$ for problem $\inf(C(\rho(\boldsymbol{\alpha}), \nu) - C(\rho(\boldsymbol{\alpha}), \mu))$ depends on the effectiveness of the indirect Dirichlet method and the Bayesian optimization procedure to explore the convex space $\mathcal{P}^1(\mathbb{R}^d)$. For this, we offer another way to model ρ as well as propose another potential optimizing procedure, as described in the sections below.

3.2 Direct Randomized Dirichlet Sampling

An alternative to Algorithm 1 is to directly draw samples from a distribution $\rho \in \mathcal{P}^1(\mathbb{R}^d)$ based on a Dirichlet distribution. In this case, ρ is

determined by the Dirichlet distribution and then randomly sampled. We refer to this as the “Direct randomized Dirichlet method”.

Algorithm 2 Basic algorithm for Direct randomized Dirichlet method

Input: samples drawn from probability measures μ, ν , maximal number of Bayesian optimization evaluations N , optimizing parameter search space lower bound l and upper bound u , dimension d of measure $\rho \in \mathcal{P}^1(\mathbb{R}^d)$, target size t of samples drawn from ρ .

Initialization: $\alpha := (\alpha_1, \dots, \alpha_{d+1})$.

for Trial 1: N **do**

Initiate an array A to contain t samples drawn from Dirichlet random variable $D(\alpha)$ with randomized signs: $A = [\mathbf{D}_1, \dots, \mathbf{D}_t]$, where each sample \mathbf{D}_i has dimension $d + 1$. Then flatten array A .

while $\sqrt{\sum_{i=1}^t (\mathbf{D}_i)^2} > 1$ **do**

Select without replacement t d -dimensional vectors $\mathbf{D}_1, \dots, \mathbf{D}_t$.

end while

Then

$$\rho = \frac{1}{t} \sum_{i=1}^t \delta_{\mathbf{D}_i}.$$

Use TPE to update $\alpha := (\alpha_1, \dots, \alpha_{d+1})$, where

$$\alpha^* = \arg \min_{\alpha \in [l, u]^{d+1}} (C(\rho(\alpha), \nu) - C(\rho(\alpha), \mu))$$

end for

return $C(\rho(\alpha^*), \nu) - C(\rho(\alpha^*), \mu)$, α^* , $\rho(\alpha^*)$.

3.3 Optimization with cvxpy

An alternative approach to optimize ρ would be to apply convex optimization algorithms in [cvxpy](#). It is well known that the Wasserstein metric can be computed as a linear program as formulated by the Kantorovich-Rubinstein duality. For the purpose of the algorithm, we consider like before discrete distributions μ, ν on a grid G on the Euclidean ball $B_1(0)$. Using the squared L^2 norm, we still use the squared-Euclidean distance for the distance matrix M . Then the p-Wasserstein distance between μ, ν can be computed numerically:

$$W_p(\mu, \nu) = (\min_{\gamma} M\gamma)^{\frac{1}{p}} = (\min_{\gamma} \sum_{i,j} M_{i,j} \gamma_{i,j})^{\frac{1}{p}},$$

such that $\gamma \mathbf{1} = \mathbf{a}$, $\gamma^T \mathbf{1} = \mathbf{b}$, and $\gamma \geq 0$. We offer here the [cvxpy](#) algorithm for computing the Wasserstein metric between probability measures $\mu, \nu \in \mathcal{P}^1(\mathbb{R}^d)$ in general dimensions using samples. We assume uniform sample weights. And that sample size of \mathbf{a} drawn from μ is n , and that of \mathbf{b} drawn from ν is m . Also, let \mathbf{m}, \mathbf{n} denote column vector of m, n respectively.

Algorithm 3 Basic cvxpy algorithm for computing the Wasserstein metric

Input: samples \mathbf{a}, \mathbf{b} drawn from probability measure μ, ν .

Initialization: optimal coupling between μ and ν is initialized as a non-negative cvxpy variable P of shape (n, m) .

Calculate distance matrix M . Apply cvxpy solver on problem

$$\min_P \text{trace}(M^T P)$$

under constraints

$$P\mathbf{1} = 1/n, \quad P^T\mathbf{1} = 1/m.$$

return $\min_P \text{trace}(M^T P), P$

For the purpose of computing our loss function, we offer the general procedures of using cvxpy to solve our 2-fold optimization problem by approximating the Wasserstein metric as a linear program computed as in the algorithm above. To see how this works, consider first the below problem:

$$\inf_{\mu \in \mathcal{P}^1(\mathbb{R}^d)} C(\mu, \rho) + C(\rho, \nu), \quad (2)$$

which is a variation of the general Wasserstein Barycenter problem:

$$\arg \min_{\mu} \sum_{i=1}^k \lambda_i \mathcal{W}(\mu, \nu_i),$$

where μ, ν_1, \dots, ν_k are probability distributions supported on \mathbb{R}^d , vector $\lambda \in \mathbb{R}^k$ of non-negative weights summing to 1, and $\mathcal{W}(\cdot)$ denotes the squared 2-Wasserstein distance. For the implementation of the algorithm, we have the below formulation of the problem, where we have chosen L^2 -norm as the distance metric:

$$\min_{\gamma^k \in \Pi(\mu, \nu^k)} \sum_{k=1}^2 \sum_{i,j} \|X_i - Y_j\|_2^2 \gamma_{i,j}^k$$

such that $\sum_{j=1}^k \gamma_{ij}^k = \mathbf{a}_i \quad \forall_{k,i}$, $\sum_{i=1}^k \gamma_{ij}^k = \mathbf{b}_j^k \quad \forall_{k,j}$, and $\gamma_{ij}^k \geq 0 \quad \forall_{k,i,j}$.

We now give a basic working algorithm to the problem (2) in 1D with distributions μ, ν inputted as histograms over a fixed grid in a cvxpy framework. The output is the histogram \mathbf{r} of the Wasserstein Barycenter ρ over the same grid.

Algorithm 4 Basic cvxpy algorithm for Wasserstein Barycenter in 1D

Input: histogram of probability measure μ (denoted \mathbf{a}), histogram of target probability measure ν (denoted \mathbf{b}) with a fixed grid g of size n .

Initialization: probability mass function of ρ over grid g initialized as a non-negative cvxpy variable of length n ; optimal transport plan π^μ between μ, ρ and π^ν between ρ, ν initialized as non-negative cvxpy variable of shape (n, n) , and non-negative cvxpy variable t^μ, t^ν .

Calculate fixed distance matrix \mathbf{M} of grid G .

Apply cvxpy solver on problem

$$\min_{\mathbf{r}, \gamma^\mu, \gamma^\nu, t^\mu, t^\nu} t^\mu + t^\nu$$

under constraints

$$t^\mu \geq \mathbf{1}^T M \gamma^\mu \mathbf{1}, \quad t^\nu \geq \mathbf{1}^T M \gamma^\nu \mathbf{1};$$

$$(\gamma^\mu \mathbf{1})^T = \mathbf{r}, \quad (\gamma^\nu \mathbf{1})^T = \mathbf{r};$$

$$\gamma^\mu \mathbf{1} = \mathbf{a}, \quad \gamma^\nu \mathbf{1} = \mathbf{b}.$$

return \mathbf{r}

We emphasize that the above algorithm only works for 1D, where inputs are histograms. This formulation of the problem can be effectively computed in cvxpy because the grid, and as a result the cost matrix computation, is fixed and does not vary with a different choice of ρ . To relax 1D histogram into general samples, one idea is to assign samples to the nearest grid points. We will leave this exploration to later work.

Since the Wasserstein Barycenter problem is a convex problem, it can be solved in cvxpy. However, our targeted optimization problem

$$\inf_{\rho \in \mathcal{P}^1 \mathbb{R}^d} (C(\rho, \nu) - C(\rho, \mu))$$

is non-convex. Similar procedures as in Algorithm 3 applies when writing algorithm for the optimization problem. Assume that μ, ν are d -dimensional probability distributions. We give below a general algorithm of how this problem would be implemented in cvxpy, were it to be convex. As before, we let $\mathbf{m}, \mathbf{n}, \mathbf{t}$ denote column vector of natural number m, n, t respectively.

Algorithm 5 Tentative cvxpy algorithm for $\min_{\rho \in \mathcal{P}^1 \mathbb{R}^d} (C(\rho, \nu) - C(\rho, \mu))$

Input: m samples drawn from probability measure μ (denoted \mathbf{a}) n samples drawn from probability measure ν (denoted \mathbf{b}), sample size t of measure ρ (its samples denoted \mathbf{r}).

Initialization: samples \mathbf{r} drawn from target distribution ρ of shape (t, d) ; non-negative optimal coupling P_a from μ and ρ of shape (m, t) and non-negative optimal coupling P_b from ρ to ν of shape (t, n) .

Compute and express in `cvxpy` variable the squared Euclidean distance matrix M_a from μ to ρ and distance matrix M_b from ρ to ν :

$$M_a = \|\mathbf{a} - \mathbf{r}\|_2^2, \quad M_b = \|\mathbf{b} - \mathbf{r}\|_2^2$$

Apply `cvxpy` solver on problem

$$\min_{P_a, P_b} \text{trace}(M_a^T P_a) + \text{trace}(M_b^T P_b)$$

under constraints

$$P_a \mathbf{1} = 1/\mathbf{m}, \quad P_a^T \mathbf{1} = 1/t, \quad P_b \mathbf{1} = 1/t, \quad P_b^T \mathbf{1} = 1/n.$$

return \mathbf{r}, P_a, P_b

It is noteworthy that this is not a running algorithm due to the non-convexity of our cost function. Two restrictions are present. One is that by formulating the problem in a two-fold optimization problem (unlike the previous 1D formulation of the barycenter problem), the cost matrix becomes convoluted and is not DPE compliant. This is a problem when one tries to write a generalized dimension Wasserstein Barycenter solver in `cvxpy`, as for Algorithm 4, a fixed distance metric is key. While this is solvable as in the proposal above, the non-convexity of the problem fundamentally prevents it from being implemented in `cvxpy`.

3.4 Complexity and Performance

We divide the section into two parts: (i). a simple overview and analysis of algorithm run-time and complexity and (ii). a presentation of experiments and results.

3.4.1 Analysis of Complexity

All three of our working algorithms are based on the Bayesian optimization solver by `HyperOpt`. We begin with a discussion of the mechanism underlying Bayesian optimization and an analysis of its complexity using the adapted tree-structured parzen estimator approach (TPE). For a more elaborate discussion, we refer to Bergstra et al.[2013] [28] and Shahriari et al.[2016] [29].

Bayesian optimization is a type of hyper-parameter optimization which tackles the problem of optimizing a loss function over a graph-structured

configuration space (or observation history) $\mathcal{H} = (x_i, f(x_i))_{i=1}^n$. In Hyper-Opt, a configuration space serves as the search space of hyper-parameters and is defined by certain generative process that draws valid samples conforming to constraints. The structure of the Bayesian optimization algorithm follows the prototype of Sequential Model-Based Global Optimization (SMBO) algorithms, which is commonly used to search for a global minimizer (or maximizer) of an unknown or costly fitness function $f : \mathcal{X} \rightarrow \mathbb{R}$:

$$\boldsymbol{\alpha}^* = \arg \min_{\boldsymbol{\alpha} \in \mathcal{X}} f(\boldsymbol{\alpha}).$$

In our case, $f(\boldsymbol{\alpha}) = C(\rho(\boldsymbol{\alpha}), \nu) - C(\rho(\boldsymbol{\alpha}), \mu)$ is a fitness function.

SMBO-based Bayesian optimization optimizes $\boldsymbol{\alpha}^*$ by prescribing a prior belief over the possible objective functions that are then sequentially refined via Bayesian posterior updating. This updating process is guided by sequentially induced acquisition functions $a_n : \mathcal{X} \rightarrow \mathbb{R}$, which evaluates the utility of candidate points for the next evaluation of f and decides which hyper-parameters to evaluate next.

For the adapted tree-structured parzen estimator approach (TPE), the acquisition functions are guided by the Expected Improvement (EI) criterion. Let y^* be certain threshold value and let $y = f(\boldsymbol{\alpha})$, Expected Improvement utilizes the utility function

$$u(y) = \max(0, y^* - y)$$

and evaluates f at the point that, in expectation under some model M of the hyper-parameters, makes the largest magnitude of improvement on the benchmark value y^* . The EI acquisition function is thus specified by the following:

$$EI_{y^*}(\boldsymbol{\alpha}) := \int_{-\infty}^{y^*} (y^* - y) p_M(y|\boldsymbol{\alpha}) dy.$$

The TPE algorithm models $p(\boldsymbol{\alpha}|y)$ through Bayes' rule $p(\boldsymbol{\alpha}|y) = \frac{p(y|\boldsymbol{\alpha})}{p(y)}$ and using non-parametric densities for the distributions of the configuration prior. Let l, u denote lower bound and upper bound respectively. In our algorithms, we describe the configuration space of target variable $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$ by modeling them as uniform variables over $[l, u]^n$. Using different observations $\{\boldsymbol{\alpha}^{(1)}, \dots, \boldsymbol{\alpha}^{(k)}\}$ in $\text{Unif}([l, u]^n)$, the TPE algorithm substitutes the configuration prior by a truncated Gaussian mixture through a learning process that defines $p(\boldsymbol{\alpha}|y)$ using two densities over the configuration space \mathcal{X} :

$$p(\boldsymbol{\alpha}|y) = \begin{cases} l(\boldsymbol{\alpha}), & \text{if } y < y^* \\ g(\boldsymbol{\alpha}), & \text{otherwise} \end{cases}$$

Here $l(\boldsymbol{\alpha})$ is the density formed via the “good” observations $\{\boldsymbol{\alpha}^{(i)}\}$ whose associated loss $f(\boldsymbol{\alpha}^{(i)}) < y^*$ whereas $g(\boldsymbol{\alpha})$ is the density formed via the “bad” remaining observations. The TPE algorithm avoid choosing a minimizing point less than the actual minimizer by choosing y^* to be certain quantile γ of the observed y values such that $p(y < y^*) = \gamma$. This parameter γ determines the size of the split between the “good” observation group and the “bad” ones. Thus, as stated in section 4.1 (Bergstra,

Bardenet et. al) the EI-based acquisition function for TPE is proportional the following:

$$EI_y^* \propto (\gamma + \frac{g(\boldsymbol{\alpha})}{l(\boldsymbol{\alpha})}(1 - \gamma))^{-1}.$$

The proof to this result is contained in the Appendix A.

Algorithm 6 Basic algorithm for Bayesian optimization via TPE

Input: fitness function f , initial model M of hyper-parameter prior distribution, threshold quantile γ , acquisition function S , maximum number of evaluations N .

Initialization: initial acquisition model $M_0 = M$, configuration space $\mathcal{H} = \emptyset$.

for Trial 1 : N **do**

 Sample a set H of parameter $\boldsymbol{\alpha}$ from M_{t-1} .

 Evaluate $f(\boldsymbol{\alpha})$ for each one of them.

 Split H into good and bad groups (defined above) based on γ .

 Update $\boldsymbol{\alpha}^*$ via the following

$$\arg \min_{\boldsymbol{\alpha}} S(M_{t-1}, \boldsymbol{\alpha}) := \arg \min_{\boldsymbol{\alpha}} (\gamma + \frac{g(\boldsymbol{\alpha})}{l(\boldsymbol{\alpha})}(1 - \gamma))^{-1},$$

where $\boldsymbol{\alpha}$ is selected from the model M_{t-1} .

 Augment $\mathcal{H} = \mathcal{H} \cup (\boldsymbol{\alpha}^*, f(\boldsymbol{\alpha}^*))$.

 Do a Bayesian update on M_t based on observation history \mathcal{H} .

end for

return $\boldsymbol{\alpha}^*, \mathcal{H}$

Alternatives to TPE include the Gaussian Process Approach (GP), mutual search, and random search. Bergtra et al. [2013] [28] study the performance of each of the Bayesian optimization alternatives and found that while on the convex dataset, both TPE and GP algorithms converge to a validation score (as defined in Bergtra et al. [2013] [28]) of 13%, in generalization, TPE’s best model had 14.1% error, outperforming GP by 2.6%, mutual search by 4.9%, and random search by 2.9%. In general Bergtra et al. found that models found by the TPE algorithm perform better on the experiment datasets than its alternatives. TPE-based Bayesian hyperparameter optimization is used widely in learning models and shows a steady performance, see Dzikienė et al. [2020] [30]. Given the typical performance of TPE-based Bayesian optimization algorithm, we choose a TPE-based approach in the numerical implementation.

The complexity of the presented algorithms depends roughly on three components: the complexity of TPE-based Bayesian optimization framework, the complexity of optimizing the acquisition function, and the complexity of the evaluation of the fitness function. From the discussion in Section 4.1 of Bergstra et al. [2011] [31], the TPE algorithm maintains sorted lists of observed variables in \mathcal{H} , and for each iteration of evaluations, the TPE algorithm scales linearly in the number of variables observed and linearly in the number of variables being optimized. In our case,

discounting the costs of numerical evaluations of fitness function, the Indirect Dirichlet algorithm (both histogram and samples-based) has a cost of $O(N|H|p^d)$, whereas the Direct Dirichlet algorithm has a cost proportional to $O(N|H|d)$, where H is the size of sampled hyper-parameters for each trial and N is the maximum number of evaluations.

The minimization of the acquisition function typically uses gradient descent. The time complexity of the gradient descent algorithm for minimizing a function depends on the size of the input and the number of iterations required to converge to a minimum. In general, the time complexity of the gradient descent algorithm is $O(kd)$, where k is the number of iterations required to converge and d is the dimensionality of the input.

The most expensive step of the computation, however, resides in the evaluation of the fitness function. As mentioned before, our fitness function is $f(\alpha) = C(\rho(\alpha), \nu) - C(\rho(\alpha), \mu)$ for $\alpha \in [l, u]^d$. In both algorithms, we phrase the problem in terms of a linear program, the details of which can be found in previous discussions. Python package [POT](#) implements one of the most efficient exact OT solvers. Our algorithms leverage the regular OT solver `ot.emd2`. When the input measures are coded as histograms, the OT problem is represented as a linear programming problem and solved with a runtime complexity of $O(n^3)$, where n is the number of bins in the histograms. The computation of the distance matrix between the histograms required by `ot.emd2` is done in $O(n^2)$ time complexity using a metric such as the Euclidean distance. On the other hand, when the input measures are coded as samples, the computation has a runtime complexity of $O(n^3 \log(n))$ with the `ot.emd2` solver, although the computation of the distance matrix between the samples remains $O(n^2)$. This is because whereas the histogram-based algorithm uses a fixed grid and thus requires only one calculation of cost matrix, sample-based algorithm calculates the cost matrix in each iteration and scales linearly in the maximum number of evaluations. As shown in the section below, for $d = 1, 2$ and a medium samples size $n = 100$, the histogram-based Indirect Dirichlet method yields consistently a much lower run-time than sample-based Indirect Dirichlet method and Direct Dirichlet method.

Discounting costs of other computations, for large n , the run-time of the Indirect Dirichlet method with histograms is $O(N|H|p^d n^3)$, the Indirect Dirichlet method with samples is $O(N|H|p^d (n^3 \log(n)))$, and the complexity of the Direct Dirichlet method is $O(N|H|d(n^3 \log(n)))$. Hence, when n is large and d small (as in our experiments), the Indirect Dirichlet method with histograms performs the best in term of runtime. However, when d is large, the Direct Dirichlet method will outperform the alternatives.

For a large scale OT problem in high dimension, POT's Sinkhorn solver `ot.sinkhorn` has a complexity of $O(n^2)$. Define

$$\Omega(\gamma) = \sum_{i,j} \gamma_{i,j} \log(\gamma_{i,j}),$$

then `ot.sinkhorn` solves the entropic regularization optimal transport problem

$$\min_{\gamma \in \Pi(\mu, \nu)} \{ \gamma M + \text{reg} \cdot \Omega(\gamma) \}$$

such that $\gamma \mathbf{1} = \mathbf{a}$, $\gamma^T \mathbf{1} = \mathbf{b}$, $\gamma \geq 0$. Generally, `ot.sinkhorn` speeds up the numerical computation. However, in our case where n and d are small, the improvement in runtime of `ot.sinkhorn` is not obvious. We thereby choose the regular OT solver `ot.emd2` for our algorithms.

3.4.2 Experiments and Results

We test the performance of algorithms presented above by looking at their ability of detecting the convex order. As in the paper by Wiesel and Zhang, we present three simple cases in which the absence or presence of the convex order is easy to see:

- **Example 1.** $\mu = \mathcal{N}(0, \sigma^2 I)$ and $\nu = \mathcal{N}(0, I)$ for $\sigma^2 \in [0, 2]$ for $d = 1, 2$.
- **Example 2.** $\mu = \frac{1}{2}(\delta_{-1-s} + \delta_{1+s})$ and $\nu = \frac{1}{2}(\delta_{-1} + \delta_1)$ for $s \in [-1, 1]$.
- **Example 3.** $\mu = \frac{1}{4}(\delta_{(-1,0)} + \delta_{(1,0)} + \delta_{(0,1)} + \delta_{(0,-1)})$ for $s \in [-1, 1]$.

Let

$$V(\mu, \nu) := \inf_{\rho \in \mathcal{P}^1(\mathbb{R}^d)} (C(\nu, \rho) - C(\mu, \rho)).$$

Then by Theorem 1.2 (Wiesel and Zhang), we have the relationship

$$\mu \preceq_c \nu \text{ iff } V(\mu, \nu) \leq 0.$$

For each example and each pair of (μ, ν) , we plot $V(\mu, \nu)$ for the Bayesian-optimization-based three methods above: indirect Dirichlet with histograms, indirect Dirichlet with samples, and direct Dirichlet with samples. We refer to the github repository for a more elaborate discussion.

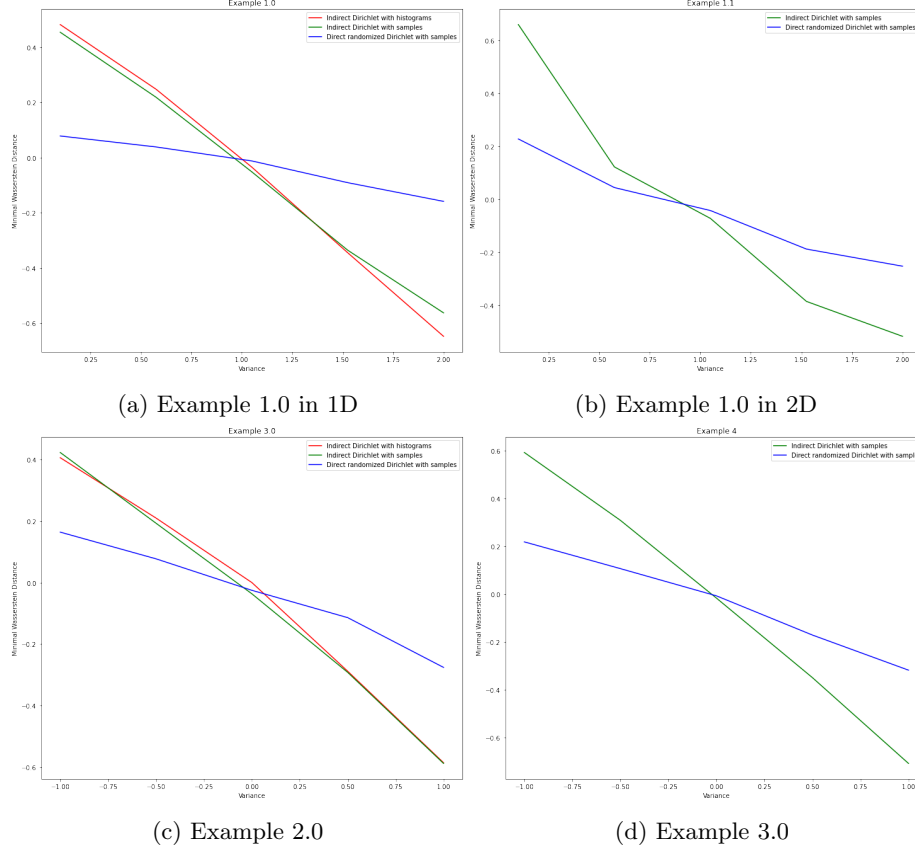


Figure 1: Values of different estimators of $V(\mu, \nu)$ plotted against σ for Examples 1-3. All plots use $N = 100$ samples.

Discounting the numerical errors in sampling, all estimators seem to detect convex order. The direct randomized Dirichlet method is numerically easier to implement; however, it does not seem to explore the $\mathcal{P}^1(\mathbb{R}^d)$ -space as well as the other two alternatives. We leave the analysis of this deficiency to another paper.

As explained in Section 3.4.1, since we only conducted experiments in dimensions we expect that the Indirect Dirichlet method with Histograms to consistently yield the lowest runtime and the runtimes of the other methods to be much higher. This is indeed the case for our experiments. To recall, this is because when working with samples, the weights of the empirical distributions are constant, while the OT cost matrices \mathbf{M}_a and \mathbf{M}_b in the implementation have to be re-computed in each iteration, which is very costly. For the histogram method, the weights ρ change, while the grid stays constant – and thus \mathbf{M}_a and \mathbf{M}_b .

4 Recovering Convex Function \hat{f}

Fix a $\rho \in \mathcal{P}^1(\mathbb{R}^d)$. The Monge-Kantorovich duality states that

$$\begin{aligned} C(\mu, \rho) &= \sup_{\pi \in \Pi(\mu, \rho)} \int \langle x, y \rangle \pi(dx, dy) \\ &= \inf_{f \oplus g \geq c} \int f d\mu + \int g d\rho \\ &= \inf_{f \oplus g \geq c, f, g \text{ proper, convex}} \int f d\mu + \int g d\rho \end{aligned}$$

Take an optimal convex pair (\hat{f}, \hat{g}) from above for $C(\mu, \rho)$. For the 2-Wasserstein distance, Brenier's theorem in the form of Villani [2003, 2008] Theorem 2.12 [12] gives that the optimal transport map is given by the gradient of a scalar convex function \hat{f} : $\rho = \nabla \hat{f}_* \nu$ if ν is absolutely continuous with the Lebesgue measure. In this section, we present algorithm and graphic example of recovering the convex function \hat{f} .

Given an optimal $\rho \in \mathcal{P}^1(\mathbb{R}^d)$ found through Section 3 in the form of m generated samples, we obtain the optimal transport plan $\pi(\nu, \rho)$. We recover the transport map $\nabla \hat{f}$ through disintegration. Specifically, ∇f is recovered through taking conditional expectation $\int x \pi_y(dx)$, where $(\pi_y)_{y \in \mathbb{R}^d}$ denotes the conditional probability distribution of π with respect to its second marginal ν , i.e.

$$\pi_y = \delta_{\nabla \hat{f}(y)} = \pi(x|y).$$

This is a standard technique (see e.g. Deb et al.[2021] [32] for details). We then use methods from `scipy.interpolate` to interpolate the gradient field $\nabla \hat{f}$ in d -dimensions.

To obtain the convex function \hat{f} , we consider the following problem. Given a gradient vector field

$$\nabla \hat{f}(x_1, \dots, x_d) := \left(\frac{\partial \hat{f}}{\partial x_1} \dots \frac{\partial \hat{f}}{\partial x_d} \right),$$

we want to compute the inverse gradient and to recover the scalar field $\hat{f} : \mathbb{R}^d \rightarrow \mathbb{R}$. In the below sections, we introduce algorithms for the computation of the inverse gradient for $d = 1, 2$.

4.1 Recovering \hat{f} in 1D

Consider a gradient $\nabla \hat{f}(x) : \mathbb{R} \rightarrow \mathbb{R}$. Then $\nabla \hat{f}(x) = \frac{d\hat{f}}{dx}$. Hence, \hat{f} is easily recovered up to an additive constant through the Fundamental Theorem of calculus by integrating $\nabla \hat{f}(x)$ with respect to x :

$$\hat{f}(x) = \int_c^x \nabla \hat{f}(y) dy$$

where c is the empirical lower bound.

4.1.1 Algorithm for 1D

To numerically compute the integral above, we use method `integrate` from `InterpolatedUnivariateSpline`. To make the pseudo-code clearer, we use list index notation. That is, given a list l , $l[i]$ denotes the i -th element of l . A matrix M in this case is represented by a two-dimensional array, and each element i, j can be accessed through calling index: $M[i][j]$. This leads to the following algorithm.

Algorithm 7 Basic algorithm for recovering \hat{f} in 1D

Input: n samples drawn from measure ν (denoted \mathbf{b}) and m samples drawn from the optimal measure ρ .

Initialization: weights of measure μ and ν are initialized to be uniform, empty list l , empty list l_y .

Compute cost matrix M from the samples and calculate optimal transport plan π via `ot.emd`.

for iteration i $1 : m$ **do**

Append $\frac{\rho[i] \cdot \mathbf{G}[i]}{\sum_j \rho[j] \cdot \mathbf{G}[i][j]}$ to l .

end for

for iteration j $1 : n$ **do**

Append $\text{mean}(l^T[j])$ to l_y .

end for

Interpolate gradient of \hat{f} :

$$\nabla \hat{f} = \text{InterpolatedUnivariateSpline}(\mathbf{b}, l_y, k = 1).$$

Obtain \hat{f} by integrating each sample in \mathbf{b} from empirical lower bound $\min(\mathbf{b})$ via `$\nabla \hat{f}.\text{integral}$` .

return $\nabla \hat{f}, \hat{f}$.

We offer two methods of smoothing the resulting graph of \hat{f} : through building an interpolated spline via `scipy.interpolate` and using lowess via `statsmodels.api.nonparametric`. Since both methods yield similar results, we will include only the lowess-based graphs in the below section.

4.1.2 Experiments and Results for 1D

We implement Algorithm 7 taking 100 samples from $\mu = \mathcal{N}(0, 1)$ and 100 samples from $\nu = \mathcal{N}(0, 5)$. We choose a partition grid of 100 for the computation of ρ and the maximum number of evaluations for Bayesian optimization to be 100.

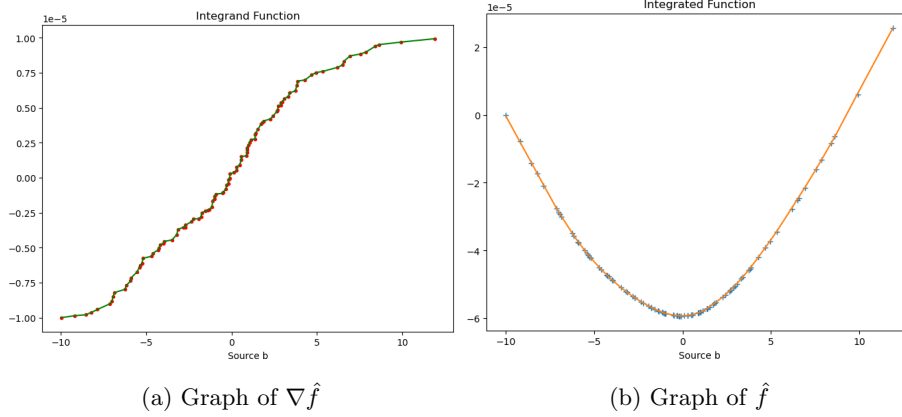


Figure 2: Graph of $\nabla \hat{f}$ and \hat{f} for example 4.1.2

As expected, the integrand $\nabla \hat{f}$ appears to be affine and the recovered function \hat{f} convex and follow the form $\hat{f}(x) = ax^2$ for some scalar $a \in \mathbb{R}$.

4.2 Recovering \hat{f} in 2D

Consider the gradient $\nabla \hat{f}(x, y) := \left(\frac{\partial \hat{f}}{\partial x}, \frac{\partial \hat{f}}{\partial y} \right)$. In implementation, we observe an empirical vector field g of \hat{f} on some bounded and connected domain $\Omega \in \mathbb{R}^n$. We expect that the real gradient field $\nabla \hat{f}(x, y)$ to be near g . Then numerically, to recover $\hat{f}(x, y)$ with its gradient field near some given observed vector field g , we consider the minimization of $\|\nabla \hat{f} - g\|$ for some choice of norm and over some class of scalar functions mapping from \mathbb{R}^2 to \mathbb{R} . Consider the $L^2(\Omega)$ -norm of the pointwise Euclidian norm, we have the below minimization problem:

$$\arg \min_{\hat{f}} \|\nabla \hat{f} - g\|^2 = \int_{\Omega} |\nabla \hat{f}(\mathbf{x}) - g(\mathbf{x})|^2 d\mathbf{x},$$

where $\Omega \subset \mathbb{R}^n$ is bounded and connected. From the calculus of variation, it is well established that (see Farnebäck[2007] [33], Song et al. [1995] [34]) the inverse gradient $\hat{f}(x, y)$ is the solution to a Poisson equation with inhomogeneous Neumann boundary conditions,

$$\begin{cases} \nabla^2 \hat{f}(\mathbf{x}) = \nabla \cdot g(\mathbf{x}), & \forall \mathbf{x} \in \Omega \\ \nabla \hat{f}(\mathbf{x}) \cdot \mathbf{n} = \mathbf{n} \cdot g(\mathbf{x}), & \forall \mathbf{x} \in \partial\Omega, \end{cases}$$

where $\partial\Omega$ is the boundary of Ω , \mathbf{n} is a normalized and outwards directed normal vector to the boundary, and ∇^2 is the Laplace operator. By Theorem 3 in Evans [2010] [35], the necessary condition for the existence of a solution to the Neumann problem, as in our case, is that

$$\int_{\Omega} -\nabla g(\mathbf{x}) d\mathbf{x} = - \int_{\partial\Omega} g dS,$$

where $-\int_{\partial\Omega} g dS$ denotes the line integral with respect to the boundary $\partial\Omega$. Note that this is satisfied because of the Generalized Stoke's theorem, which states that

$$\int_{\Omega} d\beta = \oint_{\partial\Omega} \beta dS$$

for any differential form β and domain Ω . Thus, we are guaranteed to find a solution.

4.2.1 Algorithm for 2D

Numerical methods to solve the Poisson equation with Neumann boundary conditions are well developed. Farnebäck et al. [2007] [33], for example, developed efficient multigrid-based solver to solve the above Poisson equation with an inhomogeneous Neumann boundary conditions over irregular domains. In our implementation, we work with python package [dolfin](#).

Algorithm 8 Basic algorithm for recovering \hat{f} in 2D

Input: n samples drawn from measure ν (denoted \mathbf{b}) and m samples drawn from the optimal measure ρ .

Initialization: weights of measure μ and ν are initialized to be uniform, empty list l , empty list l_y , [dolfin](#) function space with Lagrange multiplier W .

Compute cost matrix \mathbf{M} from ν to ρ and calculate optimal transport plan \mathbf{G} via `ot.emd`.

for iteration i $1 : n$ **do**
Append $\frac{\rho \cdot \mathbf{G}[i]}{\sum_j \mathbf{G}[i][j]}$ to l .

end for

Initialize gradient domain Ω based on observed gradient points. Set empirically observed gradient field $g = l$.

Set trial function (represents the unknown \hat{f}) to be (\hat{f}, \hat{c}) and test function to be (f, c) , where \hat{c}, c represent the respective constant terms. Solve variational problem using [dolfin](#):

$$\nabla \hat{f} \cdot \nabla f + \hat{c}f + c\hat{f}dx = \nabla g f dx + (\mathbf{n} \cdot g) f ds.$$

return $\nabla \hat{f}, \hat{f}$.

For graphing the resulting function $\hat{f} : \mathbb{R}^2 \rightarrow \mathbb{R}$ on irregular domains consisting of the empirical samples from ν , we use python package `matplotlib.tri`. `matplotlib.tri` interpolates and generates a 3D surface by using a Delaunay triangulation. The resulting graph is presented in the section below.

4.2.2 Experiments and Results for 2D

We implement algorithm 8 on the same example in section 4.2.1 in 2D. That is, $\mu = \mathcal{N}(0, 5I)$ and $\nu = \mathcal{N}(0, I)$. We choose a partitioned grid of

50 for the computation of ρ and the maximum number of evaluations for Bayesian optimization to be 100.

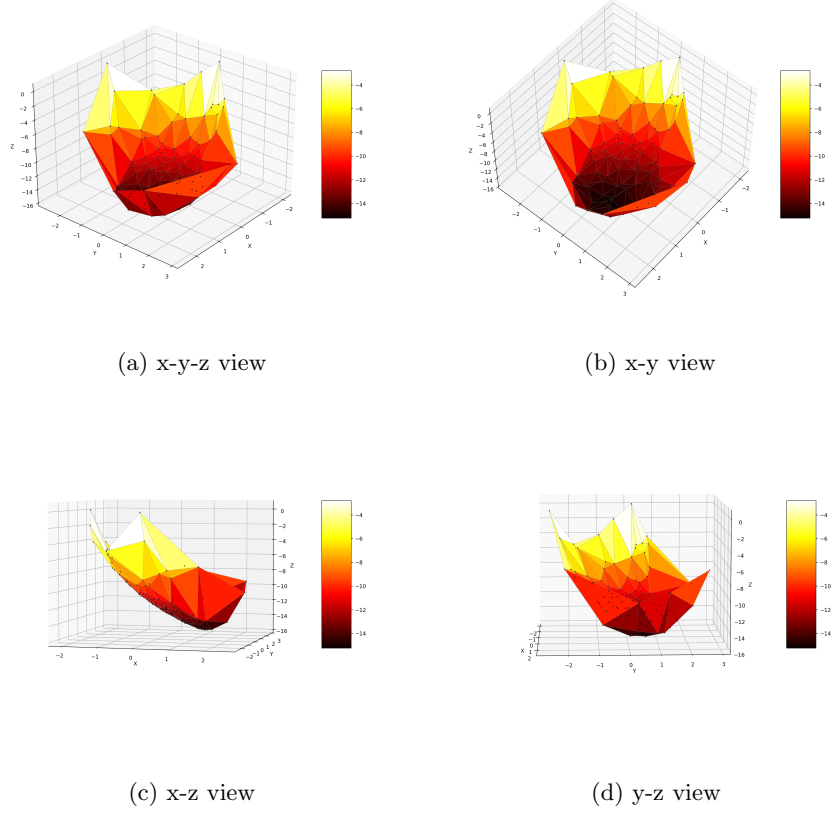


Figure 3: Graph of \hat{f} for example 4.2.2

As expected, the resulting \hat{f} is convex and has the form of a parabola, analogous to the example in 1D.

5 Model-free arbitrage strategy via \hat{f}

Recall the market setup established in Section 2 and consider the d financial assets with price processes $(S_t)_{t \geq 0} = (S_t^1, \dots, S_t^d)_{t \geq 0}$. Fix two maturities $T_1 < T_2$. Assume that call options with these maturities are traded at all strikes, then by the Breeden-Litzenberger formula (introduced in Section 1.2), the prices of the call options determine the distribution of S_{T_1} and S_{T_2} under any equivalent martingale measures. Let us denote the distributions of S_{T_1} and S_{T_2} by μ and ν respectively. We then construct model-free arbitrage strategy using the convex function recovered in Section 3 between μ and ν , where *model-independent arbitrage*

is understood w.r.t semi-static trading strategies. To establish the result, we recall a few definitions.

Definition 5.0.1 (*Martingale coupling*). Let $\mathcal{M}(\mu, \nu)$ denote the set of all martingale laws π with marginals μ, ν . That is, probability measure π such that

$$x \sim \mu, \quad y \sim \nu, \quad \mathbb{E}^\pi[y|x] = x.$$

We call $\mathcal{M}(\mu, \nu)$ the set of all martingale couplings between μ, ν .

Definition 5.0.2 (*Semi-static trading strategy*). A trading strategy $\Delta = (\Delta_t)_{t=0}^{T-1}$ is a semi-static strategy if it consists of a static portfolio in finitely many options whose prices are known at time zero, and a dynamic, self-financing strategy in the underlying assets.

As mentioned above, we define model-independent arbitrage w.r.t semi-static trading strategies. We follow definition 2.1 from Davis and Hobson [2007] [26].

Definition 5.0.3 (*Model-independent arbitrage*). A model-independent arbitrage exists when there exists a semi-static portfolio with zero initial value and with strictly positive value at the terminal date across all possible scenarios.

Our set up for model-independent arbitrage is motivated by the above definitions. Suppose that trading is only allowed at $0, T_1, T_2$, we give the following adapted version of model independent arbitrage. This is stated in Definition 5.1 by Wiesel and Zhang [2023] [1].

Definition 5.0.4 (*Adapted model-independent arbitrage*). Consider trading strategy Δ . The triple of Borel measurable functions (u_1, u_2, Δ) forms a model-independent arbitrage if $u_1 \in L^1(\mu)$ and $u_2 \in L^1(\nu)$ such that

$$u_1(x) - \int u_1 d\mu + u_2(y) - \int u_2 d\nu + \Delta(x)(y - x) > 0, \quad \forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d.$$

If no such strategies exist, then we call the market free of model-independent arbitrage.

Building on Theorem 3.4 from Guyon et al. [2017] [36], we present the following theorem connecting model-independent arbitrage and convex order between μ, ν . This is stated in Theorem 5.2 and its proof in Wiesel and Zhang [2023] [1].

Theorem 5.0.1. *The following are equivalent:*

1. *The market is free of model-independent arbitrage.*
2. $\mathcal{M}(\mu, \nu) \neq \emptyset$.
3. $\mu \preceq_c \nu$.

In particular, if $\mu \preceq_c \nu$, then there exists a convex function f , such that the triple $(-f(x), f(y), g(x))$ is a model-independent arbitrage. Here g is a measurable selector of the subdifferential of f .

We give a proof of Theorem 5.0.1.

Proof of Theorem 5.0.1. (ii) \leftrightarrow (iii) is stated in Strassen's theorem. If $\mu \not\leq_c \nu$, then by definition of convex order there exists a convex function f such that

$$\int f d\mu > \int f d\nu.$$

On the other hand, f is convex and thus satisfies

$$f(y) - f(x) \geq g(x)(y - x) \quad \forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d.$$

Combining the two equations above shows that $(-f(x), g(x), g(x))$ is a model-independent arbitrage and thus (i) \rightarrow (ii). It remains to show that (ii) \rightarrow (i). Taking expectations in the inequality

$$u_1(x) - \int u_1 d\mu + u_2(y) - \int u_2 d\nu + \Delta(x)(y - x) > 0, \quad \forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d$$

under any martingale measure with marginals μ, ν leads to a contradiction. This concludes the proof. \square

In particular, a model-independent arbitrage strategy is given by the triple $(-\hat{f}(x), \hat{f}(y), \nabla \hat{f}(x))$. This is commonly known as the calendar spread, which is a trading strategy that involves buying and selling options or futures contracts with different expiration dates but the same underlying asset and strike price. In particular, once ρ is found numerically (see section 3), we need only to estimate $\nabla f(x)$ from the optimizing transport plan of $C(\rho, \nu)$ in order to obtain an explicitly arbitrage strategy (see section 4).

6 Arbitrage between Portfolios

Another exploitation of the absence of convex order in the market for arbitrage opportunities can be found in the arbitrage between two Functionally Generated Portfolios. In this section, we introduce Functionally Generated Portfolio in Stochastic Portfolio Theory and propose a tentative link between convex order and Functionally Generated Portfolios. Specifically, we propose that the “convex order” at play in the arbitrage strategy is a different measurement of the volatility of market, captured by the constructed gamma process in Fernholz, Karatzas et al. [2017] [24]. We begin by introducing the Functionally Generated Portfolio in Stochastic Portfolio Theory and setting up relevant definitions. Note that the market setup and notation can be different from the previous sections.

6.1 Functionally generated portfolio

Functionally Generated Portfolio is the method of constructing trading strategies through functional generation. This is an important construction in Stochastic Portfolio Theory established by Robert Fernholz. Since then, Karatzas and Karadaras (see *Portfolio theory and Arbitrage* [2021] [37]) and Karatzas and Ruf [2017] [38] etc. further developed the theory systematically and simplified the Functionally Generated Portfolio method by proposing additive and multiplicative functional generation methods. We begin by setting up the relevant market model considered in the Functionally Generated Portfolio framework by defining key concepts.

6.1.1 Market model

On a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, endowed with a right continuous filtration $(\mathcal{F})_{t \geq 0}$ that satisfies $\mathcal{F}(0) = \{\emptyset, \Omega\}$, we consider a stochastic process $S = (S_1, \dots, S_d)$ of continuous non-negative semi-martingales with strictly positive initial values for each component. Now consider a frictionless equity market composed of fixed d assets and a money market. Let the vector process S_i represents the price of asset i where $i = 1, 2, \dots, d$. Define the process

$$\Sigma(t) := S_1(t) + \dots + S_d(t), \quad t \geq 0.$$

Then this process is interpreted as the market capitalizations of a fixed number $d \geq 2$ of companies in the equity market. We further impose the constraint that the total capitalization Σ of the equity market satisfies

$$\mathbb{P}[\Sigma(t) > 0, \quad \forall t \geq 0] = 1.$$

Now define the vector process $\pi = (\pi_1, \dots, \pi_d)^T$ that consists of the respective market one d companies' relative market weight processes

$$\pi_i(t) := \frac{S_i(t)}{\Sigma(t)} = \frac{S_i(t)}{S_1(t) + \dots + S_d(t)}, \quad t \geq 0, i = 1, \dots, d.$$

Then each components of π are continuous, nonnegative semimartingales, each of them takes values in the unit interval $[0, 1]$, and they satisfy $\pi_1 + \dots + \pi_d = 1$. More formally, the continuous d -dimensional semi-martingale $\pi(t) = (\pi_1(t), \dots, \pi_d(t))$ takes values in the lateral face

$$\Delta^d := \{(x_1, \dots, x_d) \in [0, 1]^d : \sum_{i=1}^d x_i = 1\} \subset \mathbb{H}^d$$

of the unit simplex, where \mathbb{H}^d denotes the hyperplane

$$\mathbb{H}^d := \{(x_1, \dots, x_d) \in \mathbb{R}^d : \sum_{i=1}^d x_i = 1\}.$$

We assume that the initial weights $\pi^i(0)$ for $i \in \{1, \dots, d\}$ are non-negative.

6.1.2 Trading strategies

We consider predictable processes $\psi = (\psi_1, \dots, \psi_d)$ and $\phi = (\phi_1, \dots, \phi_d)$ with values in \mathbb{R}^d . We define ψ_i and ϕ_i as the number of shares of stocks of company $i = 1, \dots, d$ at time $t \geq 0$. Then the associated wealth processes of the investment under strategy ψ and ϕ are

$$V^\psi(\cdot; S) := \sum_{i=1}^d \psi_i S_i$$

and

$$V^\phi(\cdot; S) := \sum_{i=1}^d \phi_i S_i.$$

We now give the definition of trading strategies in our setting.

Definition 6.1.1 (*Trading strategies*). Define $\mathcal{L}(\cdot)$ to be the set of admissible trading strategies (or portfolios) of the equity market S (or market weights π). Then for ψ and ϕ to be admissible trading strategies with respect to equity market S if they are “self-financed”, i.e.

$$V^\psi(\cdot; S) - V^\psi(0; S) = \int_0^\cdot \sum_{i=1}^d \psi_i(t) dS_i(t).$$

$$V^\phi(\cdot; S) - V^\phi(0; S) = \int_0^\cdot \sum_{i=1}^d \phi_i(t) dS_i(t).$$

In this case, we write $\psi, \phi \in \mathcal{L}(S)$. Moreover, it is important to note that by Proposition 2.3 in Karatzas and Ruf [2017] [38], $\mathcal{L}(S) = \mathcal{L}(\pi)$, i.e. a \mathbb{R}^d -valued previsible process ψ is a trading strategy with respect to the semimartingale S iff it is a trading strategy with respect to the market weight semimartingale π as defined previously. We write in this case

$$V^\psi(\cdot; S) = \Sigma(\cdot) V^\psi(\cdot; \pi)$$

and

$$V^\phi(\cdot; S) = \Sigma(\cdot) V^\phi(\cdot; \pi).$$

6.1.3 Proportional Investment

By our setup, for an admissible trading strategy $v = (v_1, \dots, v_d) \in \mathcal{L}(S)$, the associated wealth process

$$V^v(\cdot; S) = V^v(0; S) + \int_0^\cdot \sum_{i=1}^d v_i(t) dS_i(t)$$

is a numeraire by definition. Since $V^\psi(\cdot; S), V^\phi(\cdot; S) > 0$, we can define new predictable, vector-valued processes $\mu = (\mu_1, \dots, \mu_d)^T \in [0, 1]^d$ and $\nu = (\nu_1, \dots, \nu_d)^T \in [0, 1]^d$ with respect to ψ, ϕ in the following manner:

$$\mu_i(t) := \frac{S_i(t) \psi_i(t)}{V^\psi(t; S)}$$

$$\nu_i(t) := \frac{S_i(t) \phi_i(t)}{V^\phi(t; S)}$$

for each $i \in \{1, \dots, d\}$. We interpret μ and ν as two trading portfolios in the d -asset market, where μ_i and ν_i each represents the proportion of current wealth at time $t \geq 0$ invested in the stock of company $i = 1, \dots, d$.

6.1.4 Additively generated portfolios

Fernholtz, Karatzas, and Ruf [2018] [24] establish a special class of trading strategies for which the representation in 6.1.2 takes a very simple form. These trading strategies are classified in two groups: the additive functional generation and the multiplicative functional generation. In this paper, we concern ourselves with only the additive Functionally Generated Portfolios. To introduce this concept, we begin with a definition of regular function following Definition 3.1 in Karatzas and Ruf [2017] [38].

Definition 6.1.2 (*Regular function*). A continuous mapping $G : \Delta^d \rightarrow \mathbb{R}$ is called a regular function with respect to some vector process $(\pi(t))_{t \geq 0}$ of relative market weights, if the process $G(\pi(t))$ is a semimartingale of the form

$$G(\pi(t)) = G(\pi(0)) + \int_0^t \sum_{i=1}^d D_i G(\pi(s)) d\pi_i(s) - \Gamma^G(t)$$

for some measurable function $DG = (D_1 G, \dots, D_d G) : \Delta^d \rightarrow \mathbb{R}^d$ and a continuous, adapted process $(\Gamma^G(t))_{t \geq 0}$ of finite variation on compact time intervals. If the continuous function G can be extended to a twice continuously differentiable function, elementary stochastic calculus then expresses the gamma process as:

$$\Gamma^G(t) = \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \int_0^t D_{ij}^2 G(\pi(s)) d\langle \pi_i, \pi_j \rangle(s)$$

using the notation

$$D_i G = \frac{\partial G}{\partial x_i}, \quad D_{ij}^2 = \frac{\partial^2 G}{\partial x_i \partial x_j}.$$

If the function G is concave, then the process $\Gamma^G(t)$ is non-decreasing.

Definition 6.1.3 (*Lyapunov Function*). A regular function G is a Lyapunov function if the process $\Gamma^G(\cdot)$ in Definition 6.1.2 is nondecreasing.

Whereas Fernholtz, Karatzas, and Ruf [2018] developed both the notion of the additively generated portfolios and that of the multiplicatively generated portfolios, in this paper we restrict ourselves to the discussion of only the additively generated portfolios. We follow Definition 4.1 in Karatzas and Ruf [2017] [38]

Definition 6.1.4 (*Additively generated portfolios*). For any given regular function $G : \Delta^d \rightarrow \mathbb{R}$ w.r.t. the vector process $(\pi(t))_{t \geq 0}$ of market weights, we consider d -dimensional vector $(\phi^G(t))_{t \geq 0} = (\phi_1^G(t), \dots, \phi_d^G(t))_{t \geq 0}$ with components

$$\phi_i^G(t) := D_i G(\pi(t)) + \Gamma^G(t) + G(\pi(t)) - \sum_{j=1}^d \pi_j(t) D_j G(\pi(t)), \quad \text{for } i = 1, \dots, d.$$

This process is said to be a portfolio *additively generated by G* .

By Proposition 4.3 of Karatzas and Ruf [2017] [38], the additively generated portfolio $(\pi(t))_{t \geq 0}$ from Definition 6.1.2 is a trading strategy in the sense of Definition 6.1.1 and has the below *relative value process*:

$$V^\phi(t) = G(\pi(t)) + \Gamma^G(t).$$

6.2 Functionally generated arbitrage

For previous works on functionally generated portfolios (see Karatzas and Ruf [2017] [38] and Fernholz, Karatzas, and Ruf [2018] [24]), arbitrage and trading strategies of functionally generated portfolios are studied with

respect to the benchmark market portfolio. In this section, we consider a generalization of the theory by considering arbitrage of a functionally generated portfolio with respect to another functionally generated portfolio.

6.2.1 Relative arbitrage

We begin by defining relative arbitrage of a functionally generated portfolio (or its respective trading strategy) with respect to another portfolio. For this, we adapt and extend Definition 4.1 in Fernholtz et al. [2018] [24].

Definition 6.2.1 (*Relative Arbitrage*). Let us fix a real number $T > 0$ and a frictionless equity market S with d stocks. Let the process π represent the relative market weight process. Given trading strategies $\psi, \phi \in \mathcal{L}(S)$, we define their respective trading portfolios $\mu := (\mu_1, \dots, \mu_d)$ and $\nu := (\nu_1, \dots, \nu_d)$ by

$$\mu_i(t) := \frac{S_i(t)\psi_i(t)}{V^\psi(t; S)}$$

$$\nu_i(t) := \frac{S_i(t)\phi_i(t)}{V^\phi(t; S)}$$

for $i = 1, 2, \dots, d$ (see Section 6.1.3). Then portfolio μ is an arbitrage relative to portfolio ν over the time horizon $[0, T]$ if we have

$$V^\psi(t; \pi), V^\phi(t; \pi) \geq 0, \quad \forall t \in [0, T]; \quad V^\psi(0; \pi) = V^\phi(0; \pi) = 1$$

along with

$$\mathbb{P}[V^\psi(T; \pi) - V^\phi(T; \pi) \geq 0] = 1$$

and

$$\mathbb{P}[V^\psi(T; \pi) - V^\phi(T; \pi) > 0] > 0.$$

Moreover, whenever a given portfolio μ satisfies these conditions, and if the last probability is not just positive but actually equal to 1, that is, if

$$\mathbb{P}[V^\psi(T; \pi) - V^\phi(T; \pi) > 0] = 1,$$

we say that the portfolio μ generated by ψ is a strong arbitrage relative to the portfolio ν generated by ϕ over the time horizon $[0, T]$. Equivalently, in terms of trading strategies, we say that trading strategy ψ is a strong arbitrage relative to the trading strategy ϕ over the time horizon $[0, T]$.

6.2.2 Relative arbitrage between additively generated portfolios

In this section, we examine relative arbitrage between two additively generated portfolios. This is a generalization of the result in Theorem 4.3 by Karatzas and Ruf [2017] [38]. We leave the analysis of multiplicatively generated portfolio to future work.

Proposition 6.2.1. *Let continuous functions $G_1, G_2 : \Delta^d \rightarrow [0, \infty)$ be Lyapunov functions with $G_1(\pi(0)), G_2(\pi(0)) > 0$. Further impose G_1, G_2 to be non-negative to ensure non-negative wealth processes generated by the trading strategies derived from G_1, G_2 . Assume that $G_2(\pi(t))$ is bounded above by some constant C for all $t \geq 0$. Let Γ^{G_k} denote the Gamma process generated by functions G_k for $k = 1, 2$. If (continuous, adapted) process*

$$\kappa(\cdot) := \Gamma^{G_1}(\cdot) - \Gamma^{G_2}(\cdot)$$

is such that

$$\kappa(t) > \eta t$$

for some constant $\eta > 0$, then there exists a time horizon over which the trading strategy $\psi = (\psi_1, \dots, \psi_d)$ additively generated by G_1 is a strong arbitrage relative to the trading strategy $\phi = (\phi_1, \dots, \phi_d)$ additively generated by G_2 .

Proof. The following proof simplifies the notation for relative value process notation $V(\cdot; \pi)$ to $V(\cdot)$. Let trading strategies ψ, ϕ be additively generated with respect to functions G_1, G_2 . By Proposition 4.3 in “Lyapunov functions”,

$$V^\psi(t) = G_1(\pi(t)) + \Gamma^{G_1}(t)$$

$$V^\phi(t) = G_2(\pi(t)) + \Gamma^{G_2}(t)$$

for $t \geq 0$. It follows that given a fixed $T > 0$,

$$\begin{aligned} V^\psi(T) - V^\phi(T) &= (\Gamma^{G_1}(T) - \Gamma^{G_2}(T)) + (G_1(\pi(T)) - G_2(\pi(T))) \\ &\geq (\Gamma^{G_1}(T) - \Gamma^{G_2}(T)) - G_2(\pi(T)) \end{aligned} \quad (3)$$

Here we have used the bound $G_1 \geq 0$. From this, we can derive a stronger condition for strong relative arbitrage. Specifically, if

$$\mathbb{P}[\Gamma^{G_1}(T) - \Gamma^{G_2}(T) > G_2(\pi(T))] = 1$$

then

$$\mathbb{P}[V^\psi(T) - V^\phi(T) > 0] = 1$$

holds almost surely. Recall that $\kappa(t) = \Gamma^{G_1}(t) - \Gamma^{G_2}(t) > \eta t$ for some $\eta > 0$ and $\forall t \geq 0$. Choose $T^* > 0$ to be such that

$$T^* > \frac{C}{\eta},$$

it follows that

$$\Gamma^{G_1}(T^*) - \Gamma^{G_2}(T^*) > \eta T^* > C \geq G_2(\pi(T^*)).$$

Then for all $T \geq T^*$, we have that

$$\Gamma^{G_1}(T) - \Gamma^{G_2}(T) > \eta T > \eta T^* > C \geq G_2(\pi(T)).$$

Therefore, the additively generated strategy $\psi = (\psi_1, \dots, \psi_d)^T$ is a strong arbitrage relative to the additively generated strategy $\phi = (\phi_1, \dots, \phi_d)^T$ over every time horizon $[0, T]$ with $T \geq T^*$. \square

Remark 6.2.1. Note that by the definition of regular functions, the trading strategy additively generated by a regular function G need not be unique. This is because DG is not necessarily unique. This may result in different wealth process additively generated by the same regular function G . We note that our results in Proposition 1.3 hold in general, without having to assume the uniqueness of additively generated trading strategies nor wealth process. In addition, to impose further that the wealth process additively generated by the same regular function be unique, we invoke Proposition 3.4 from [38] and observe that if there exists a local martingale deflator for the market weight process $\pi(\cdot)$, then the process $\Gamma^G(\cdot)$ does not depend on the choice of DG and is uniquely determined up to indistinguishability.

6.3 Convex order and the gamma process

In this section, we present a conjecture on the link between convex order and the gamma process $\Gamma(\cdot)$ in the theory of Functionally Generated Portfolio.

Conjecture 6.3.1. *Consider a generic d -dimensional semimartingale of market weights $\pi(\cdot)$ with continuous paths. Let $G_1, G_2 : \Delta^d \rightarrow [0, \infty)$ be a continuous regular function in C^2 . For all $t \geq 0$ and for $k = 1, 2$, recall that the (continuous, adapted) gamma process functionally generated by G_1 and G_2 is defined as*

$$\Gamma^{G_k}(t) := G_k(X(0)) - G_k(X(t)) + \int_0^t \sum_{i=1}^d v_{k_i}(t) dX_i(t)$$

for $k = 1, 2$, where

$$v_{k_i}(t) := D_i G_k(X(t)), \quad i = 1, \dots, d, t \geq 0$$

is integrable with respect to the Δ^d -valued semimartingale $\pi(\cdot)$.

Now let ψ, ϕ be trading strategies **additively** generated by functions G_1, G_2 . Fix a $t \geq 0$. If (continuous, adapted) process

$$\kappa(\cdot) := \Gamma^{G_1}(\cdot) - \Gamma^{G_2}(\cdot)$$

satisfies some form of concavity*, then the probability measures μ, ν defined by the investment proportion processes, i.e.

$$\mu_i(\cdot) := \frac{S_i(\cdot)\psi_i(\cdot)}{V^\psi(\cdot; S)}$$

$$\nu_i(\cdot) := \frac{S_i(\cdot)\phi_i(\cdot)}{V^\phi(\cdot; S)}$$

for $i = 1, 2, \dots, d$ have finite second moments and are in a weaker form of convex order, which we call “quasi-convex order” or the “T-order”.

Below we give a motivation for Conjecture 6.3.1.

Let ψ, ϕ be trading strategies additively generated by G_1, G_2 . Then by Proposition 2.2 and Proposition 4.3 in [38],

$$\begin{aligned}\mu_i(t) &:= \frac{S_i(t)\psi_i(t)}{V^\psi(t; S)} = \frac{S_i(t)\psi_i(t)}{\Sigma(t)V^\psi(t; \pi)} = \frac{S_i(t)\psi_i(t)}{\Sigma(t)(G_1(\pi(t)) + \Gamma^{G_1}(t))} \\ \nu_i(t) &:= \frac{S_i(t)\phi_i(t)}{V^\phi(t; S)} = \frac{S_i(t)\phi_i(t)}{\Sigma(t)V^\phi(t; \pi)} = \frac{S_i(t)\phi_i(t)}{\Sigma(t)(G_2(\pi(t)) + \Gamma^{G_2}(t))}\end{aligned}$$

Again, by Proposition 4.3 of Karatzas and Ruf [2017] [38] we can write

$$\begin{aligned}\psi_i(t) &= V^\psi(t; \pi) + D_i G_1(\pi(t)) - \sum_{j=1}^d \pi_j(t) D_j G_1(\pi(t)) \\ &= G_1(\pi(t)) + \Gamma^{G_1}(t) + D_i G_1(\pi(t)) - \sum_{j=1}^d \pi_j(t) D_j G_1(\pi(t))\end{aligned}\tag{4}$$

and

$$\begin{aligned}\phi_i(t) &= V^\phi(t; \pi) + D_i G_2(\pi(t)) - \sum_{j=1}^d \pi_j(t) D_j G_2(\pi(t)) \\ &= G_2(\pi(t)) + \Gamma^{G_2}(t) + D_i G_2(\pi(t)) - \sum_{j=1}^d \pi_j(t) D_j G_2(\pi(t))\end{aligned}\tag{5}$$

Then we write

$$\begin{aligned}\mu_i(t) &= \pi_i(t) \frac{V^\psi(t; \pi) + D_i G_1(\pi(t)) - \sum_{j=1}^d \pi_j(t) D_j G_1(\pi(t))}{V^\psi(t; \pi)} \\ &= \pi_i(t) \left(1 + \frac{D_i G_1(\pi(t)) - \sum_{j=1}^d \pi_j(t) D_j G_1(\pi(t))}{G_1(\pi(t)) + \Gamma^{G_1}(t)} \right)\end{aligned}\tag{6}$$

and so similarly,

$$\nu_i(t) = \pi_i(t) \left(1 + \frac{D_i G_2(\pi(t)) - \sum_{j=1}^d \pi_j(t) D_j G_2(\pi(t))}{G_2(\pi(t)) + \Gamma^{G_2}(t)} \right)\tag{7}$$

since

$$\frac{S_i(t)}{\Sigma(t)} = \pi_i(t).$$

In the theorem, we suppose process $\kappa(\cdot) = \Gamma^{G_1}(\cdot) - \Gamma^{G_2}(\cdot)$ to follow certain form of concavity. It is not yet clear what this concavity may be. Some ideas for the concavity include that it be defined as a logarithm transform of a random process X that is stochastically concave, i.e.

$$\mathbb{E}[wX(t_1) + (1-w)X(t_2)] \leq w\mathbb{E}[X(t_1)] + (1-w)\mathbb{E}[X(t_2)]$$

for all $t_1, t_2 \geq 0$ and $w \in [0, 1]$. Intuitively, we want a form of bounded concavity in $\kappa(\cdot)$ because we want to existence of a dominating constant $D \in \mathbb{R}$ or another bounded function $y(t)$ such that for all values in $\text{supp}(\kappa(t))$, $\kappa(t) \leq D$ or $\kappa(t) \leq y(t)$. Since the Γ^G -process is interpreted

as the quadratic variation term of its generating function G , it is a measure of spread. By saying that the difference process $(\kappa_t)_{t \geq 0}$ between two Gamma processes is bounded above by some constant (or some bounded function) means that the spread of the first portfolio generated by G_1 is no bigger than the second one generated by G_2 to a certain extent. This is weaker than the strict definition of convex order, since for Γ -order, it allows a certain amount of “leeway” up to the dominating constant D (or dominating bounded function $\gamma(t)$). In this sense, we call the form of “relaxed” convex order the Γ -order a quasi convex order.

7 Conclusion

We have introduced how an arbitrage strategy can be determined through leveraging the absence of convex order in the market. We developed model-independent arbitrage strategy in trading with European options via optimal transport and also offered an overview of relative arbitrage between Functionally Generated Portfolios in Stochastic Portfolio Theory as well as how this could be related to a form of convex order in the market. Future works include formally proving or disproving Conjecture 6.3.1, back-testing model-independent trading strategies in Section 5 with real-world data, as well as developing more efficient numerical methods for the calculation of optimal ρ as well as recovering convex function \hat{f} in multi-dimension.

8 Appendix

8.1 A. Derivation of the TPE Acquisition Function

We derive here the EI-based acquisition function for TPE. That is,

$$EI_y^* \propto (\gamma + \frac{g(x)}{l(x)}(1 - \gamma))^{-1}.$$

Proof. Recall that the TPE algorithm parametrizes $p(x, y)$ as $p(y)p(x|y)$. Using Bayes’ rule and by definition of the EI criterion, we have that

$$EI_y^*(x) = \int_{-\infty}^{y^*} (y^* - y) \frac{p(x|y)p(y)}{p(x)} dy.$$

Let γ be the cut-off quantile. Recall that $p(y < y^*) = \gamma$ and that TPE defines $p(x|y)$ using two densities over the configuration space \mathcal{X} :

$$p(x|y) = \begin{cases} l(x), & \text{if } y < y^* \\ g(x), & \text{otherwise} \end{cases}$$

Then we can write

$$p(x) = \int_{\mathbb{R}} p(x|y)p(y)dy = \gamma y^* l(x) + (1 - \gamma)g(x).$$

Therefore,

$$\int_{-\infty}^{y^*} (y^* - y)p(x|y)p(y)dy = l(x) \int_{-\infty}^{y^*} (y^* - y)p(y)dy = \gamma y^* l(x) - l(x) \int_{-\infty}^{y^*} p(y)dy.$$

It follows that

$$\begin{aligned} EI_y^* &= \int_{-\infty}^{y^*} (y^* - y) \frac{p(x|y)p(y)}{p(x)} dy \\ &= \int_{-\infty}^{y^*} (y^* - y) \frac{p(x|y)p(y)}{\gamma y^* l(x) + (1 - \gamma)g(x)} dy \\ &= \frac{\gamma y^* l(x) - l(x) \int_{-\infty}^{y^*} p(y)dy}{\gamma l(x) + (1 - \gamma)g(x)} \\ &\propto \left(\gamma + \frac{g(x)}{l(x)}(1 - \gamma)\right)^{-1}. \end{aligned}$$

□

8.2 B. Derivation of the Poisson Equation with Neumann Boundary Conditions

Assume that the domain $\Omega \subset \mathbb{R}^n$ is bounded and connected. Assume that $\hat{f} : \mathbb{R}^2 \rightarrow \mathbb{R}$ is sufficiently smooth. Given an empirically observed gradient field g in 2-dimensional, we consider the general case where the values of \hat{f} on the boundary are not known. Let $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ be an arbitrary continuously differentiable function representing the random noise. We provide a proof for the equivalence of inverse gradient calculation in 2D with the Poisson equation with Neumann boundary conditions.

Proof. Consider

$$\begin{aligned} \|\nabla(\hat{f} + \phi) - g\|^2 &= \|\nabla\hat{f} + \nabla\phi(\mathbf{x}) - g(\mathbf{x})\|^2 \\ &= \int_{\Omega} |\nabla\hat{f} - g(\mathbf{x})|^2 d\mathbf{x} + 2 \int_{\Omega} (\nabla\hat{f}(\mathbf{x}) - g(\mathbf{x})) \cdot \nabla\phi(\mathbf{x}) d\mathbf{x} \\ &\quad + \int_{\Omega} |\nabla\phi(\mathbf{x})|^2 d\mathbf{x}. \end{aligned}$$

Since $\arg \min_{\hat{f}} \|\nabla\hat{f} - g\|$ must also minimize $\arg \min_{\hat{f}} \|\nabla(\hat{f} + \phi) - g\|$ (consider taking a constant function $\phi(\mathbf{x}) = 0$), this means that the following condition must be satisfied for the optimal \hat{f} :

$$\int_{\Omega} (\nabla\hat{f}(\mathbf{x}) - g(\mathbf{x})) \cdot \nabla\phi(\mathbf{x}) d\mathbf{x} = 0$$

Decomposing a line integral from the LHS, we have

$$\begin{aligned} 0 &= \int_{\Omega} (\nabla\hat{f}(\mathbf{x}) - g(\mathbf{x})) \cdot \nabla\phi(\mathbf{x}) d\mathbf{x} \\ &= \int_{\partial\Omega} \left(\frac{\partial\hat{f}}{\partial n}(\mathbf{x}) - \mathbf{n} \cdot g(\mathbf{x})\right) \phi(\mathbf{x}) dS \\ &\quad - \int_{\Omega} (\nabla^2 \hat{f}(\mathbf{x}) - \mathbf{n} \cdot \nabla g(\mathbf{x})) \phi(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

Since we want the above equality to hold for every function $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$, we recovered the Neumann boundary condition:

$$\begin{cases} \nabla^2 \hat{f}(\mathbf{x}) = \nabla \cdot g(\mathbf{x}), & \forall \mathbf{x} \in \Omega \\ \nabla \hat{f}(\mathbf{x}) \cdot \mathbf{n} = \mathbf{n} \cdot g(\mathbf{x}), & \forall \mathbf{x} \in \partial\Omega, \end{cases}$$

□

References

- [1] Johannes Wiesel and Erica Zhang. *An optimal transport based characterization of convex order*. 2023. arXiv: [2207.01235](https://arxiv.org/abs/2207.01235) [[math.PR](#)].
- [2] Volker Strassen. “The Existence of Probability Measures with Given Marginals”. In: *The Annals of Mathematical Statistics* 36.2 (1965), pp. 423–439. URL: <http://www.jstor.org/stable/2238148>.
- [3] Douglas T Breeden and Robert H Litzenberger. “Prices of State-Contingent Claims Implicit in Option Prices”. In: *The Journal of Business* 51.4 (1978), pp. 621–651. URL: <https://faculty.baruch.cuny.edu/lwu/890/BreedenLitzenberger78.pdf>.
- [4] Julien Guyon. “Inversion of Convex Ordering in the VIX Market”. In: *Quantitative Finance* (2019). Forthcoming. URL: <https://ssrn.com/abstract=3504022%20or%20http://dx.doi.org/10.2139/ssrn.3504022>.
- [5] Peter Carr, Roger Lee and Zura Kakushadze. “Options on Realized Variance and Convex Orders”. In: *International Journal of Theoretical and Applied Finance* 21.2 (2018), p. 1850018. URL: <https://engineering.nyu.edu/sites/default/files/2019-03/Carr-options-on-realized-variance-and-convex-orders.pdf>.
- [6] Andre H. Tchen. “Inequalities for Distributions with Given Marginals”. In: *The Annals of Probability* 8.4 (1980), pp. 814–827. DOI: [10.1214/aop/1176994668](https://doi.org/10.1214/aop/1176994668). URL: <https://doi.org/10.1214/aop/1176994668>.
- [7] Ken L Bechmann and Rolf R Olesen. “Variance Swap Valuation with Time-Varying Volatility and Interest Rate”. In: *Journal of Risk and Insurance* 83.3 (2016), pp. 633–658. DOI: [10.1111/jori.12108](https://doi.org/10.1111/jori.12108). URL: <https://doi.org/10.1111/jori.12108>.
- [8] Yann Brenier. “Polar factorization and monotone rearrangement of vector-valued functions”. In: *Communications on Pure and Applied Mathematics* 44.4 (1991), pp. 375–417. DOI: [10.1002/cpa.3160440402](https://doi.org/10.1002/cpa.3160440402). URL: <https://doi.org/10.1002/cpa.3160440402>.
- [9] Gaspard Monge. *Mémoire sur la théorie des déblais et des remblais*. De l’Imprimerie Royale, 1781.
- [10] Leonid Kantorovich. “On the translocation of masses”. In: *Management Science* 5.1 (1958), pp. 1–4. DOI: [10.1287/mnsc.5.1.1](https://doi.org/10.1287/mnsc.5.1.1). URL: <https://doi.org/10.1287/mnsc.5.1.1>.

- [11] Svetlozar Rachev and Ludger Rüschendorf. *Mass Transportation Problems: Volume I: Theory, volume 1*. Springer Science & Business Media, 1998. DOI: [10.1007/978-3-662-03612-2](https://doi.org/10.1007/978-3-662-03612-2). URL: <https://doi.org/10.1007/978-3-662-03612-2>.
- [12] Cédric Villani. *Optimal Transport: Old and New*. Vol. 338. Springer Berlin, 2008.
- [13] Guillaume Carlier. “Remarks on Toland’s duality, convexity constraint and optimal transport”. In: *Pacific Journal of Optimization* 4.3 (2008), pp. 423–432.
- [14] Mark Rubinstein. “Implied Binomial Trees”. In: *Journal of Finance* 49.3 (1994), pp. 771–818.
- [15] Bernard Dumas, Jeff Fleming and Robert E. Whaley. “Implied Volatility Functions: Empirical Tests”. In: *Journal of Finance* 53.6 (1998), pp. 2059–2106.
- [16] James B. Wiggins. “Option Values Under Stochastic Volatility: Theory”. In: *Journal of Finance* 42.2 (1987), pp. 281–300.
- [17] David Hobson. “Robust Hedging of the Lookback Option”. In: *Finance & Stochastics* 2.3 (1998), pp. 329–347.
- [18] Laurent Cousot. *Necessary and Sufficient Conditions for No Static Arbitrage among European Calls*. Tech. rep. Courant Institute, New York University, 2004.
- [19] Hans Buehler. “Expensive Martingales”. In: *Quantitative Finance* 6.3 (2006), pp. 207–218.
- [20] Peter Carr and Dilip B. Madan. “A Note on Sufficient Conditions for No Arbitrage”. In: *Finance Research Letters* 2.3 (2005), pp. 125–130.
- [21] Frank Riedel. “The Fundamental Theorem of Asset Pricing Without Probabilistic Prior Assumptions”. In: *SSRN Electronic Journal* (2011). DOI: [10.2139/ssrn.1892706](https://ssrn.com/abstract=1892706). URL: <https://ssrn.com/abstract=1892706>.
- [22] Alexander M. G. Cox and Jan Oblój. “Robust Pricing and Hedging of Double No-Touch Options”. In: *Finance and Stochastics* 15.3 (2011), pp. 573–605.
- [23] Marcel Nutz. “Superreplication Under Model Uncertainty in Discrete Time”. In: *Preprint* (2013). eprint: [arXiv:1301.3227](https://arxiv.org/abs/1301.3227).
- [24] E. Robert Fernholz, Ioannis Karatzas and Johannes Ruf. *Volatility and Arbitrage*. 2016. arXiv: [1608.06121](https://arxiv.org/abs/1608.06121) [q-fin.PM].
- [25] Patryk Gierjadowicz et al. “Robust Pricing and Hedging via Neural SDEs”. In: *SSRN Electronic Journal* (Jan. 2020). DOI: [10.2139/ssrn.3646241](https://ssrn.com/abstract=3646241).
- [26] Mark H. A. Davis and David G. Hobson. “The Range of Traded Option Prices”. In: *Mathematical Finance* 17.1 (2007), pp. 1–14. DOI: [10.1111/j.1467-9965.2007.00290.x](https://doi.org/10.1111/j.1467-9965.2007.00290.x).

- [27] Beatrice Acciaio et al. “A Model-Free Version of the Fundamental Theorem of Asset Pricing and the Super-Replication Theorem”. In: *Mathematical Finance* 28.2 (2018), pp. 125–134. DOI: [10.1111/mafi.12060](https://doi.org/10.1111/mafi.12060).
- [28] James Bergstra, Daniel Yamins and David D. Cox. “Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures”. In: *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)*. 2013, pp. 115–123.
- [29] Bobak Shahriari et al. “Taking the Human Out of the Loop: A Review of Bayesian Optimization”. In: *Proceedings of the IEEE* 104.1 (2016), pp. 148–175.
- [30] Jurgita Kapociūtė-Dzikiene, Kaspars Balodis and Raivis Skadiņš. “Intent Detection Problem Solving via Automatic DNN Hyperparameter Optimization”. In: *Applied Sciences* 10.12 (2020), p. 4113.
- [31] James Bergstra et al. “Algorithms for hyper-parameter optimization”. In: *Proceedings of the 24th Annual Conference on Neural Information Processing Systems (NIPS 2011)*. Granada, Spain, 2011, pp. 2546–2554.
- [32] Nabarun Deb, Promit Ghosal and Bodhisattva Sen. “Rates of estimation of optimal transport maps using plug-in estimators via barycentric projections”. In: *Advances in Neural Information Processing Systems*. 2021, pp. 29736–29753.
- [33] Gunnar Färneback et al. “Efficient Computation of the Inverse Gradient on Irregular Domains”. In: *2007 IEEE 11th International Conference on Computer Vision*. 2007, pp. 1–8. DOI: [10.1109/ICCV.2007.4409176](https://doi.org/10.1109/ICCV.2007.4409176).
- [34] S. M.-H. Song et al. “Phase unwrapping of MR phase images using Poisson equation”. In: *IEEE Transactions on Image Processing* 4.5 (May 1995), pp. 667–676. DOI: [10.1109/83.370388](https://doi.org/10.1109/83.370388).
- [35] Lawrence C. Evans. *Partial Differential Equations*. 2nd ed. Chapter 5, Theorem 3 (p. 376). American Mathematical Society, 2010.
- [36] Julien Guyon, Romain Menegaux and Marcel Nutz. “Bounds for VIX futures given S&P 500 smiles”. In: *Finance and Stochastics* 21 (2017), pp. 593–630.
- [37] Ioannis Karatzas and Constantinos Kardaras. *Portfolio Theory and Arbitrage: A Course in Mathematical Finance*. New York: Springer, 2012.
- [38] Ioannis Karatzas and Johannes Ruf. *Trading Strategies Generated by Lyapunov Functions*. 2016. arXiv: [1603.08245](https://arxiv.org/abs/1603.08245) [[q-fin.MF](https://arxiv.org/archive/finance)].