Martingale optimal transport: towards robust model-independent bounds on option prices

Vivien Tisserand

May 6, 2024

Abstract

We review the advances and financial applications of martingale optimal transport. After introducing the optimization problem, we discuss three ways for solving equivalent formulations before implementing it to derive model-free bounds on the price of forward start options on the SPX.
An implementation in `jax` is available at this repository.

Contents

1 A primer on martingale optimal transport

Martingale optimal transport was first introduced internally at Société Générale in 2009 by Pierre Henry-Labordère. [Hen17] extensively covers this subject.

It is crucial to design hedging strategies when selling derivatives (whether it is vanilla options or exotic products) – this is corporate banks businesses. This involves finding replication prices (or even super-replication prices) that can then be referred to as the "fair price" of the derivative.

Below, we highlight the main differences that stem from martingale optimal transport as opposed to classic optimal transport.

1.1 Model setup, motivation

The main problem we are facing when valuing an exotic options is to choose a "correct" pricing model: indeed it should be consistent with the exotic market prices as well as with the liquid hedging instruments – that is vanilla options.
The martingale optimal transport setup allows us to derive model-free bounds for the prices of such options.

The Breeden-Litzenberger formula gives us the market implied probability density: let's fix a maturity T and assume we have access to a strip of (call) option prices at different strikes, the probability density of the asset price at time T is given by

$$f(K) = \partial_{K^2} C(K).$$

For financial assets, this is how we will build the **marginals**. Prescribing the one-dimensional marginal distributions is equivalent to assuming that our model is calibrated to vanilla options with payoffs $(S_T - K)^+$.

We focus on a two-period model: with one asset S at two times t_1 and t_2 and a payoff (or cost to stay in the optimal transport lexical field) $c(x, y)$ at time T, we are interested in the problem

$$\sup_{\mathbb{Q} \in \mathcal{M}(\mathbb{P}^1, \mathbb{P}^2)} \mathbb{E}^{\mathbb{Q}} \left[c(S_{t_1}, S_{t_2}) \right], \tag{P}$$

where $\mathbb{P}^i$'s are the densities derived from vanilla options on the assets S_{t_i}'s, and $\mathcal{M}(\mathbb{P}^1, \mathbb{P}^2) \subseteq \mathcal{P}(\mathbb{P}^1, \mathbb{P}^2)$ is the **martingale transport plan** set:

$$\mathcal{M}(\mathbb{P}^1, \mathbb{P}^2) \coloneqq \{\mathbb{Q} \in \mathcal{P}(\mathbb{P}^1, \mathbb{P}^2) : \mathbb{E}^{\mathbb{P}} \left[X_{t_2} \mid X_{t_1} \right] = X_{t_1} \}.$$

This extra condition on top of the classic one on the marginals restrains the setup to martingale optimal transport.

The problem P is the dualized formulation[1] of the Monge-Kantorovitch linear program (classicly solved with a simplex algorithm).
$\mathcal{M}(\mathbb{P}^1, \mathbb{P}^2)$ is a closed subset of the compact $\mathcal{P}(\mathbb{P}^1, \mathbb{P}^2)$, and Strassen proves it is non-empty iff $\mathbb{P}^1 \preceq_{\mathrm{cx}} \mathbb{P}^2$, *i.e.* the longer term density is noisier than the sorter one[2], thus there exists a solution to our problem.
P is interpreted as a robust upper bound of the fair value of the option with payoff c at time $T = t_2$.

With the Fenchel-Rockafellar duality theorem, we can transport the problem P:

$$\inf_{(u_1, u_2) \in \mathcal{M}^*(\mathbb{P}^1, \mathbb{P}^2), h \in \mathcal{B}} \{\mathbb{E}^{\mathbb{P}^1} \left[u_1(S_{t_1}) \right] + \mathbb{E}^{\mathbb{P}^2} \left[u_2(S_{t_2}) \right]\}, \tag{D}$$

[1] In the financial optimal transport literature, Monge's primal problem (resp. Kantorovitch's dual problem) is referred to as the dual (resp. primal), thus P is our primal problem.

[2] This also translates as: $(\mathbb{P}^1, \mathbb{P}^2)$ must be a *peacock process*. This *convex order* means in financial terms that we exclude calendar spread arbitrage opportunities

looking for solution on the subset where for every $(x, y) \in \mathbb{R}_+^2$:

$$\underbrace{u_1(x) + u_2(y)}_{\text{replication strategy payoff}} + \underbrace{h(x)(y - x)}_{\text{strategy PnL, } h \text{ a bounded function}} \geq \underbrace{c(x, y)}_{\text{payoff of the contingent claim}},$$

this corresponds to a semi-static hedging strategy.

Assuming the u_i's are twice differentiable, the Carr-Madan formula gives us a replication strategy in terms of calls and puts at times t_1 and t_2. With this dual formulation we minimize a linear cost function with respect to the vanilla calls and puts.

The primal problem gives a model-independent upper bound for arbitrage-free prices while the dual problem solves for the lowest super-replication price. Under duality, we have a model-independent super-replication price (under some regularity conditions on the payoff c).
Overall we are looking for a measure $\mathbb{Q}^* \in \mathcal{M}(\mathbb{P}^1, \mathbb{P}^2)$ which achieves the upper bounds.

In general, this linear programming problem is not solved by traditional simplex or primal-dual interior point methods [Hen13; HT13; GO19],

1.2 Solving the primal, or the dual?

The primal, a bit messy? On the one hand there are some claims that the MOT primal P involves discretization for the asset prices at both times and mostly a smart quantization of both measures (naive methods often break the Strassen therorem and leave the martingale transport plan empty [Bak12]) before having a linear program solvable using the simplex algorithm. On the other hand, when it comes to numerical methods to solve such a problem, [ACJ19] uses the linear programming solver GLPK (when dealing with problems where the MOT is not known explicitly) to solve a discretized version of the primal:

$$\sup_{r_{ij} \geq 0, \sum_i r_{ij} = q_j, \sum_j r_{ij} = p_i, \sum_j r_{ij} y_j = p_i x_i} \sum_i \sum_j r_{ij} c(x_i, y_j),$$

A dual with financial insights? This motivates a further study of the MOT dual D. We can approximate the quantity to minimize by a position in cash, the underlying asset and available call options at time t_1 and t_2:

$$\mathbb{E}^{\mathbb{P}^1}\left[u_1(S_{t_1})\right] + \mathbb{E}^{\mathbb{P}^2}\left[u_2(S_{t_2})\right] \approx \beta + \alpha S_{t_0} + \sum_{l=0}^{N} \omega_1^l C(t_1, K_1^l) + \sum_{k=0}^{M} \omega_2^k C(t_2, K_2^k),$$

with the super-replication constraint being also discretisable, we have a linear program that approximates the upper bound.

A dual with convex transformation? Some other ways to solve the dual problem involve neural networks [EK21a; EK21b]. In particular, there exists a formulation of the dual problem in the martingale setup involving the convex biconjugates and alleviates the problems constraints [BHP13; Ses24]. In the classical OT, such a reformulation with c-transforms was introduced by Cédric Villani [Vil+09, Theorem 5.10].

The *Legendre-Fenchel conjugate* for a function $f\colon \mathcal{X} \mapsto \bar{\mathbb{R}}$ is written:

$$f^*(y) := \sup_{x\in\mathcal{X}} \langle y, x\rangle - f(x),$$

thus the *biconjugate* is $f^{**}(y) := \sup_{x\in\mathcal{X}} \langle y, x\rangle - f^*(y)$. This defines the *convex envelope* of the function f.

Proposition 1.0.1. *Assuming the payoff function fulfills the linear growth condition $c(x,y) \geq -K(1+|x|+|y|)$ for some $K \in \mathbb{R}$ and that there exists a market measure $\mathbb{Q} \in \mathcal{M}(\mathbb{P}^1,\mathbb{P}^2)$ such that $\mathbb{E}^{\mathbb{Q}}[c] < +\infty$, we have an equivalent formulation of P:*

$$\sup_{\mathbb{Q}\in\mathcal{M}(\mathbb{P}^1,\mathbb{P}^2)} \mathbb{E}^{\mathbb{Q}}\left[c(S_{t_1},S_{t_2})\right] = \inf_{u\in L^1(\mathbb{P}^2)} \mathbb{E}^{\mathbb{P}^1}\left[(c(S_{t_1},\cdot)-u(\cdot))^{**}(S_{t_1})\right] + \mathbb{E}^{\mathbb{P}^2}\left[u(S_{t_2})\right]. \tag{Biconjugate dual}$$

Proof. Following [BHP13, Proposition 4.4], let $\mathbb{Q} \in \mathcal{M}(\mathbb{P}^1,\mathbb{P}^2), u \in L^1(\mathbb{P}^2)$, then:

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left[c(S_{t_1},S_{t_2})\right] &= \mathbb{E}^{\mathbb{Q}}\left[c(S_{t_1},S_{t_2}) - u(S_{t_2})\right] + \mathbb{E}^{\mathbb{P}^2}\left[u(S_{t_2})\right] \\
&\geq \mathbb{E}^{\mathbb{Q}}\left[(c(S_{t_1},\cdot)-u(\cdot))^{**}(S_{t_2})\right] + \mathbb{E}^{\mathbb{P}^2}\left[u(S_{t_2})\right] \quad \text{using the convex envelope} \\
&= \mathbb{E}^{\mathbb{P}^1}\left[\mathbb{E}^{\mathbb{Q}}\left[(c(S_{t_1},\cdot)-u(\cdot))^{**}(S_{t_2})\right] \mid S_{t_1}\right] + \mathbb{E}^{\mathbb{P}^2}\left[u(S_{t_2})\right] \\
&\geq \mathbb{E}^{\mathbb{P}^1}\left[(c(S_{t_1},\cdot)-u(\cdot))^{**}(\mathbb{E}^{\mathbb{Q}}\left[S_{t_2} \mid S_{t_1}\right])\right] + \mathbb{E}^{\mathbb{P}^2}\left[u(S_{t_2})\right] \quad \text{by Jensen} \\
&= \mathbb{E}^{\mathbb{P}^1}\left[(c(S_{t_1},\cdot)-u(\cdot))^{**}(S_{t_1})\right] + \mathbb{E}^{\mathbb{P}^2}\left[u(S_{t_2})\right] \quad \text{by martingality,}
\end{aligned}$$

hence the first inequality.

For the reverse inequality let's first assume that for $x_1 \in \mathbb{R}$, $u_1 \in L^1(\mathbb{P}^1)$ and $\Delta \in \mathcal{C}_b(\mathbb{R})$ such that

$$u_1(x_1) + \Delta(x_2 - x_1) \leq c(x_1,x_2)$$

for all $(x_1,x_2) \in \mathbb{R}^2$, then $u_1(x_1) \leq (c(x_1,\cdot) - u_1(\cdot))^{**}(x_1)$

Using this additional hypothesis,

$$\begin{aligned}
&\inf_{u\in L^1(\mathbb{P}^2)} \mathbb{E}^{\mathbb{P}^1}\left[(c(S_{t_1},\cdot)-u(\cdot))^{**}(S_{t_1})\right] + \mathbb{E}^{\mathbb{P}^2}\left[u(S_{t_2})\right] \\
&\qquad \leq \inf_{u\in L^1(\mathbb{P}^2)} \inf_{u_1 : u_1(x_1)+\Delta(x_2-x_1)\leq c(x_1,x_2)} \mathbb{E}^{\mathbb{P}^1}\left[(u_1(S_{t_1})\right] + \mathbb{E}^{\mathbb{P}^2}\left[u(S_{t_2})\right] \\
&\qquad = \inf_{(u_1,u_2)\in\mathcal{M}^*(\mathbb{P}^1,\mathbb{P}^2), h\in\mathcal{B}} \{\mathbb{E}^{\mathbb{P}^1}\left[u_1(S_{t_1})\right] + \mathbb{E}^{\mathbb{P}^2}\left[u_2(S_{t_2})\right]\} \\
&\qquad = D = P \quad \text{by duality.}
\end{aligned}$$

Hence the equivalence between the problems. This can be extended to d marginals swiftly [Ses24].

$\square$

1.3 Some variants: entropy regularization penalties

Without further constraints, the optimal martingale measure which achieves the super-replication strategy is often far from the measures produced by stochastic volatility models. Adding an entropic penalty (here we can impose the Kullback-Leibler divergence – or the Wasserstein distance – between $\mathbb{P}$ and a prior measure to be bounded).

One can also constrain the behaviour of the transport by adding VIX futures data: this information acts as a bound on the conditional dispersion of S_1/S_2, thus restricts the set of potential martingale transports [Guy24].

1.4 Financial applications: forward start options, among others

Among other exotic payoffs, we choose to focus on **forward start options**: given we are at time t and see two expiries in the future, t_1 and t_2, we wish to enter the option expiring at t_2 delievering the payoff $(S_{t_2} - KS_{t_1})^+$.

1.5 Other applications: Skokhorod embedding theorem

More academic-inclined problems can be solved with MOT, for instance the Skokhorod embedding theorem that aims at representing a probability measure μ as the law of a Brownian motion stopped at a well-chosen stopping-time τ [Cou+11].

There are also interesting links with other mathematical objects that model dependence: Fréchet-Hoeffding bounds and copulas among others.

2 Implementing MOT

Below we restrict ourselves to a setup with one asset and two marginals corresponding to the implied densities of this stock in the future. We wish to find model-free bounds for the forward start option.

2.1 Data processing

We retrieve options data on the SPX, an equity index on US stocks that became the prevailing benchmark due to its established presence in financial analysis. We look at options quoted on November 4$^{\text{th}}$ 2022 when the SPX closed at 3770.

The implied volatilities of the liquid options in the (strike, time to maturity) space reveal a classic equity vol surface, with smile and skew. In our setup we refer to the two measures as the shorter one and the longer one.

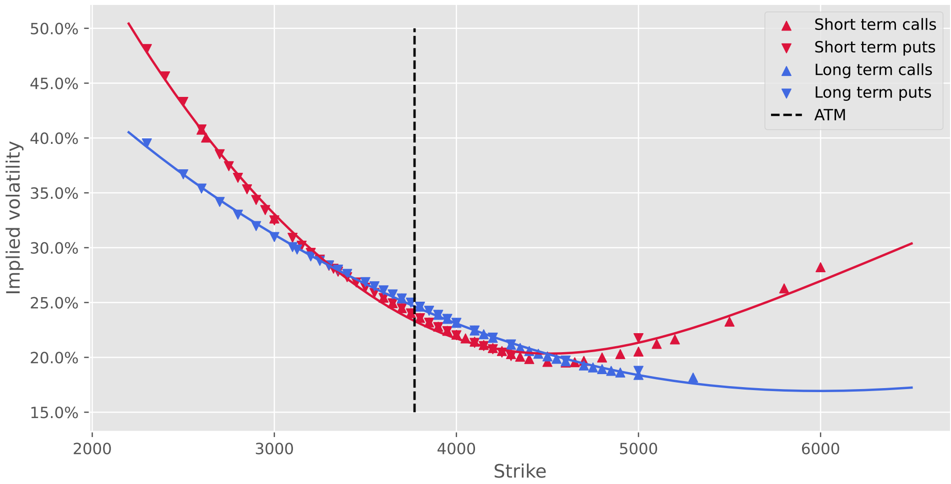

Figure 1: Implied volatility at liquid strikes for two maturities: option smiles with SVI interpolation.

As seen above, we will be deriving the call prices with regards to the strike twice in order to get the implied density of the underlying asset at two maturities (a shorter and a longer one). Using finite differences one raw prices, we obtain nor credible neither smooth densities (we observe spikes and even negative values).

Due to the high discretization in strikes we observe in the market-available instruments, we decide to first interpolate the volatility smiles before differentiating. Namely we will be using the stochastic volatility inspired (SVI) parametrization [GJ14]: for a given expiry we calibrate the implied variance with regard to the log-strike according to:

$$w(k, \Theta) = a + b\left(\rho(k-m) + \sqrt{(k-m)^2 + \sigma^2}\right), \quad \Theta = \{a, b, \rho, m, \sigma\}.$$

This yields a pretty good fit, especially around the money where the majority of the density lies, see Figure 1.

Now we are ready to get smooth densities: we convert the implied volatilities into call prices and then differentiate twice with regards to the strike.

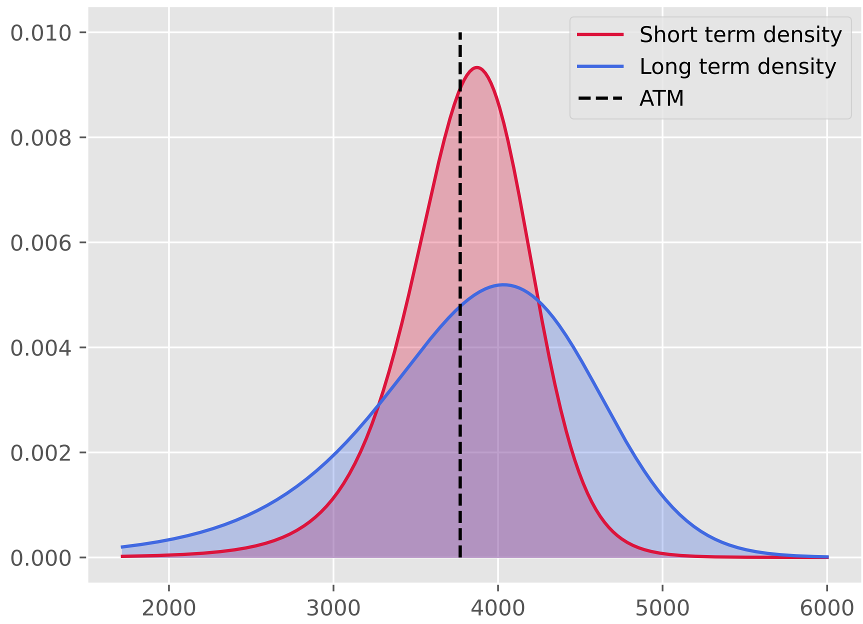

Figure 2: Implied probability densities: the two marginals in between we have to find a martingale transport that minimizes a predefined cost function.

For a financial interpretation of these – expected – results: we observe that the densities tend to shift upward as there are risk-free instruments available on the market (think US bonds) whose appreciation has to be reflected by equities. Moreover the longer term density appears flatter than the shorter term one: more time means more potentialities before the expiry of the option thus accrued variance priced in.

We save the cumulative distribution functions of these interpolated densities in order to sample from them alter.

2.2 Classic OT results

Let's see how classic optimal transport without the martingality condition solves this problem. We rely on `ott-jax` for the implementation [Cut+22].

Using Sinkhorn:

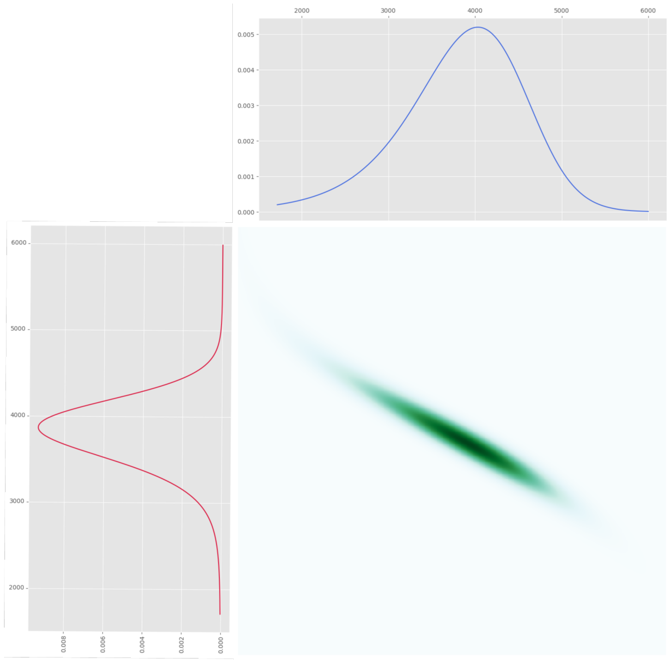

Figure 3: Optimal transport plan – not necessarily a martingale one – between the short term and long term densities, using Sinkhorn with $\varepsilon = 10^{-3}$.

2.3 Programming

First, we have to get familiar with the `jax` framework[3] and how `ott-jax` leverages it.

It looks like we should be defining a new problem type as well as a new solver: we would have something like:

```
from ott.problems.linear import martingale_problem
from ott.solvers.linear import mot
```

Let's stay away from this formalism for now.

We decide to solve the equivalent formulation of the problem Biconjugate dual.

Computing the convex biconjugate

We need to compute the convex enveloppe or a function in the objective. After looking for several ways of computing this feature, we decided to go with a pointwise computation with `numpy` (rather than relying on `tensorflow`'s utility math functions (like `reduce_max` that works across tensors, faster but adds dependencies).

[3] We hone our skills with the quickstart and further documentation.

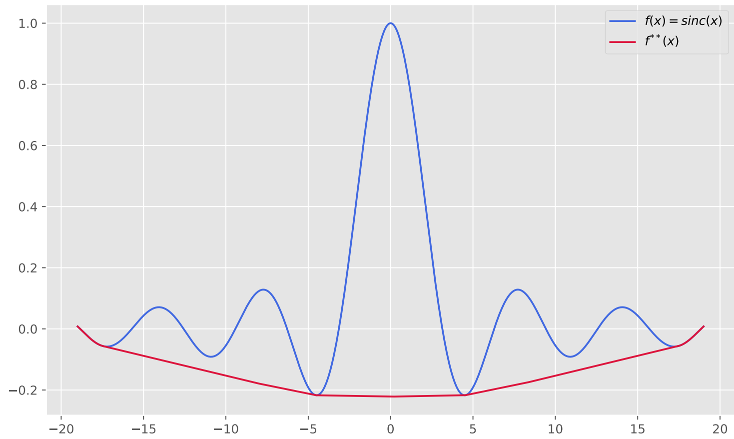

Figure 4: Example of the double application of the Legendre-Fenchel transform to the `sinc` function – we get the convex enveloppe.

Defining a payoff function

The cost function involved in the optimization problem is the payoff. For simplicity, we set the strike of our forward start option at $K = 1$ such that we are entitled to the contingent claim $(S_{t_2} - S_{t_1})^+$ at time t_2.

Minimizing the objective

Back to:

$$\inf_{u \in L^1(\mathbb{P}^2)} \mathbb{E}^{\mathbb{P}^1}\left[(c(S_{t_1}, \cdot) - u(\cdot))^{**}(S_{t_1})\right] + \mathbb{E}^{\mathbb{P}^2}\left[u(S_{t_2})\right] \ldots$$

First thing we notice is that we'd rather work with samples from the two distributions rather than actual densities. To sample from the distribution, we compute the cumulative distribution function, use a PRNG to get a sample for a standard uniform and invert the cdf.

Then, we decide to try and minimize the objective with a neural network, relying on `flax` and `optax`[4]. Basically, we are saying that $u = \mathcal{L}(\mathbf{W}, \mathbf{b})$, where $\mathbf{W}, \mathbf{b}$ are learnable parameters of a neural network, leveraging the universal approximation theorem.

In steps, we need to define:

- A dense network, on which we'll learn u,
- A loss function, involving the convex biconjugate and the sampling from the target distributions,
- A (just-in-time) gradient function,

[4]We take inspiration from this basic `flax` example.

- An optimizer to minimize the loss, we'll go with Adam.

MLP Summary

path	module	inputs	outputs	params
	MLP	float32[500,50…	float32[500,1]	
Dense_0	Dense	float32[500,50…	float32[500,64]	bias: float32[64] kernel: float32[500,6… **32,064 (128.3 KB)**
dense_layers_0	Dense	float32[500,64]	float32[500,64]	bias: float32[64] kernel: float32[64,64] **4,160 (16.6 KB)**
dense_layers_1	Dense	float32[500,64]	float32[500,32]	bias: float32[32] kernel: float32[64,32] **2,080 (8.3 KB)**
Dense_1	Dense	float32[500,32]	float32[500,1]	bias: float32[1] kernel: float32[32,1] **33 (132 B)**
			Total	**38,337 (153.3 KB)**

Total Parameters: 38,337 (153.3 KB)

Figure 5: The `flax` neural network architecture

Once all this is done, we can run the neural network for a defined number of epochs and get the model-free lower price bound for the forward start option. We can get the upper bound in a similar way.

Results

The loss function that we are trying to minimize (our objective in Biconjugate dual) does not give sensible values, and looks quite unstable. We cannot reach convergence of the lower bound, neither of the higher bound.
The most plausible reason for this explosion is in the handling of the convex biconjugate function, as well as the high values we are considering (the strikes gravitate around 4000).

```
❯ python3 src/mot.py
Iteration: 0, Avg. Loss: -976.1793212890625
Iteration: 1, Avg. Loss: -574.822509765625
Iteration: 2, Avg. Loss: -708.5950927734375
Iteration: 3, Avg. Loss: -1823.5506591796875
Iteration: 4, Avg. Loss: -2114.95263671875
Iteration: 100, Avg. Loss: -72275512.0
Iteration: 200, Avg. Loss: -6669571584.0
Iteration: 300, Avg. Loss: 2770108672.0
Iteration: 400, Avg. Loss: 188534816768.0
Iteration: 500, Avg. Loss: -785404592128.0
Iteration: 600, Avg. Loss: 901244846080.0
Iteration: 700, Avg. Loss: 3558679773184.0
Iteration: 800, Avg. Loss: -1015457841152.0
Iteration: 900, Avg. Loss: -2821206310912.0
```

Figure 6: The "loss"... not quite convincing, neither converging.

Navigating the code

The repository tree looks like this:

```
/
├── src
│   ├── __init__.py
│   ├── config.py
│   ├── data_utils.py
│   ├── mot.py
│   └── svi.py
└── data
    ├── long_cdf.npy
    ├── short_cdf.npy
    ├── xs.npy
    └── spx_options.parquet
```

Figure 7: Directory tree

where we will execute the `mot.py` file, where all the neural network machinery lies, while the `config` holds the network parameters. Most of the other files had to do with data preprocessing.

References

[Vil+09] Cédric Villani et al. *Optimal transport: old and new*. Vol. 338. Springer, 2009.

[Cou+11] Areski Cousin et al. "The Skorokhod embedding problem and model-independent bounds for option prices". In: *Paris-Princeton lectures on mathematical finance 2010* (2011), pp. 267–318.

[Bak12] David Baker. "Martingales with specified marginals". In: *Theses, Université Pierre et Marie Curie-Paris VI* (2012).

[BHP13] Mathias Beiglböck, Pierre Henry-Labordere, and Friedrich Penkner. "Model-independent bounds for option prices—a mass transport approach". In: *Finance and Stochastics* 17 (2013), pp. 477–501.

[Hen13] Pierre Henry-Labordere. "Automated option pricing: Numerical methods". In: *International Journal of Theoretical and Applied Finance* 16.08 (2013), p. 1350042.

[HT13] Pierre Henry-Labordere and Nizar Touzi. "An explicit martingale version of Brenier's theorem". In: *arXiv preprint arXiv:1302.4854* (2013).

[GJ14] Jim Gatheral and Antoine Jacquier. "Arbitrage-free SVI volatility surfaces". In: *Quantitative Finance* 14.1 (2014), pp. 59–71.

[Hen17] Pierre Henry-Labordère. *Model-free hedging: A martingale optimal transport viewpoint*. Chapman and Hall/CRC, 2017.

[ACJ19] Aurélien Alfonsi, Jacopo Corbetta, and Benjamin Jourdain. "Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds". In: *International Journal of Theoretical and Applied Finance* 22.03 (2019), p. 1950002.

[GO19] Gaoyue Guo and Jan Obłój. "Computational methods for martingale optimal transport problems". In: *The Annals of Applied Probability* 29.6 (2019), pp. 3311–3347.

[EK21a] Stephan Eckstein and Michael Kupper. "Computation of optimal transport and related hedging problems via penalization and neural networks". In: *Applied Mathematics & Optimization* 83.2 (2021), pp. 639–667.

[EK21b] Stephan Eckstein and Michael Kupper. "Martingale transport with homogeneous stock movements". In: *Quantitative Finance* 21.2 (2021), pp. 271–280.

[Cut+22] Marco Cuturi et al. "Optimal transport tools (ott): A jax toolbox for all things wasserstein". In: *arXiv preprint arXiv:2201.12324* (2022).

[Guy24] Julien Guyon. "Dispersion-constrained martingale Schrödinger problems and the exact joint S&P 500/VIX smile calibration puzzle". In: *Finance and Stochastics* 28.1 (2024), pp. 27–79.

[Ses24] Julian Sester. "A multi-marginal c-convex duality theorem for martingale optimal transport". In: *Statistics & Probability Letters* (2024), p. 110112.