From entropic transport to martingale transport, and applications to model calibration

Jean-David Benamou, Guillaume Chazareix, Grégoire Loeper

From entropic transport to martingale transport, and applications to model calibration

Jean-David Benamou[1], Guillaume Chazareix[1,2], and Grégoire Loeper[2]

[1]INRIA Paris
[2]BNP Paribas Global Markets

June 17, 2024

Abstract

We propose a discrete time formulation of the semi martingale optimal transport problem based on multi-marginal entropic transport. This approach offers a new way to formulate and solve numerically the calibration problem proposed by [19], using a multi-marginal extension of Sinkhorn algorithm as in [6, 11, 7]. In the limit when the time step goes to zero we recover, as detailed in the companion paper [9], a semi-martingale process, solution to a semi-martingale optimal transport problem, with a cost function involving the so-called specific entropy introduced in [14], see also [13] and [3].

1 Introduction

Applications of Semi Martingale Optimal Transport (SMOT) in finance have been the object of several recent studies ([25], [19], [20] amongst others). This framework is particularly well adapted to the problem of model calibration: Find a diffusion model that is compatible with observed option prices. SMOT is the stochastic version of Dynamic Optimal Transport (DOT), that was introduced by [5], as a generalization of OT where transport is achieved by a time dependent flow minimizing the kinetic energy.

While the theoretical aspects of these problems are now well understood, the numerical implementation remains challenging.

In the meantime, a stochastic relaxation of static (i.e. not dynamic) optimal transport, known as *Entropic* Optimal Transport (EOT), has shown to be solvable very efficiently, by the so-called *Sinkhorn* algorithm (see [24] for a review). Interestingly, while there is equivalence between the Static OT problem and its dynamic version, the Entropic regularisation of OT can also be seen either as a static problem, or as variant of the DOT adding a constant volatility diffusion to the governing model, this dynamic problem is known as the Schrödinger problem, see [21].

From a mathematical perspective, all problems (DOT, SMOT and EOT) can be seen as a variant of the same problem: find a process described by the SDE

$$\mathrm{d}X_t = \mu_t \, \mathrm{d}t + \sigma_t \, \mathrm{d}W_t,$$

the induced probability $\mathbb{P}$ on the space of continuous paths, with distribution constraints of the form of moment constraints $\mathbb{E}_{\mathbb{P}}[\psi_i(X_{t_i})] = c_i$[1] , and minimising a Lagrangian

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^T F(\mu_t, \sigma_t^2)\, \mathrm{d}t\right] := \mathcal{F}(\mathbb{P}). \tag{1}$$

[1]Note that prescribing the whole distribution is equivalent to prescribing enough moment constraints

Classical DOT [5] corresponds to the particular case $F = |\mu|^2$ if $\sigma \equiv 0$, $+\infty$ otherwise.

The Semi-Martingale Transport will handle general forms of F, as long as it is convex with respect to (μ, σ^2).

Classical Entropic OT [21] corresponds to the particular case $F = |\mu|^2$ if $\sigma \equiv \overline{\sigma}$ where $\overline{\sigma}$ is constant, $+\infty$ otherwise. It is also one of the formulations of the *Schrödinger's* problem , i.e. minimizing the relative entropy (aka Kullblack Leibler divergence) of $\mathbb{P}$: $\mathrm{KL}(\mathbb{P}|W_{\overline{\sigma}})$, with respect to the Wiener measure $W_{\overline{\sigma}}$ (with a given constant volatility $\overline{\sigma}$), under initial and terminal conditions on the law of $X_{t=0}$ and $X_{t=T}$. Thanks to the properties of the relative entropy, Classical EOT in its static formulation can be solved very efficiently by the Sinkhorn's algorithm.

A general approach (see [4] for a review) is a time discretization that leads to a so-called Multi-Marginal OT problem . In this setting the minimization is performed over the law $\mathbb{P}^h$ (h is the time step) of a vector-valued random variable whose marginals represent densities at each time step. In this paper, we use this time discretization method and an entropic penalization to solve problem (1): We minimize the sum of a discretized form of $\mathcal{F}$ in (1), $\mathcal{F}_h(\mathbb{P}^h)$ (by taking the obvious discrete versions of μ and σ) and the discrete time relative entropy regularization $\mathrm{KL}(\mathbb{P}^h|W^h_{\overline{\sigma}})$.

The drawback of minimising an energy in the form $\mathrm{KL}(\mathbb{P}|W_{\overline{\sigma}})$ is that by essence the minimizer is constrained at $\sigma = \overline{\sigma}$ and cannot satisfy constraints on μ (for instance $\mu \equiv 0$ or $\mu = r$, the interest rate) familiar in finance, since μ is precisely the only degree of freedom used to comply with the distribution constraints. We propose to overcome this issue by considering a proper scaling of the discrete relative entropy and its convergence property : if $\mathbb{P}^h$ a sequence of Markov chains converging to the law of a diffusion $\mathbb{P}$ with (possibly local) drift μ and volatility σ then

$$\lim_{h \searrow 0} h\ \mathrm{KL}(\mathbb{P}^h|W^h_{\overline{\sigma}}) = \mathbb{E}_{\mathbb{P}}\left[\int_0^T \frac{\sigma^2}{\overline{\sigma}^2} - 1 - \log\frac{\sigma^2}{\overline{\sigma}^2}\ \mathrm{d}t\right] =: \mathcal{S}(\mathbb{P}|W_{\overline{\sigma}}), \tag{2}$$

or in short $\mathcal{S}(\sigma|\bar{\sigma})$.

The "specific relative entropy" $\mathcal{S}$ defined above has been introduced in [14] see also [13] [3].

It is shown in [9] that minimizers of $\mathcal{F}_h(\mathbb{P}^h) + h\,KL(\mathbb{P}^h|W^h_{\overline{\sigma}})$ converge in the limit $h \searrow 0$ to a diffusion process $\mathbb{P}$ minimizing the modified cost

$$\mathcal{F}(\mu, \sigma^2) + \mathcal{S}(\mathbb{P}|W_{\overline{\sigma}}).$$

For $h > 0$ we recover a Multi-Marginal EOT, and a discrete Markov chain that can still be used for simulations.

The interest of this approach is not only theoretical: classical methods to solve (1) involve maximizing the dual problem through gradient ascent or primal-dual approaches. These methods imply solving a fully non-linear Hamilton-Jacobi-Bellman equation at each iteration ([5], [23], [18]).

Our approach by Multi-Marginal Sinkhorn's algorithm, extending [6] and [7], computes the same object with the usual convergence guarantees of classical EOT [10] [12].

This paper describes the dual formulation of the problem in the context of local volatility calibration, the associated Sinkhorn algorithm and its practical implementation, with numerical examples.

2 Martingale Optimal Transport for model calibration

The continuous formulation of (Semi-)Martingale Optimal Transport was introduced in [25], and extended for multiple calibration applications as presented in the survey [19]. We are interested here in the one-dimensional formulation of this problem, for the calibration of a local volatility in space and time using a finite number of discrete constraints.

Let $\Omega = C([0, T], \mathbb{R})$, $T > 0$ be the set of continuous paths, and $\mathcal{P}$ the set (or a convex subset of) probability measures on Ω.

The input of the calibration problem is a set of discrete constraints indexed by $i \in \mathcal{I} := \{1, \dots, N_c\}$ described by $(\tau_i, c_i, G_i)_{i\in\mathcal{I}}$ where for each i the triplet (τ_i, c_i, G_i) is maturity, price and payoff function of an observed derivative price on the market.
We will seek for an element $\mathbb{P} \in \mathcal{P}$ such that

$$\mathbb{E}_{\mathbb{P}}[G_i(X_{\tau_i})] = c_i.$$

As an example, calibrating a set of call options at a fixed maturity T would lead to $G_i(x) = (x - K_i)^+$, where K_i is the strike of the i-th option, and $\tau_i = T$ for all i. Moreover, we assume that there are only a finite set of maturities τ_i and thus the set $\mathcal{I}$ can be partitioned as $\mathcal{I} = \bigcup_k \mathcal{I}_k$, $\mathcal{I}_k := \{i \in \mathcal{I}, \tau_i = t_k\}$, where t_k is the k-th distinct maturity in the set of constraints.
To formulate the problem as a constrained minimization problem on $\mathcal{P}$, we restrain our search to the set $\mathcal{P}_s^0 \subset \mathcal{P}$ such that, for each $\mathbb{P} \in \mathcal{P}^0$, $X \in \Omega$ is an $(\mathcal{F}, \mathbb{P})$-semimartingale on $[0, 1]$ given by

$$X_t = X_0 + A_t^{\mathbb{P}} + M_t^{\mathbb{P}}, \quad \langle X\rangle_t = \langle M_t^{\mathbb{P}}\rangle = B^{\mathbb{P}}, \quad \mathbb{P}\text{-a.s.}, \quad t \in [0, 1],$$

where $M^{\mathbb{P}}$ is an $(\mathcal{F}, \mathbb{P})$-martingale on $[0, 1]$ and $(A^{\mathbb{P}}, B^{\mathbb{P}})$ is $\mathcal{F}$-adapted and $\mathbb{P}$-a.s. absolutely continuous with respect to time. In particular, $\mathbb{P}$ is said to be have characteristics $(\mu, \sigma^2)(\mathbb{P})$, which are defined in the following way,

$$\mu = \frac{\mathrm{d}A_t^{\mathbb{P}}}{\mathrm{d}t}, \sigma_t^2 = \frac{\mathrm{d}B_t^{\mathbb{P}}}{\mathrm{d}t},$$

Note that (μ, σ^2) is $\mathcal{F}$-adapted and determined up to $d\mathbf{P} \times dt$, almost everywhere. We now let $F(t, x, a, b)$ be convex with respect to (a, b) for every (t, x), and seek for

$$\mathcal{V} = \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \int_0^T F(t, X_t, \mu, \sigma^2)\ \mathrm{d}t, \tag{CMOT}$$

In the calibration case we alsos impose $X_0 = x_0 \in \mathbb{R}$ (i.e. $X_0 \sim \delta_{x_0}$) as the derivative price is known at time 0.

At this stage, the processes μ, σ are very general and can be generally path-dependent, however, as showed in [17], they can be chosen as local processes, i.e. functions of (t, X_t) only: indeed, for any choice of μ, σ, there exists a local version $\mu(t, x), \sigma(t, x)$ that preserves the constraints (i.e. option prices) and that can only reduce the cost (CMOT). The minimization problem can therefore be reduced to X_t solutions of the stochastic differential equation:

$$\mathrm{d}X_t = \mu(X_t, t)\ \mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t, \tag{SDE}$$

where B_t is a standard Brownian motion.

The function F can be decomposed into a sum of model constraints F_{mc}, calibration constraints F_{cc} and regularization F_r components:

- Model constraints: for instance, if we want to impose that the underlying X_t follows a pure diffusion model, i.e. $\mu \equiv 0$ this can be imposed by choosing F_{mc} as :

$$F_{mc}(\mu) = \begin{cases} 0 & \text{if } \mu = 0 \\ +\infty & \text{otherwise.} \end{cases}$$

  (or a soft version).

- Calibration constraints expressed as

$$F_{cc}(X_t) = \begin{cases} 0 & \text{if } \mathbb{E}[G_i(X_{\tau_i})] = c_i. \\ +\infty & \text{otherwise.} \end{cases}$$

  (or a penalisation of the constraint).

- Regularization/model assumptions: F_r helps enforce qualitative properties of the model. We might want for instance σ to be close to a prescribed guess $\overline{\sigma}$, which can be enforced by choosing F as a penalty function of the form $F(\sigma^2/\bar{\sigma}^2)$, for instance Loeper [22] uses

$$F = (\sigma - \bar{\sigma})^2$$

(which is convex in σ^2 and is linked to the Bass martingale problem), or in [19] they use

$$F_r(\sigma^2) = a\left(\frac{\sigma^2}{\overline{\sigma^2}}\right)^p + b\left(\frac{\sigma^2}{\overline{\sigma^2}}\right)^{-q} + c \tag{3}$$

with $p, q, a, b > 0$ and $c \in \mathbb{R}$ such that F is convex with minimum at $\sigma = \overline{\sigma}$. It is also a barrier as σ^2 goes to 0.

In this paper we choose

$$F_r(\sigma^2) = \mathcal{S}(\sigma|\overline{\sigma}).$$

with $\mathcal{S}$ defined in (2). As explained in the introduction it allows to discretize (in time) the problem as a multi-marginal EOT problem. It is again convex with minimum at $\sigma = \overline{\sigma}$ and a barrrier as σ^2 goes to 0 but unlike (3) it is sublinear as $\sigma \nearrow +\infty$. This difficulty is discussed in [9].

3 Discretisation into a Multi-Marginal Martingale Transport

3.1 Notations

We will discretize our problem in time, replacing the interval $[0, T]$ with a regular grid of $N_T + 1$ timesteps $t_k = k\,h$ for $k \in \{0, \dots, N_T\} =: \mathcal{K}^h$, where $h := T/N_T$ is the time step. We impose that all the calibration times τ_i are included in the grid, i.e. $\tau_i = t_{k_i}$ for some $k_i \in \mathcal{K}^h$.

Instead of functions $t \mapsto \omega(t)$, we consider their discrete path counterparts, which are n-tuples $(\omega_0, \dots, \omega_{N_T}) \in \mathbb{R}^{N_T+1}$ for $k \in \mathcal{K}^h$, in which ω_k corresponds to the value of the path at time t_k. Instead of $\mathbb{R}^{N_T}$, we denote by $\mathcal{X}_k$ the space of values that ω_k can take, and by $\Omega^h := \Pi_{i=0}^{N_T}\mathcal{X}_i$ the space of discrete paths.

An element $(t \mapsto \omega(t)) \in \Omega$ is hence replaced by a n-tuple $(\omega_0, \dots, \omega_{N_T}) \in \Omega^h$ with $\omega_k \in \mathcal{X}_k$ for $k \in \mathcal{K}^h$.

We are hence searching for a probability measure $\mathbb{P}^h$ on Ω^h. We denote $(X_k)_{k\in\mathcal{K}^h}$ the canonical process of $\mathbb{P}^h$ on Ω^h. We will denote by $\mathbb{P}^h_k := X_k\#\mathbb{P}^h$ the marginal law of $\mathbb{P}^h$ at timestep $k \in \mathcal{K}^h$, and by $\mathbb{P}^h_{k,l} := (X_k, X_l)\#\mathbb{P}^h$ the joint law of time steps $k \in \mathcal{K}^h$ and $l \in \mathcal{K}^h$.

We note $\mathcal{K}^h_{-i} = k \in \mathcal{K}^h \setminus \{i\}$ the set of timesteps except timestep i, and $\mathrm{d}x_{-i} = \prod_{\mathcal{K}^h_{-i}} \mathrm{d}x_k$, which allows to write the marginal law $\mathbb{P}^h_k$ as $\mathbb{P}^h_k = \int \mathbb{P}^h(x_k, \mathrm{d}x_{-k})$ and joint laws in a similar fashion. Similarly, we note $\mathrm{d}x_{[i,j]} = \Pi_{k=i}^{j}\,\mathrm{d}x_k$.

We note ρ_0 the initial marginal of our process, which is imposed, $X_0 \sim \rho_0$. It may or may not be a Dirac in our case.

We denote by $\overline{\mathbb{P}^h}$ the reference measure on Ω^h that we will use to regularize the problem. We will denote by $(Y_k)_{k\in\mathcal{K}^h}$ the canonical process of $\overline{\mathbb{P}^h}$ on Ω^h. It's law is determined by a Euler-Maruyama discretisation of the continuous reference process :

$$Y_{k+1} = Y_k + \overline{\mu}(Y_k, kh)\,h + \overline{\sigma}(Y_k, kh)\,h^{1/2}\,Z_k, \quad \forall k \in \mathcal{K}^h_{-0},\ Y_0 \sim \rho_0.$$

We write $\mathcal{P}^h_{\mathrm{EM}}$ the set of probability measures on Ω^h whose canonical process Y_k can be writen as a Euler-Maruyama discretisation as such, with $\mu_k = \mu(Y_k, kh)$ and $\sigma_k = \sigma(Y_k, kh)$.

For any probability measure $\mathbb{P}^h \in \mathcal{P}(\Omega^h)$ and $\overline{\mathbb{P}^h} \in \mathcal{P}(\Omega^h)$, we note $\mathrm{KL}(\mathbb{P}^h|\overline{\mathbb{P}}) = \mathbb{E}_{\mathbb{P}^h}\left[\log\left(\frac{\mathrm{d}\mathbb{P}^h}{\mathrm{d}\overline{\mathbb{P}^h}}\right) - 1\right]$ the Kullback-Leibler divergence between $\mathbb{P}^h$ and $\overline{\mathbb{P}^h}$ if $\mathbb{P}^h \ll \overline{\mathbb{P}^h}$. By convention, if $\mathbb{P}^h \not\ll \overline{\mathbb{P}^h}$, we set $\mathrm{KL}(\mathbb{P}^h|\overline{\mathbb{P}^h}) = +\infty$.

For two continuous diffusion processes $\mathbb{P}$ with volatility σ^2 and $\overline{\mathbb{P}}$ with volatility $\overline{\sigma}^2$, we note $\mathcal{S}(\sigma^2|\overline{\sigma}^2) = \mathbb{E}_{\mathbb{P}}\left[\int \frac{\sigma^2}{\overline{\sigma}^2} - 1 - \log\left(\frac{\sigma^2}{\overline{\sigma}^2}\right) \mathrm{d}t\right]$ the specific entropy between $\mathbb{P}$ and $\overline{\mathbb{P}}$. For any two discrete probabilities in $\mathcal{P}^h_{\mathrm{EM}}$, we will denote $\mathcal{S}^h(\mathbb{P}^h|\overline{\mathbb{P}^h}) = \mathbb{E}_{\mathbb{P}^h}\left[h\sum_{k=0}^{N_T} \frac{\sigma_k^2}{\overline{\sigma_k}^2} - 1 - \log\left(\frac{\sigma_k^2}{\overline{\sigma_k}^2}\right)\right]$ the discrete specific entropy between $\mathbb{P}^h$ and $\overline{\mathbb{P}^h}$, which is a Riemann sum discretizing the continuous specific entropy.

3.2 Discrete drifts and diffusions coefficients

As opposed to the continuous-time approach, which uses Markovian projections of the processes, and as such the variable being optimised are functions representing the drift and volatility, in this discrete-time approach, we will directly optimise on $\mathbb{P}^h$. In order to justify the choice of moment variables in the discrete problem, we first consider Euler-Maruyama discretization of a diffusion process. Let a diffusion process X_t with drift μ and volatility σ, following the SDE :

$$\mathrm{d}X_t = \mu(X_t, t)\ \mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}W_t.$$

Consider the Euler-Maruyama time discretization of the process:

$$X^h_{k+1} = X^h_k + \mu(X^h_k, kh)\,h + \sigma(X^h_k, kh)\,h^{1/2}\,Z_k$$

where $\forall k \in \{0, \dots, N_T\} := \mathcal{K}^h$, Z_k is a standard normal random variable, of which we note $\mathbb{P}^h$ the law. We have $\mathbb{P}^h \in \mathcal{P}^h_{\mathrm{EM}}$.

For such a process, we can compute the following quantities from conditional expectations :

$$\beta_k(x) = \frac{1}{h}\,\mathbb{E}[X_{k+1} - X_k \mid X_k = x] = \mu(x, kh), \tag{4}$$

$$\alpha_k(x) = \frac{1}{h}\,\mathbb{E}\big[(X_{k+1} - X_k)^2 \;\big|\; X_k = x\big] = \mu(x, kh)^2\,h + \sigma^2(x, kh) \xrightarrow[h\to 0]{} \sigma^2(x, kh). \tag{5}$$

These variables are computed from the law $\mathbb{P}^h$ and are the discrete counterpart of the drift and volatility of the continuous process. They can hence be used to compute the discretisation of $F(t, X_t, \mu, \sigma^2)$.

A more general framework is to consider a variable $b_k : \mathcal{X} \to \mathbb{R}^K$ defined by taking the conditional expectation of a general function $B : (\mathcal{X}, \mathcal{X}) \to \mathbb{R}^K$ depending on two consecutive timesteps :

$$b_k(x) = \frac{1}{h}\,\mathbb{E}[B(X_k, X_{k+1}) \mid X_k = x].$$

The vector formed with variables β_k and α_k can be computed as such by using the function $B(X, Y) = \begin{bmatrix}(Y - X)\\(Y - X)^2\end{bmatrix}$.

We might alternatively want to control other types of moments, such as the skewness or the kurtosis of the process. For this reason, we will consider a general function B and the corresponding b_k in the following, and not specifically β_k and α_k.

3.3 Specific relative entropy

We give a formal derivation of the Specific Entropy (see [14] [13] [3]).

The Kullback-Leibler divergence between two normal laws $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$ is equal to :

$$\mathrm{KL}(\mathcal{N}(\mu_1, \sigma_1^2)|\mathcal{N}(\mu_2, \sigma_2^2)) = \frac{1}{2}\left(\frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{\sigma_2^2} - 1 - \log\left(\frac{\sigma_1^2}{\sigma_2^2}\right)\right).$$

Consider two diffusion measures $\mathbb{P}$ and $\overline{\mathbb{P}}$ defined by the following SDEs on their respective canonical processes X and Y :

$$\mathrm{d}X_t = \mu(X_t, t)\ \mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}W_t,\ X_0 \sim \rho_0, \tag{6}$$

$$\mathrm{d}Y_t = \overline{\mu}(Y_t, t)\ \mathrm{d}t + \overline{\sigma}(Y_t, t)\,\mathrm{d}W_t,\ Y_0 \sim \rho_0. \tag{7}$$

We can discretize on a grid of step h as law $\mathbb{P}^h$ and $\overline{\mathbb{P}^h}$ using the Euler-Maruyama discretization from previous section, giving their respective canonical processes X^h and Y^h :

$$X^h_{k+1} = X^h_k + \mu(X^h_k, kh)\,h + \sigma(X^h_k, kh)\,h^{1/2}\,Z_k, \quad \forall k \in \mathcal{K}^h,\ X^h_0 \sim \rho_0, \tag{8}$$

$$Y^h_{k+1} = Y^h_k + \overline{\mu}(Y^h_k, kh)\,h + \overline{\sigma}(Y^h_k, kh)\,h^{1/2}\,Z_k, \quad \forall k \in \mathcal{K}^h, Y^h_0 \sim \rho_0. \tag{9}$$

Noting $\mu_k(x) = \mu(x, kh)$, $\sigma_k(x) = \sigma(x, kh)$, and $\overline{\mu}_k(x) = \overline{\mu}(x, kh)$ and $\overline{\sigma}_k(x) = \overline{\sigma}(x, kh)$, we can conclude the transitions laws $\mathbb{P}^h_{k,k+1}$ and $\overline{\mathbb{P}^h}_{k,k+1}$ are normal laws, and the Kullback-Leibler divergence can then be decomposed as follows :

$$\begin{aligned}
h\ \mathrm{KL}(\mathbb{P}^h|\overline{\mathbb{P}^h}) &= h\ \textstyle\sum_{k=0}^{N_T-1} \int \mathrm{KL}(\mathcal{N}(\mu_k(x)h, \sigma_k(x)^2 h)|\overline{\mathbb{P}^h}_{k,k+1})\,\rho_k(dx) \\
&= \tfrac{h}{2}\ \textstyle\sum_{k=0}^{N_T-1} \int \left(\frac{\sigma_k(x)^2 h + ((\mu_k(x) - \overline{\mu}_k(x))h)^2}{\overline{\sigma}_k(x)^2 h} - 1 - \log\left(\frac{\sigma_k(x)^2 h}{\overline{\sigma}_k(x)^2 h}\right)\right)\,\rho_k(dx) \\
&= \tfrac{h}{2}\ \textstyle\sum_{k=0}^{N_T-1} \int \left(\frac{\sigma_k(x)^2 + (\mu_k(x) - \overline{\mu}_k(x))^2 h}{\overline{\sigma}_k(x)^2} - 1 - \log\left(\frac{\sigma_k(x)^2}{\overline{\sigma}_k(x)^2}\right)\right)\,\rho_k(dx) \\
&\xrightarrow[h\to 0]{} \mathcal{S}(\sigma^2|\overline{\sigma}^2) := \tfrac{1}{2}\int \mathbb{E}_{\mathbb{P}^h}\left[\frac{\sigma^2}{\overline{\sigma}^2} - 1 - \log\left(\frac{\sigma^2}{\overline{\sigma}^2}\right)\right]\mathrm{d}t
\end{aligned} \tag{10}$$

More generally, it is shown in [9] that the discrete Kullback-Leibler Divergence allows to control the approximation of the volatility of the discrete process (5). This motivates to use the specific entropy as a regulariser of the continuous problem, because it is linked to a natural discretization in terms of the Kullback-Leibler divergence, and entropy-regularized optimal transport is a thoroughly studied problem.

3.4 Discretisation

We will now use the previously defined tools to discretize the continuous problem (CMOT) in time, in the case of a specific entropy regularizer, since it has a natural discretization in terms of the Kullback-Leibler divergence.

We first directly discretize (CMOT) in time as a Riemann sum and using (10) and the variables β_k and α_k defined in equations (4) and (5), we obtain :

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^T (F + \mathcal{S})(t, X_t, \mu, \sigma^2)\,\mathrm{d}t\right] \approx h\ \mathbb{E}_{\mathbb{P}^h}\left[\sum_{k=0}^{N_T} F(t, X_k, \beta_k(X_k), \alpha_k(X_k))\right] + h\ \mathrm{KL}(\mathbb{P}^h|\overline{\mathbb{P}^h})$$

where $\mathcal{S}$ the specific entropy introduced in section 3.3.

We can hence formulate our discretization of Problem CMOT as :

$$\inf_{\mathbb{P}^h \in \mathcal{P}^h} h\ \mathbb{E}_{\mathbb{P}^h}\left[\sum_{k=0}^{N_T} F(t, X_k, \beta(X_k), \alpha(X_k))\right] + h\,\mathrm{KL}(\mathbb{P}^h|\overline{\mathbb{P}^h}) \tag{11}$$

where $\mathcal{P}^h$ is the set of probability measures on n-uplets respecting the constraints, that is :

$$\mathcal{P}^h = \{\mathbb{P}^h \in \mathcal{P}(\mathbb{R}^{N_T}),\ \text{s.t. } X_0 \# \mathbb{P}^h = \mu_0 \text{ and } \forall i \in \mathcal{I},\ \mathbb{E}_{\mathbb{P}^h}[G_i(X_{\tau_i})] = c_i\}$$

We emphasize the fact that while $\mathbb{P}^{h,\star}$ solution of this problem is a measure that respects the initial condition and the price constraints similarly to the continuous problem, it is not obvious that it is a discrete diffusion Markov chain, as we used in the previous informal derivation. This is discussed in [8].

We will generalize this problem in the next section to allow for more general constraints on the marginals and the model prices, and we will then find its dual problem using Fenchel-Rockafellar duality.

4 Duality

We first recall the Fenchel-Rockafellar theorem and hence the form of the primal problem we aim to formulate.

Theorem 4.1 (Fenchel-Rockafellar)**.** *Let (E, E^*) and (F, F^*) be two couples of topologically paired spaces. Let $\Delta : E \to F$ be a continuous linear operator and $\Delta^\dagger : F^* \to E^*$ be its adjoint. Let $\mathcal{F} : E^* \to \overline{\mathbb{R}}$ and $\mathcal{G} : F^* \to \overline{\mathbb{R}}$ be two lower semicontinuous and proper convex functions. If there exists $\mathbb{P} \in F^*$ such that $\mathcal{G}(\mathbb{P}) < +\infty$ and $\mathcal{F}$ is continuous at $\Delta^\dagger \mathbb{P}$, then :*

$$\sup_{\Phi \in E} -\mathcal{F}^\star(-\Phi) - \mathcal{G}^\star(\Delta\Phi) = \inf_{\mathbb{P} \in F^*} \mathcal{F}(\Delta^\dagger \mathbb{P}) + \mathcal{G}(\mathbb{P}),$$

and the inf *is attained. Moreover, if there exists a maximizer $\Phi^\star \in E$, then there exists $\mathbb{P}^* \in F^*$ satisfying $\Delta\Phi^\star \in \partial\mathcal{G}(\mathbb{P})$ and $\Delta^\dagger \mathbb{P} \in -\partial\mathcal{F}^\star(-\Phi^\star)$.*

We note the primal problem :

$$\mathcal{V} := \min_{\mathbb{P} \in F^*} \mathcal{F}(\Delta^\dagger \mathbb{P}) + \mathcal{G}(\mathbb{P}), \tag{PRIMAL}$$

and the dual :

$$\mathcal{D} := \sup_{\Phi \in E} -\mathcal{F}^\star(-\Phi) - \mathcal{G}^\star(\Delta\Phi). \tag{DUAL}$$

We will define their corresponding objects in the next sections.

4.1 Primal problem

In this part, we will formulate a generalization of the problem (11), in a standard form compatible with the Fenchel-Rockafellar duality. We will first introduce the variables that we will control, and then define the primal problem.

We aim to define the objects corresponding to the primal problem, that is $\mathcal{F}$, $\mathcal{G}$ and the linear operator $\Delta^\dagger$. The operator $\Delta^\dagger$ will include variables of interest that we will control. We will define these variables in the following remark.

Definition 4.1. *For each timestep k, we define the following variables :*

$$\begin{aligned} \nu_k(\mathrm{d}x) &= X_k \# \mathbb{P}^h(\mathrm{d}x) \\ b_k(x) &= \frac{1}{h}\, \mathbb{E}_{\mathbb{P}^h_{k,k+1}}\left[B(X^h_k, X^h_{k+1}) \mid X^h_k = x\right] \\ g_i &= \mathbb{E}_{\mathbb{P}^h_{\tau_i}}[G_i(X_{\tau_i})]. \end{aligned}$$

and We define $\Delta^\dagger$ as the following linear operator :

$$\begin{aligned} \Delta^\dagger : \quad E^* := \mathcal{P}(\Omega^h) \quad &\to \quad F^* := (\otimes_{k=1}^{N_T} \mathcal{P}(\mathcal{X}_k)) \times (\otimes_{k=1}^{N_T} \mathcal{M}(\mathcal{X}_k)) \times \mathbb{R}^K \\ \mathbb{P}^h \quad &\to \quad \Delta^\dagger \mathbb{P}^h := (\nu_k,\, \nu_k b_k, g_i) \end{aligned}$$

ν_k, representing the marginal laws of $\mathbb{P}^h$, are linear with respect to $\mathbb{P}^h$, as they are a projection on a basis vector. g_i are the model prices at the calibration times and are also linear with respect to $\mathbb{P}^h$, since they are the expectation of a function of the state at a given time. b_k are general variables corresponding to moments defined as the conditional expectation of a function B between two consecutive timesteps as mentioned in section 3.2. Since they are a conditional expectation, they are not linear with respect to $\mathbb{P}^h$. For this reason, we instead consider the product $\nu_k b_k$ that is linear with respect to $\mathbb{P}^h$, as is done in the classical and martingale optimal transport [5] [25]. Indeed, this product is equal to:

$$\nu_k b_k(x) = \frac{1}{h}\,\mathbb{E}_{\mathbb{P}^h_{k,k+1}(x,\cdot)}\big[B(x, X^h_{k+1})\big]$$

which is linear with respect to $\mathbb{P}^h$ as a simple expectation with respect to a projection.

Next, we define our function $\mathcal{G}$ as the scaled Kullback-Leibler divergence between $\mathbb{P}^h$ and a reference measure $\overline{\mathbb{P}^h}$, as it is the only function in our problem that depends on the measure itself.

Definition 4.2. *We define $\mathcal{G}$ the following functional :*

$$\mathcal{G} : \mathbb{P} \in F^* \to h\,\mathrm{KL}(\mathbb{P}|\overline{\mathbb{P}^h}).$$

Finally, $\mathcal{F}$ will be a functional of the variables defined in remark 4.1, and will be a sum of convex functions of these variables, as presented for the continuous case at the end of Section 3.2. These functions may be regularization, soft, or hard constraints.

Remark 4.1. *Let us say we want to impose a variable x to be close to a target x_0.*

*A **hard constraint** is of the form :*

$$F(x) = \begin{cases} 0 & \text{if } x = x_0 \\ +\infty & \text{otherwise} \end{cases}$$

*Some example of **soft constraints** are :*

- $F(x) = C(x - x_0)^2$ *: the quadratic penalty*
- $F(x) = x\log(x/x_0) + x - x_0$ *: the Kullback-Leibler divergence*

We want to keep the formulation of our problem general and be able to express at least problem (11) in this framework. We hence want to be able to express :

- The constraint $X_0 \# \mathbb{P}^h = \delta_{x_0}$, but more generally we will consider any constraint on the marginals of $\mathbb{P}^h$, leading to the definition of a term $\mathcal{F}_{\text{marg}} = \sum_{k=0}^{N_T} M_k(\nu_k)$ which is a sum of convex, l.s.c., potentially null functions of the marginals.
- The constraint $\mathbb{E}_{\mathbb{P}^h}[G_i(X_{\tau_i})] = g_i$, but more generally we will consider any constraint on the model prices, leading to defining a term $\mathcal{F}_{\text{prices}} = \sum_{i=1}^{N_C} C_i(g_i)$ which is a sum of convex, l.s.c., potentially null functions of the model prices.
- The constraints on moments of the process lead to the definition of a term $\mathcal{F}_{\text{mom}} = \sum_{k=0}^{N_T-1} \mathcal{F}_{\text{mom},k}(\nu_k, \nu_k b_k)$ that is a sum of perspective functions of the marginals and the product of the marginals and the moments.

which leads to the following definition of $\mathcal{F}$:

Definition 4.3. *We define $\mathcal{F}$ the following functional :*

$$\begin{aligned}\mathcal{F}(\nu, b, g) &= \mathcal{F}_{marg} + \mathcal{F}_{mom} + \mathcal{F}_{prices} \\ &= \sum_{k=0}^{N_T-1} \mathcal{F}_{mom,k}(\nu_k, \nu_k b_k) + \sum_{k=0}^{N_T} M_k(\nu_k) + \sum_{i=1}^{N_C} C_i(g_i)\end{aligned}$$

Remark 4.2. *Problem* (11) *may be retrieved from problem the problem we built with* $\mathcal{F}_{mom,k}(\nu_k, \nu_k b_k) = h\,\mathbb{E}_{\mathbb{P}^h}[F(\nu_k b_k/h\nu_k)]$, $\forall k$, M_0 *and* C_i, $\forall i \in \mathcal{I}$ *hard constraints and* $M_k = 0$, $\forall k \neq 0$, *and by taking* b_k *to be the vector of variables* β_k *and* α_k *as explained at the end of section 3.2. This function* $\mathcal{F}_{mom,k}$ *is then a perspective function and is well studied in the theory of convex optimization.*

4.2 Dual

In order to apply the Fenchel-Rockafellar duality, we need to compute the Legendre transforms of $\mathcal{F}$ and $\mathcal{G}$, and the adjoint operator of $\Delta^\dagger$.

Lemma 4.1. $\mathcal{G}^\star$, *the Legendre-Fenchel transform of* $\mathcal{G}$ *is given by :*

$$\begin{aligned} \mathcal{G}^\star : \quad F = \mathcal{C}_b(\Omega^h) &\to \overline{\mathbb{R}} \\ f &\to h\mathbb{E}_{\overline{\mathbb{P}^h}}(\exp(f/h)) \end{aligned}$$

Proposition 4.1. *The Legendre transform of* $\mathcal{F}$ *is given by :*

$$\begin{aligned} \mathcal{F}^\star(\phi_\nu, \phi_b, \lambda_g) = &\sum_k \inf_\psi M_k^\star(\phi_{\nu_k} - \psi) + F_k^\star(\psi, \phi_b) \\ &+ \sum_{i=1}^{N_C} C_i^\star(\lambda_{g_i}) + M_{N_T}^\star(\phi_{\nu_{N_T}}) \end{aligned} \tag{$\mathcal{F}$-dual}$$

Proof. We first notice that the expression of $\mathcal{F}$ is separable in k and i and that g_i can be separated from $(\nu_k, \nu_k b_k)$. We can then rewrite the function as :

$$\mathcal{F}(\nu, \nu b, g) = \sum_{k=0}^{N_T - 1} \mathcal{F}_k^1(\nu_k, \nu_k b_k) + \sum_{i=0}^{N_C} \mathcal{F}_i^2(g_i) + M_{N_T}(\nu_{N_T}) \tag{12}$$

where :

$$\begin{aligned} \mathcal{F}_k^1(\nu_k, \nu_k b_k) &= \mathcal{F}_{\mathrm{mom},k}(\nu_k, \nu_k b_k) + M_k(\nu_k) \\ \mathcal{F}_i^2(g_i) &= C_i(g_i). \end{aligned}$$

The Legendre transform of $\mathcal{F}_i^2$ are the Legendre transform of C_i and the Legendre transform of $\mathcal{F}_k^1$ is given by :

$$\begin{aligned} \mathcal{F}_k^{1\star}(\phi_\nu, \phi_b) &= \sup_{\nu, b} \phi_\nu \nu + \phi_b b - \mathcal{F}_k^1(\nu, b) - M_k(\nu) \\ &= \sup_{\nu, \nu', b} \inf_\psi \psi(\nu' - \nu) + \phi_\nu \nu + \phi_b b - \mathcal{F}_k^1(\nu', b) - M_k(\nu) \\ &= \inf_\psi \sup_{\nu, \nu', b} (\phi_\nu - \psi)\nu + \phi_b b + \psi\nu' - \mathcal{F}_k^1(\nu', b) - M_k(\nu) \\ &= \inf_\psi \sup_\nu (\phi_\nu - \psi)\nu - M_k(\nu) + \sup_{\nu', b} \psi\nu' + \phi_b b - \mathcal{F}_k^1(\nu', b) \\ &= \inf_\psi M_k^\star(\phi_\nu - \psi) + F_k^\star(\psi, \phi_b) \end{aligned}$$

We can conclude by summing over k and i the separable parts. □

We are particularly interested in the case $\mathcal{F}_k(\nu_k, \nu_k b_k) = h\,\mathbb{E}_{\nu_k}(F(\nu_k b_k/h\nu_k))$ where F is a convex function, so $\mathcal{F}_k$ is a perspective function. This is because it is a direct discretisation of the continuous

problem. For this case, we first find the Legendre-Transform of $\mathcal{F}_k$ coordinate-wise :

$$\begin{aligned}
\mathcal{F}_k^\star(\phi_\nu, \phi_b) &= \sup_{\nu,b} \ \phi_\nu \nu + \phi_b b - \mathcal{F}_k(\nu, b) \\
&= \sup_{\nu,b} \ \phi_\nu \nu + \phi_b b - h\nu F(b/h\nu) \\
&= \sup_{\nu} \ \phi_\nu \nu + h\nu \left(\sup_b \phi_b (b/h\nu) - F(b/h\nu) \right) \\
&= \sup_{\nu} \ \phi_\nu \nu + h\nu F^\star(\phi_b) \\
&= \sup_{\nu} \ \left(\phi_\nu + hF^\star(\phi_b)\right) \nu \\
&= \begin{cases} 0 & \text{if } \phi_\nu + hF^\star(\phi_b) < 0 \\ +\infty & \text{otherwise} \end{cases}
\end{aligned}$$

We can hence obtain the Legendre-Transform of $\mathcal{F}_k^1$ defined above as :

$$\begin{aligned}
\mathcal{F}_k^1(\phi_\nu, \phi_b) &= \inf_\psi M_k^\star(\phi_\nu - \psi) + \mathcal{F}_k^\star(\psi, \phi_{b_k}) \\
&= M_k^\star(\phi_\nu + hF^\star(\phi_b))
\end{aligned}$$

The dual of $\mathcal{F}$ is given by the following simpler expression :

$$\begin{aligned}
\mathcal{F}^\star(\phi_\nu, \phi_b, \lambda_g) = &\sum_k M_k^\star(\phi_{\nu_k} + hF^\star(\phi_b)) \\
&+ \sum_{i=1}^{N_C} C_i^\star(\lambda_{g_i}) + M_{N_T}^\star(\phi_{\nu_{N_T}})
\end{aligned}$$

Lemma 4.2. *The adjoint operator to $\Delta^\dagger$ is given by :*

$$\begin{aligned}
\Delta : \Phi = (\phi_{\nu_k},\, \phi_{b_k},\, \lambda_{g_i}) \to &\bigoplus_k h\phi_{\nu_k} + B(x_k, x_{k+1})\phi_{b_k} \\
&+ \sum_i G_i(x_{\tau_i})\lambda_{g_i}
\end{aligned}$$

Proof. We can easily check that $\langle \Delta\Phi, \mathbb{P}^h \rangle = \langle \Phi, \Delta^\dagger \mathbb{P}^h \rangle$. □

As customary in optimal transport, through the Monge-Kantorovitch dual formulation , optimal solutions (but for non-optimal constraints) can be obtained from the dual potentials. Thus, by constructing a maximizing sequence of the dual problem, we hope to find a sequence of measures converging to a solution of the primal problem. This is the approach we follow in this work.

Proposition 4.2. *Let $\Phi^\star = (\phi_\nu^\star, \phi_b^\star, \lambda_g^\star)$ be a solution of DUAL. Then $\Phi^\star$ induces a measure $\mathbb{P}^{h,\star}$ through*

$$\frac{\mathrm{d}\mathbb{P}^{h,\star}}{\mathrm{d}\overline{\mathbb{P}^h}} = \exp\left(\frac{\Delta\Phi^\star}{h}\right),$$

$\mathbb{P}^{h,\star}$ is the optimal solution of (PRIMAL) for the constraints functions M, C that are finite under $\mathbb{P}^{h,\star}$.

Proof. The first optimality condition is given by :

$$\Delta\Phi^\star \in h\, \partial_{\mathbb{P}^{h,\star}} \mathrm{KL}(\mathbb{P}^{h,\star} | \overline{\mathbb{P}^h})$$

which leads to :

$$\Delta\Phi^\star = h \log\left(\frac{\mathrm{d}\mathbb{P}^{h,\star}}{\mathrm{d}\overline{\mathbb{P}^h}}\right) \tag{13}$$

□

Definition 4.4. *Let $\Delta_{i,i+1}$ the transitional part of Δ given by:*

$$\Delta_{i,i+1}(x_i, x_{i+1}) = B(x_i, x_{i+1})\phi_{b_i}(x_i),$$

Let ψ_i^u, ψ_i^d potential functions computing the forward and backward influence :

$$\psi_i^u(x_i) = \log\left(\int \rho_0 \prod_{k=0}^{i-1} \exp\left(\frac{\Delta_{k,k+1}}{h} + \phi_{\nu_k} + \overrightarrow{\Lambda_k} \cdot \overrightarrow{G_k}\right) \overline{\mathbb{P}^h_{k,k+1}}\, \mathrm{d}x_{[0,i-1]}\right)$$

$$\psi_i^d(x_i) = \log\left(\int \prod_{k=i}^{N_T-1} \exp\left(\frac{\Delta_{k,k+1}}{h} + \phi_{\nu_{k+1}} + \overrightarrow{\Lambda_{k+1}} \cdot \overrightarrow{G_{k+1}}\right) \overline{\mathbb{P}^h_{k,k+1}}\, \mathrm{d}x_{[i+1,N_T]}\right)$$

and let $\overrightarrow{\Lambda}_k$ the vector of Lagrange multipliers and $\overrightarrow{G_k}$ the vector of their corresponding payoff functions associated with timestep k:

$$\overrightarrow{\Lambda}_k = (\lambda_{g_i})_{i \in \mathcal{I}_k}$$
$$\overrightarrow{G_k} = (G_i)_{i \in \mathcal{I}_k}$$

One interesting property of this formulation is that the Markovianity of the solution directly arises from the structure of the linear operator Δ.

Proposition 4.3. *Let $\mathbb{P}^{h,\star}$ an optimal solution of* (PRIMAL)*. The following properties hold true :*

- *Its joint density with respect to the reference measure is given by :*

$$\begin{aligned}\mathrm{d}\mathbb{P}^{h,\star}_{i,i+1}(x_i, x_{i+1}) = &\exp(\psi_i^u(x_i) + \phi_{\nu_i}(x_i) + \overrightarrow{\Lambda_i} \cdot \overrightarrow{G_i}(x_i) \\ &+ \Delta_{i,i+1}(x_i, x_{i+1})/h \\ &+ \overrightarrow{\Lambda_{i+1}} \cdot \overrightarrow{G_{i+1}}(x_i) + \phi_{\nu_{i+1}}(x_{i+1}) + \psi_{i+1}^d(x_{i+1}))\, \mathrm{d}\overline{\mathbb{P}^h_{i,i+1}}\end{aligned}$$

  From the form of its transition density, and the Markovianity of the reference measure, we deduce that it is Markovian.

- *Its marginal are given by :*

$$\nu_k^\star = \mathbb{P}_k^{h,\star}(x_k) = \exp(\psi_k^u(x_k) + \phi_{\nu_k}(x_k) + \overrightarrow{\Lambda_k} \cdot \overrightarrow{G_k}(x_k) + \psi_k^d(x_k))$$

Proof. In Appendix .1 □

Proposition 4.4. *The quantities ψ_k^u, ψ_k^d can be computed iteratively using the following updates :*

$$\psi_{k+1}^u = \log\left(\int \exp\left(\psi_k^u + \frac{\Delta_{k,k+1}}{h} + \phi_{\nu_k}^n + \overrightarrow{\Lambda_k^n} \cdot \overrightarrow{G_k}\right) \overline{\mathbb{P}^h_{k,k+1}}\, \mathrm{d}x_k\right)$$

$$\psi_{k-1}^d = \log\left(\int \exp\left(\psi_k^d + \frac{\Delta_{k-1,k}}{h} + \phi_{m_k}^n + \overrightarrow{\Lambda_k^n} \cdot \overrightarrow{G_k}\right) \overline{\mathbb{P}^h_{k-1,k}}\, \mathrm{d}x_k\right)$$

with $\psi_0^u = \log \rho_0$ and $\psi_{N_T}^d = 0$.

Proof. Let $i \in \mathcal{K}$ and $x_i \in \mathcal{X}_i$. We can compute the following integral :

$$\begin{aligned}\psi_i^u(x_i) &= \log\left(\int \rho_0 \prod_{k=0}^{i-1} \exp\left(\frac{\Delta_{k,k+1}}{h} + \phi_{\nu_k} + \Lambda_k \cdot \overrightarrow{G_k}\right) \overline{\mathbb{P}^h_{k,k+1}}\, \mathrm{d}x_{[0,i-1]}\right) \\ &= \log\left(\int \left(\int \rho_0 \prod_{k=0}^{i-2} \exp\left(\frac{\Delta_{k,k+1}}{h} + \phi_{\nu_k} + \Lambda_k \cdot \overrightarrow{G_k}\right) \overline{\mathbb{P}^h_{k,k+1}}\, \mathrm{d}x_{[0,i-2]}\right)\right. \\ &\qquad \left.\exp\left(\frac{\Delta_{i-1,i}}{h} + \phi_{m_{i-1}} + \Lambda_{i-1} \cdot \overrightarrow{G_{i-1}}\right) \overline{\mathbb{P}^h_{i-1,i}}\, \mathrm{d}x_{i-1}\right) \\ &= \log\left(\int \exp\left(\psi_{i-1}^u + \frac{\Delta_{i-1,i}}{h} + \phi_{m_{i-1}} + \Lambda_{i-1} \cdot \overrightarrow{G_{i-1}}\right) \overline{\mathbb{P}^h_{i-1,i}}\, \mathrm{d}x_{i-1}\right).\end{aligned}$$

A symmetric reasoning can be done for ψ_i^d. $\square$

5 Sinkhorn algorithm

In order to numerically solve the dual problem, we propose to use an extension of the Sinkhorn algorithm. The Sinkhorn algorithm is a well-known algorithm to solve optimal transport problems. In order to do such numeric computations, we need to discretise our problem in space as well.

5.1 Algorithm

We first describe the principle of the algorithm.

We initialise the dual potentials ϕ_{ν_k}, ϕ_{b_k} and $\overrightarrow{\Lambda_k}$ as 0 which corresponds to being equal to the reference neasure, and we then iteratively the following updates :

$$\begin{cases} \phi_{\nu_k}^{n+1} = \arg\min_\phi \inf_\psi -M_k^\star(-\phi-\psi) - F_k^\star(\psi,-\phi_b) \\ \qquad\qquad - h\,\mathbb{E}_{\overline{\mathbb{P}^h_k}}\Big[\exp(\psi_k^{u,n+1} + \phi + \overrightarrow{\Lambda_k^n}\cdot\overrightarrow{G_k} + \psi_k^{d,n})\Big] \\ \overrightarrow{\Lambda_k^{n+1}} = \arg\min_{\vec{\Lambda}} -C_k^\star(-\Lambda) - h\,\mathbb{E}_{\overline{\mathbb{P}^h_k}}\Big[\exp(\psi_k^{u,n+1} + \phi_{\nu_k}^{n+1} + \vec{\Lambda}\cdot\overrightarrow{G_k} + \psi_k^{d,n})\Big] \\ \phi_{b_k}^{n+1} = \arg\min_\phi \inf_\psi -M_k^\star(-\phi_{\nu_k}-\psi) - F_k^\star(\psi,-\phi) \\ \qquad\qquad - h\,\mathbb{E}_{\overline{\mathbb{P}^h_{k,k+1}}}\Big[\exp(\psi_k^{u,n+1} + \phi_{\nu_k}^{n+1} + \overrightarrow{\Lambda_k^{n+1}}\cdot\overrightarrow{G_k} + \Delta_{i,i+1}(\phi) + \psi_k^{d,n})\Big] \end{cases} \tag{SK1}$$

and for the last marginal :

$$\begin{cases} \phi_{m_{N_T}}^{n+1} = \arg\min_\phi -M_{N_T}^\star(-h\phi) \\ \qquad\qquad - h\,\mathbb{E}_{\overline{\mathbb{P}^h_{N_T}}}\Big[\exp(\psi_{N_T}^{u,n+1} + \phi + \Lambda_{N_T}^n\cdot\overrightarrow{G_{N_T}})\Big] \\ \Lambda_i^{n+1} = \arg\min_\Lambda -C_i^\star(-\Lambda) - h\,\mathbb{E}_{\overline{\mathbb{P}^h_{N_T}}}\Big[\exp(\psi_{N_T}^{u,n+1} + \phi_{m_{N_T}}^{n+1} + \Lambda\cdot\overrightarrow{G_{N_T}})\Big] \end{cases} \tag{SK2}$$

In practice, before each step we compute every downward ψ_k^d, iteratively as described in the previous section. The pseudocode algorithm is described in Algorithm 1.

Algorithm 1: Function Definitions for UpdatePsiUp and UpdatePsiDown

Function UpdatePsiUp($\psi_{k+1}^u, \phi_{\nu_k}^n, \Lambda_k^n, \overline{\mathbb{P}^h_{k,k+1}}, x_k, x_{k+1}, \phi_{b_k}, h$)
begin
  Compute $\Delta_{k,k+1} = (x_{k+1}-x_k)\phi_{b_k}(x_k)$
  return $\log\left(\int \exp\left(\psi_{k+1}^u + \frac{\Delta_{k,k+1}}{h} + \phi_{\nu_k}^n + \Lambda_k^n\cdot\overrightarrow{G_k}\right)\overline{\mathbb{P}^h}_{k,k+1}\,\mathrm{d}x_{k-1}\right)$
Function UpdatePsiDown($\psi_{k-1}^d, \phi_{\nu_k}^n, \Lambda_k^n, \overline{\mathbb{P}^h}_{k-1,k}, x_k, x_{k-1}, \phi_{b_k}, h$)
begin
  Compute $\Delta_{k-1,k} = (x_k-x_{k-1})\phi_{b_k}(x_k)$
  return $\log\left(\int \exp\left(\psi_{k-1}^d + \frac{\Delta_{k-1,k}}{h} + \phi_{m_k}^n + \Lambda_k^n\cdot\overrightarrow{G_k}\right)\overline{\mathbb{P}^h}_{k-1,k}\,\mathrm{d}x_k\right)$

We describe the pseudo-code of the Multi-Marginal Sinkhorn algorithm in Algorithm 2.

Algorithm 2: Sinkhorn algorithm for problem DUAL

Input: Number of timesteps N_T, support of each space $\mathcal{X}_k^{\mathrm{d}x}$
Input: Stopping tolerance ϵ, reference measure $\overline{\mathbb{P}^h}$
Input: Initial potentials $\phi^0_{\nu_k}$, $\phi^0_{b_k}$, Λ^0_i
Result: Numerical solution of problem DUAL

$\psi_0^{u,0} \leftarrow \log \rho_0$;
$\psi_{N_T}^{d,0} \leftarrow 0$;
for $n \leftarrow 0$ **to** N **do**
  for $k \leftarrow N_T - 1$ **to** 0 **do**
    $\psi_k^{d,n} \leftarrow \mathrm{UpdatePsiDown}(\psi_k^{d,n}, \phi^n_{\nu_k}, \phi^n_{b_k}, \Lambda^n_i, \overline{\mathbb{P}^h})$;
  for $k \leftarrow 0$ **to** $N_T - 1$ **do**
    $\phi^{n+1}_{\nu_k} \leftarrow \mathrm{SolveMarginal}(\phi^n_{\nu_k}, \phi^n_{b_k}, \Lambda^n_i, \psi_k^{u,n}, \psi_k^{d,n}, \overline{\mathbb{P}^h})$;
    $\Lambda^{n+1}_k \leftarrow \mathrm{SolvePrices}(\phi^n_{\nu_k}, \phi^n_{b_k}, \Lambda^n_i, \psi_k^{u,n}, \psi_k^{d,n}, \overline{\mathbb{P}^h})$;
    $\phi^{n+1}_{b_k} \leftarrow \mathrm{SolveDriftVol}(\phi^{n+1}_{\nu_k}, \phi^n_{b_k}, \Lambda^{n+1}_i, \psi_k^{u,n}, \psi_k^{d,n}, \overline{\mathbb{P}^h})$;
    $\psi_{k+1}^{u,n+1} \leftarrow \mathrm{UpdatePsiUp}(\psi_{k+1}^{u,n+1}, \phi^{n+1}_{\nu_k}, \phi^{n+1}_{b_k}, \Lambda^{n+1}_i, \overline{\mathbb{P}^h})$;
  $\phi^{n+1}_{m_{N_T}} \leftarrow \mathrm{SolveMarginal}(\phi^n_{m_{N_T}}, \Lambda^n_i, \psi_{N_T}^{u,n}, \psi_{N_T}^{d,n}, \overline{\mathbb{P}^h})$;
  $\Lambda^{n+1}_{N_T} \leftarrow \mathrm{SolvePrices}(\phi^{n+1}_{m_{N_T}}, \Lambda^n_i, \psi_{N_T}^{u,n}, \psi_{N_T}^{d,n}, \overline{\mathbb{P}^h})$;
  $e_{\max} \leftarrow \frac{\|\Phi^{n+1} - \Phi^n\|_\infty}{\|\Phi^n\|_\infty}$;
  if $e_{max} < \epsilon$ **then**
    return Φ, Ψ;
return Φ, Ψ;

The functions SolveMarginal, SolvePrices and SolveDriftVol are functions that solve the minimization problems in (SK1) and (SK2), and might have different implementations depending on the structure of the problem.

In order to provide a numerical implementation of the method, we provide multiple details in the next sections.

5.2 Truncation in space

First, we need to truncate the support so that it has a finite width. We take advantage of the fact that our reference measure $\mathbb{P}^h$ is a diffusion measure and that its density tends quickly to 0 as we move away from the mean. We want to create an interval for each marginal ν_k that contains most of the mass. Because the marginals are not known beforehand, we instead propose to truncate on $\overline{\nu_k}$, the marginals of the reference measure $\overline{\mathbb{P}^h}$.

First, in the case of a reference measure with constant drift and volatility, for each timestep $t_k = k\,h \quad i = 0., \ldots, N_T$, we restrict the computational domain to

$$\overline{\mathcal{X}_k} = [m_k - \delta\, v_k,\, m_k + \delta\, v_k]$$

where :

- $m_k = m_0 + h \sum_{l=0}^{k} \overline{\mu_l}$ is the mean of the k-th marginal, where m_0 is the initial mean of the reference measure.
- δv_i is a multiple of the standard deviation of the reference measure: $v_k = \sqrt{v_0^2 + h \sum_{l=0}^{k} \overline{\sigma_l}^2}$ where v_0 is the standard deviation of the initial marginal of the reference measure. For a sufficiently large δ (in general $\delta = 5$) we expect $\mathbb{P}^h_k$ the solution to be negligible outside of $\mathcal{X}_k$, i.e. the mass transported by the drift (small as we are solving soft martingale constraint problems) and the diffusion further than the enlarged domain is negligible.

When the reference measure has a non-constant drift and volatility, we can use the same truncation as above, but with the maximum in space of the drift and volatility.

5.3 Discretisation

On this compact supports, we discretize the potentials on a regular grid. We can hence represent the potentials ϕ_{ν_k}, ϕ_{b_k}, as vectors $\phi_{b_k}^{\mathrm{d}x}$, of size $N_{\mathcal{X}_k}$ and $N_{\mathcal{X}_k} \times d$ respectively. We can also represent the quantities ψ_k^u, ψ_k^d as vectors $\psi_k^{u,\mathrm{d}x}$, $\psi_k^{d,\mathrm{d}x}$ of size $N_{\mathcal{X}_k}$. The integrals can be replaced by discrete sums over the grid $\mathcal{X}_k^{\mathrm{d}x}$. For clarity, we will drop the superscript $\mathrm{d}x$ in the following and keep the notations with integrals as they can be used interchangeably.

An important parameter is the space discretisation step $\mathrm{d}x$. We want to choose it so that the transition law $\overline{\mathbb{P}_{i,i+1}}$ is non-zero on enough points. To ensure this, we want :

$$\mathrm{d}x < K\overline{\sigma}_i\sqrt{h}$$

which links $N_{\mathcal{X}_k}$ to N_T, for a constant K to be determined which ensure the minimum of points in the width of the kernel.

5.4 Multiscale Strategy

Because the complexity of algorithm 2 scales with the number of timesteps N_T, it can be interesting to start with a coarse time discretisation and to refine it iteratively.

One way to do so is presented in [8] and consists in interpolating the potentials ϕ_{b_k} and using them as initialisation for the next level of discretisation.

Here, we propose an alternative method, which consists in using the result of optimization at a coarse scale as the reference measure for a finer scale. This is done as such :

1. Solve the problem at a scale $h = \frac{T}{N_T - 1}$ with N_T timesteps.
2. Interpolate the obtained volatilities $\sigma_k^2(x,t)$ for a new scale $h' = \frac{T}{N_T' - 1}$ with $N_T' > N_T$ timesteps. Multiple interpolation technique can be used, in our case, we use Unbalanced Optimal Transport Barycenters. We then obtain a new reference measure $\overline{\mathbb{P}^{h'}}$.
3. $h = h'$, $N_T = N_T'$, and return at step 1 if the desired scale is not reached.

We can, for instance, start with N_T equal to the number of calibration steps.

5.5 Anderson Acceleration

Entropic Sinkhorn iterations are known to converge slowly when the regularization parameter $\epsilon \to 0$. Since one Sinkhorn iteration $\Phi^{k+1} = s(\Phi^k)$ is a fixed-point iteration, in order to accelerate convergence, we propose to use Anderson acceleration [2] [26]. We consider our variable Φ as a variable of dimension N_Φ and denote $g(\Phi) = s(\Phi) - \Phi$ is the residual of the Sinkhorn iteration. In particular, given a vector of m_k iterates $\Phi_{m_k}^k = (\Phi^{k-m_k+1}, \dots, \Phi^k)$, we denote $G = \{g(\Phi^k)\}_k = \{s(\Phi^k) - \Phi^k\}_k$ the vector of their residuals. We compute the next iterate Φ^{k+1} as a combination of the m_k previous iterates as a solution to the following problem :

$$\min_{\alpha \in \Delta(m_k)} ||G\alpha||_2 \tag{AA}$$

where $\Delta(m_k) = \{v \in \mathbb{R}^{m_k}, \sum v_i = 1\}$. The idea is that by first-order approximation, $||g(\Phi_{m_k}^k \alpha)||_2 \approx ||G\alpha||_2$. The new iterate is then given by $\Phi^{k+1} = s(\Phi_{m_k}^k)\alpha \approx s(\Phi_{m_k}^k \alpha)$ by first order approximation.

The problem (AA) is reformulated with the following linear least square problem as suggested in [2] [26]:

$$\min_{\gamma \in \mathbb{R}^{m_k - 1}} ||\mathcal{G}\gamma - g_k||_2 \tag{AA-LS}$$

where $\mathcal{G} = \begin{pmatrix} g_k - g_{k-1} & \dots & g_{k-m_k+1} - g_{k-m_k} \end{pmatrix} \in \mathbb{R}^{N_\Phi \times (m_k-1)}$ with $g_k = g(\Phi^k) \in \mathbb{R}^{N_\Phi}$. This provides a solution to (AA) as : $\alpha = \begin{pmatrix} 1 - \gamma_{m_k-1} & \gamma_{m_k-1} - \gamma_{m_k-2} & \dots & \gamma_2 - \gamma_1 & \gamma_1 \end{pmatrix}$.

We can then rewrite the iterate Φ^{k+1} as :

$$\Phi^{k+1} = s(\Phi^k_{m_k})\alpha = \sum_{i=k-m_k}^{k} \alpha_i s(\Phi^i) = s(\Phi^k) + \sum_{i=k-m_k}^{k-1} \gamma_i (s(\Phi^{i+1}) - s(\Phi^i))$$

Noting $d\Phi^k = \begin{pmatrix} \Phi^k - \Phi^{k-1} & \dots & \Phi^{k-m_k+1} - \Phi^{k-m_k} \end{pmatrix} \in \mathbb{R}^{N_\Phi \times (m_k-1)}$ the matrix of iterate differences, we can rewrite the iterate Φ^{k+1} as :

$$\begin{aligned} \Phi^{k+1} &= s(\Phi^k) + d\Phi^k \gamma \\ &= \Phi^k + g_k - (\mathcal{G} + d\Phi^k)\gamma \end{aligned} \tag{AA-iter}$$

The problem (AA-LS) is a linear least square problem that can be solved efficiently. We solve it by solving the linear system $\mathcal{G}^T\mathcal{G}\gamma = \mathcal{G}^T g_k$ since the Gram matrix $\mathcal{G}^T\mathcal{G}$ can be easily computed. To avoid cases where $\mathcal{G}^T\mathcal{G}$ is ill-conditioned, we add a ridge regularization term ϵI to the Gram matrix, where ϵ is a small positive constant. When $\epsilon \to 0$, the solution of the linear system converges to the solution of (AA-LS), while when $\epsilon \to \infty$, the solution of the linear system converges to 0 and hence the steps become simple Sinkhorn iterations.

To avoid convergence problems, we also employ a safeguarding mechanism as described, for instance, in [15] to enhance the stability and performance of the Anderson acceleration technique. At each iteration k, an accelerated candidate $\Phi^{k,\text{acc}}$ is generated using iteration (AA-iter). To assess the quality of this candidate, we compare its residual $g(\Phi^{k,\text{acc}})$ with the residual of the last accepted iterate $g(\Phi^k)$. Specifically, we impose a safeguarding condition:

$$\|g(\Phi^{k+1,\text{acc}})\|_2 \leq \tau \|g(\Phi^k)\|_2 \tag{14}$$

Here, τ is a tolerance parameter. If the condition is met, the accelerated candidate $\Phi^{k,\text{acc}}$ is accepted for the next iteration. Otherwise, the candidate is declined, and the algorithm proceeds without acceleration for that step.

Finally, similarly to what is done for Sinkhorn algorithm, we stop the algorithm when the l^∞ norm of the residual is below a given tolerance ϵ_{stop}.

6 Application: Local volatility calibration

We want to minimize over martingale positive process, hence we are interested in the process of the form $X_t = \log S_t$ where S_t is a martingale diffusion process:

$$\frac{\mathrm{d}S_t}{S_t} = \sigma(S_t, t), \mathrm{d}W_t.$$

Applying Ito's lemma, in terms of $X_t = \log S_t$, we obtain the following SDE:

$$\mathrm{d}X_t = -\frac{1}{2}\sigma^2(X_t, t)\ \mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}W_t.$$

As noted in article [19], in terms of the characteristics of the SDE μ and σ^2, this reads:

$$2\mu(x, kh) = -\sigma^2(x, kh),$$

Using our defined variables β_k and α_k in Section 3.2 equations (4) (5), this can be written as

$$2\beta_k = -\alpha_k.$$

Notice however that both equations are not exactly equivalent due to the first-order term $\mu^2\, h$ appearing in equation (5).

Instead of the variables β and α, we will use the variable b_k corresponding to the choice $B(X,Y) = 1 - e^{Y-X}$. We compute b_k :

$$\begin{aligned} b_k(x) &= \frac{1}{h}\,\mathbb{E}\big[1 - e^{X_{k+1}-X_k} \;\big|\; X_k = x\big] \\ &= \frac{1}{h}\frac{1}{e^{X_k}}\,\mathbb{E}\big[e^{X_k} - e^{X_{k+1}} \;\big|\; X_k = x\big] \\ &= \frac{1}{h}\frac{1}{S_k}\,\mathbb{E}[S_k - S_{k+1} \mid X_k = x], \end{aligned}$$

which directly corresponds to an at martingale constraint on the process S_t. This is the moment we will consider in this application and the above strategy method is easily applied.

We penalize this variable with $\mathcal{F}$ a soft penalization $F = c_{\text{mart}}\|\cdot\|_{L^2}$ with a constant penalization parameter c_{mart}. This is to overcome the convex ordering problem mentioned in [1] when working on a discretized grid: the discrete approximation on a grid of continuous measures in convex order might not be in convex order.

For the price constraints, let $c_i \in \mathbb{R}^+$ be an observed price, we use the soft constraint C_i a convex function with minima in c_i, for instance, $C_i = \frac{1}{2}(\cdot - c_i)^2$. We use the payoff function $g_i(x) = \max(0, e^x - K_i)$ for a call option with strike K_i, and $g_i(x) = \max(0, K_i - e^x)$ for a put option with strike K_i.

For the first marginal constraint, we propose using a hard constraint $M_0 = \iota_{\mu_0}$ whose dual is $\langle \phi_{\nu_0}, \mu_0 \rangle$, with $\mu_0 = \delta_{\log S_0}$.

Finally, we obtain the following problem:

$$\begin{aligned} \mathcal{V} = \inf_{\mathbb{P}^h} \sum h\,\mathbb{E}_{\nu_k}[F(b_k)] + h\,\mathrm{KL}(\mathbb{P}^h|\overline{\mathbb{P}^h}) \\ + \iota_{\mu_0}\left(\nu_0\right) + \sum_{i=1}^{n} C_i\left(g_i\right), \end{aligned}$$

and in its dual form :

$$\begin{aligned} \mathcal{D} = \sup_{\phi_{\nu_0}, \phi_{b_k}, \lambda_{g_i}} J(\phi) \overset{\text{def}}{=} \mathbb{E}_{\mu_0}[\phi_{\nu_0}] + \sum_{i=0}^{N_C} C_i^\star(\lambda_{g_i}) \\ + h\,\mathbb{E}_{\overline{\mathbb{P}^h}}\left[\exp\left(\frac{\Delta(\phi_m, \phi_b, \lambda_g)}{h}\right)\right], \end{aligned}$$

under the constraints $\phi_{\nu_k} = hH^\star(-\phi_{b_k})$, $\forall k \notin \{0, N_T\}$, $\phi_{m_{N_T}} = 0$.

The first equation of the Sinkhorn system (SK1) becomes :

$$\begin{aligned} \mu_0 &= \int \mathbb{P}^h(x_0, \mathrm{d}x_{-0}) \\ &= \exp(\psi_0^u(x_0) + \phi_{\nu_0}(x_0)/h + \psi_0^d(x_0)), \end{aligned}$$

which can be solved using classical Sinkhorn iteration as :

$$\phi_{\nu_0} = -h(\psi_0^u + \psi_0^d - \log(\mu_0)).$$

We derivate our functional J with respect to ϕ_{b_k} to obtain the following optimality equation :

$$0 = \mathbb{E}_{\mathbb{P}^h_{k,k+1}}\Big[(1 - e^{X_{k+1}-X_k}) + hH^{\star'}(-\phi_{\nu_k}(x_k))\Big]$$

The expression of the Hessian is simple as well, as it is diagonal.

In a similar fashion, we can obtain optimality equations for Λ_i :

$$0 = -C_i^{\star'}(-\lambda_i) + \mathbb{E}_{\mathbb{P}_k^h}[g_i^{\mathbb{P}^h}], \forall i \in \{0, N_T - 1\}$$

We can compute the Hessian which is this time of size $|Ic_k|$.

To optimize on these two variables, we can perform a Newton method.

The calculation of the derivatives on ϕ_{b_k} is of complexity $O(N_{\mathcal{X}_k} \times N_{\mathcal{X}_{k+1}})$, while the other derivatives are of complexity $O(N_{\mathcal{X}_k})$. The computation of ψ^u and ψ^d are of complexity $O(N_{\mathcal{X}_k} \times N_{\mathcal{X}_{k+1}})$. Hence, the complexity of the whole algorithm is of complexity $O(N_T \times \max N_{\mathcal{X}_k}^2)$.

At the coarsest scale, the reference measure is chosen to match the ATM price at the calibration time. Hence, given the ATM volatility $\sigma_{\text{BS}}(F, \tau_i)$ for all calibration time (τ_i), where F is the forward price, we set our reference volatility to be $\overline{\sigma}_0^2(x) = \sigma_{\text{BS}}^2(F, \tau_0)\tau_0$ and $\overline{\sigma}_i^2(x) = \sigma_{\text{BS}}^2(F, \tau_i)\tau_i - \overline{\sigma}_{i-1}^2(x)$ for all $i \in \{1, \dots, \}$.

6.1 Numerical results

Here, we first generate prices using a parametric local volatility surface. The local volatility surface that we choose is the SSVI surface as presented in [16]. We choose the at-the-money implied total variance for the money to be $\theta_t = 0.04t$. We choose a power-law parameterization of the function ϕ described in [16] as $\phi(\theta) = \eta\theta^{-\lambda}$. The at-the-money total implied variance is then

$$\sigma_{\text{BS}}^2(k, T) = \frac{\theta_t}{2}\left(1 + \rho\phi(\theta_t)k + \sqrt{(\phi(\theta_t)k + \rho)^2 + (1 - \rho^2)}\right),$$

where k is the log-moneyness $\log(K/F)$. The parameters are chosen as $\eta = 1.6$, $\lambda = 0.4$ and $\rho = -0.15$. The resulting surface is shown in Figure 1. We produce prices using the Black-Scholes formula.

We select five times in which we will calibrate the model on generated prices : $t \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$. On time τ_i, we select the calls with strikes $K \in \{S_0{+}1{+}4k_i\}$ and the puts with strikes $K \in \{S_0{-}1{-}4k_i\}$ for $k_i \in \{0, 1, \dots, N_{C,i}\}$, with $N_C = (5, 7, 9, 10, 12)$. We calibrate less points for the earlier maturities as mass is almost nonexistent far from the at-the-money price at these maturities.

We perform the multiscale strategy described in Section 5.4 with up to $N_T = 81$.

We choose the constant $c_{\text{mart}} = 1 \times 10^4$ as the penalization term for the martingale constraint.

We implement the algorithm in Python using the PyKeops library.

The program runs in approximately 10 minutes on a V100 GPU with 24GB of GDDR5 memory and an Intel Xeon 5217 8 core CPU with 192GB or DDR4 RAM.

We show the convergence curves at the last scale $N_T = 81$ in Figure 2. Figure 2a shows the L^2 norm of the relative iterate errors $\frac{\|\Phi^{k+1} - \Phi^k\|}{\|\Phi^k\|}$ at each iteration. Figure 2b shows the L^2 norm of the martingale error at each iteration. Figure 2c shows the L^2 norm of the price errors $\sum_i \|c_i - \mathbb{E}(G(X_{\tau_i}))\|^2$ by iteration for every scale. Different scales are separated by a dashed line.

Finally, Figure 3 shows the calibration results at each time. For each of the calibration times, we show the reference implied volatility, the calibrated implied volatility, and the implied volatility generated by the forward diffusion process with the same number of timesteps and the volatility of the solution.

References

[1] Aurélien Alfonsi, Jacopo Corbetta, and Benjamin Jourdain. *Sampling of Probability Measures in the Convex Order and Approximation of Martingale Optimal Transport Problems.* SSRN Scholarly Paper. Rochester, NY, Nov. 2017.

[2] Donald G. Anderson. "Iterative Procedures for Nonlinear Integral Equations". In: *J. ACM* 12.4 (1965), 547–560. ISSN: 0004-5411.

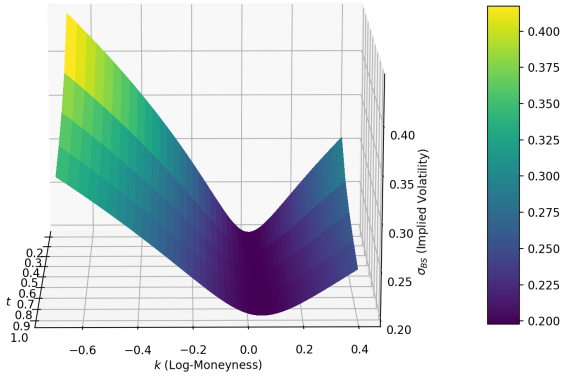


Figure 1: Generating model implied volatility

[3] Julio Backhoff-Veraguas and Clara Unterberger. *On the specific relative entropy between martingale diffusions on the line*. 2023.

[4] Jean-David Benamou. "Optimal transportation, modelling and numerical simulation". In: *Acta Numerica* 30 (2021), pp. 249–325.

[5] Jean-David Benamou and Yann Brenier. "A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem". In: *Numer. Math.* 84 (2000), pp. 375–393.

[6] Jean-David Benamou, Guillaume Carlier, and Luca Nenna. "Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm". In: *Numer. Math.* 142 (2019), pp. 33–54.

[7] Jean-David Benamou et al. "An entropy minimization approach to second-order variational mean-field games". In: *Math. Models Methods Appl. Sci.* 29 (2019), pp. 1553–1583.

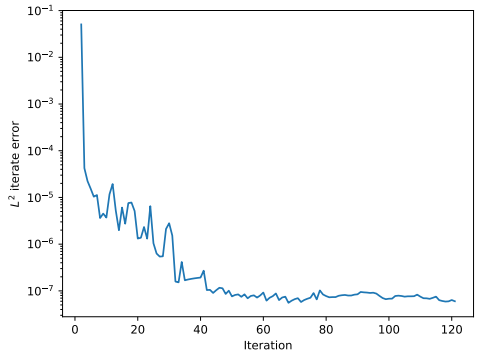
[8] Jean-David Benamou et al. "Entropic Martingale Optimal Transport". In: (2024).

[9] Jean-David Benamou et al. "Entropic optimal martingale transport". in preparation. 2024.

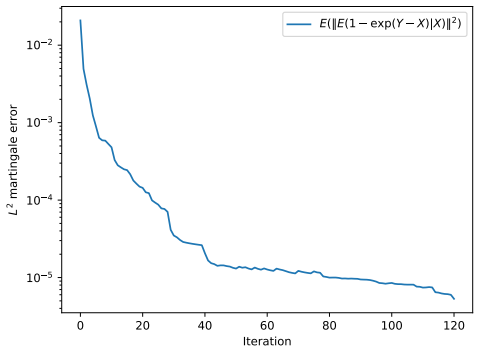
[10] Guillaume Carlier. "On the Linear Convergence of the Multimarginal Sinkhorn Algorithm". In: *SIAM Journal on Optimization* 32.2 (2022), pp. 786–794.

[11] Guillaume Carlier et al. "Convergence of Entropic Schemes for Optimal Transport and Gradient Flows". In: *SIAM Journal on Mathematical Analysis* 49.2 (2017), pp. 1385–1418.

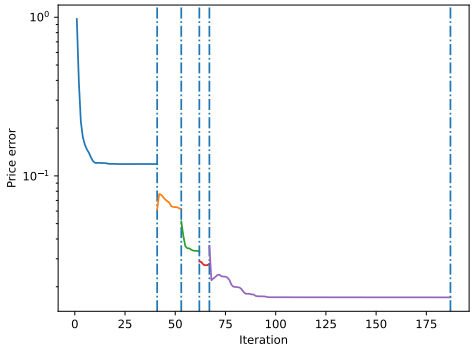
[12] Simone Di Marino and Augusto Gerolin. "An Optimal Transport Approach for the Schrödinger Bridge Problem and Convergence of Sinkhorn Algorithm". In: *Journal of Scientific Computing* 85 (Nov. 2020), p. 27.



(a) L^2 norm of iterate errors



(b) L^2 norm of the martingale error



(c) L^2 of the price errors

Figure 2: Convergence curves

[13] Hans Föllmer. "Doob Decomposition, Dirichlet Processes, and Entropies on Wiener Space". In: *Festschrift in honour of Masatoshi Fukushima's Beiju*. Springer Proceedings in Mathematics & Statistics. 2022.

[14] Nina Gantert. *Some large deviations of Brownian motion*. 1991.

[15] Michael Garstka, Mark Cannon, and Paul Goulart. *Safeguarded Anderson acceleration for parametric nonexpansive operators*. 2022.

[16] Jim Gatheral and Antoine Jacquier. "Arbitrage-free SVI volatility surfaces". In: *SSRN Electronic Journal* (2012).

[17] Ivan Guo, Gregoire Loeper, and Shiyi Wang. "Calibration of Local-Stochastic Volatility Models by Optimal Transport". In: arXiv:1906.06478 (July 2021).

[18] Ivan Guo, Grégoire Loeper, and Shiyi Wang. "Calibration of local-stochastic volatility models by optimal transport". In: *Mathematical Finance* 32.1 (2022), pp. 46–77.

[19] Ivan Guo et al. "Optimal transport for model calibration". In: *Risk* (2022).

[20] Julien Guyon. *Dispersion-Constrained Martingale Schrödinger Bridges: Joint Entropic Calibration of Stochastic Volatility Models to S&P 500 and VIX Smiles*. en. 2022.

[21] Christian Léonard. "A survey of the Schrödinger problem and some of its connections with optimal transport". In: *Discrete Contin. Dyn. Syst.* 34 (2014), pp. 1533–1574.

[22] Gregoire Loeper. "Option Pricing with Linear Market Impact and Non-Linear Black-Scholes Equations". In: *The Annals of Applied Probability, 2018* (2016).

[23] Nicolas Papadakis, Gabriel Peyré, and Edouard Oudet. "Optimal Transport with Proximal Splitting". In: *SIAM J. Imaging Sci.* 7 (2014), pp. 212–238. ISSN: 1936-4954.

[24] G. Peyré and M. Cuturi. "Computational optimal transport". In: *Found. Trends Mach. Learning* 11 (Mar. 2019), pp. 355–607.

[25] Xiaolu Tan and Nizar Touzi. "Optimal transportation under controlled stochastic dynamics". In: *The Annals of Probability* 41.5 (2013).

[26] Homer F. Walker and Peng Ni. "Anderson Acceleration for Fixed-Point Iterations". In: *SIAM Journal on Numerical Analysis* 49.4 (2011), pp. 1715–1735.

.1 Proof of Proposition 4.3

First, we separate the sum of λ_{g_i} per timesteps using the values defined above :

$$\sum_{i=0}^{N_C} \lambda_{g_i} G_i(x_i) = \sum_{k=0}^{N_T} \sum_{i=0}^{N_C} \lambda_{g_i} \mathbb{1}_{\tau_i = k} G_i(x_i) = \sum_{k=0}^{N_T} \vec{\Lambda}_k \cdot \overrightarrow{G_k}(x_k)$$

We can rewrite the operator Δ as a sum :

$$\begin{aligned}\Delta(\phi_m, \phi_b, \lambda_g) = \sum_{k=0}^{N_T-1} & \Delta_{k,k+1}(x_k, x_{k+1}) + \overrightarrow{\Lambda_k} \cdot \overrightarrow{G_k} + \phi_{m_k}(x_k) \\ & + \phi_{m_{N_T}}(x_{N_T}) + \overrightarrow{\Lambda_{N_T}} \cdot \overrightarrow{G_{N_T}}(x_{N_T})\end{aligned}$$

where only consecutive timesteps are grouped together. In particular, for a given k, we can separate this sum into three parts :

$$\begin{aligned}\Delta(\phi_m, \phi_p, \phi_d, \lambda_g) = & \Delta_{k,k+1}(x_k, x_{k+1}) + \overrightarrow{\Lambda_k} \cdot \overrightarrow{G_k}(x_k) + \phi_{m_k}(x_k) \\ & + \overrightarrow{\Lambda_{k+1}} \cdot \overrightarrow{G_{k+1}}(x_{k+1}) + \phi_{m_{k+1}}(x_{k+1}) \\ & + \Delta_k^u(x_k) + \Delta_{k+1}^d(x_{k+1})\end{aligned}$$

where Δ_k^u and Δ_k^d are given by :

$$\begin{aligned}\Delta_k^u(x_k) &= \sum_{i=0}^{k-1} \Delta_{i,i+1}(x_i, x_{i+1}) + \vec{\Lambda_i} \cdot \overrightarrow{G_i}(x_i) + \phi_{m_i}(x_i) \\ \Delta_k^d(x_k) &= \sum_{i=k}^{N_T-1} \Delta_{i,i+1}(x_i, x_{i+1}) + \overrightarrow{\Lambda_{i+1}} \cdot \overrightarrow{G_{i+1}}(x_i) + \phi_{m_{i+1}}(x_{i+1}).\end{aligned}$$

We further note :

$$\begin{aligned}\overline{\Delta}_{k,k+1}(x_k, x_{k+1}) = & \Delta_{k,k+1}(x_k, x_{k+1}) + \overrightarrow{\Lambda_k} \cdot \overrightarrow{G_k}(x_k) + \phi_{m_k}(x_k) \\ & + \overrightarrow{\Lambda_{k+1}} \cdot \overrightarrow{G_{k+1}}(x_{k+1}) + \phi_{m_{k+1}}(x_{k+1})\end{aligned}$$

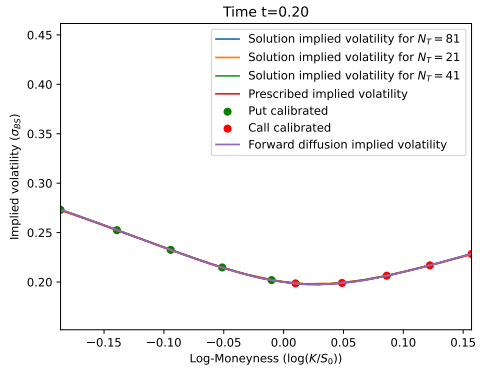
for simplicity.

Given that $\overline{\mathbb{P}^h}$ is separable in the same fashion, we can compute the joint probability between steps k and $k+1$ as :

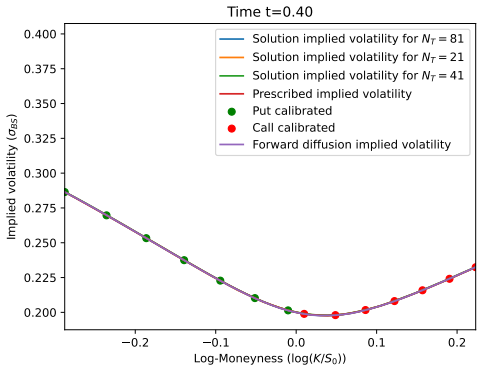
$$\begin{aligned}
\mathbb{P}^{h,\star}_{k,k+1}(x_k, x_{k+1}) &= \int \mathbb{P}^{h,\star}(dx_{[0,k-1]}, x_k, x_{k+1}, dx_{[k+2,N_T]}) \\
&= \int e^{(\Delta^u_k + \overline{\Delta}_{k,k+1} + \Delta^d_{k+1})/h} \rho_0 \prod_{i=0}^{N_T} \overline{\mathbb{P}^h_{i,i+1}} dx_{[0,k-1]} dx_{[k+2,N_T]} \\
&= \left(\int e^{\Delta^u_k/h} \rho_0 \prod_{i=0}^{k-1} \overline{\mathbb{P}^h_{i,i+1}} dx_{[0,k-1]} \right) \\
&\quad \times e^{\overline{\Delta}_{k,k+1}/h} \overline{\mathbb{P}^h}_{k,k+1} \\
&\quad \times \left(\int e^{\Delta^d_{k+1}/h} \prod_{i=k+1}^{N_T} \overline{\mathbb{P}^h_{i,i+1}} dx_{[k+2,N_T]} \right) \\
&= \exp(\psi^u_k(x_k) + \overline{\Delta}_{k,k+1}(x_k, x_{k+1})/h + \psi^d_{k+1}(x_{k+1})) \overline{\mathbb{P}^h_{k,k+1}}(x_k, x_{k+1})
\end{aligned}$$

Similarly as in the previous proof, we can compute the marginal as:

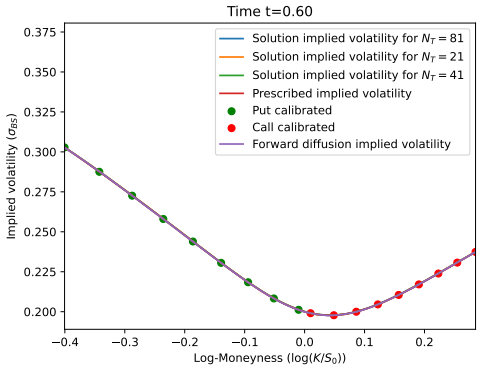
$$\begin{aligned}
\mathbb{P}^{h,\star}_k(x_k) &= \int \mathbb{P}^{h,\star}(dx_{-k}, x_k) \\
&= \int e^{(\Delta^u_k + \phi_{m_k}(x_k) + \overrightarrow{\Lambda_k} \cdot \overrightarrow{G_k}(x_k) + \Delta^d_k)/h} \rho_0 \prod_{i=0}^{N_T} \overline{\mathbb{P}^h_{i,i+1}} dx_{[0,k-1]} dx_{[k+2,N_T]} \\
&= \left(\int e^{\Delta^u_k/h} \rho_0 \prod_{i=0}^{k-1} \overline{\mathbb{P}^h_{i,i+1}} dx_{[0,k-1]} \right) \\
&\quad \times e^{\phi_{m_k}(x_k) + \overrightarrow{\Lambda_k} \cdot \overrightarrow{G_k}(x_k)/h} \\
&\quad \times \left(\int e^{\Delta^d_k/h} \prod_{i=k}^{N_T} \overline{\mathbb{P}^h_{i,i+1}} dx_{[k+1,N_T]} \right) \\
&= \exp(\psi^u_k(x_k) + \phi_{m_k}(x_k) + \overrightarrow{\Lambda_k} \cdot \overrightarrow{G_k}(x_k) + \psi^d_k(x_k))
\end{aligned}$$



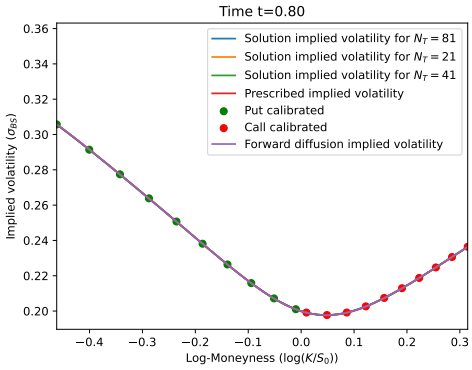
(a) Calibration at time 0.2.



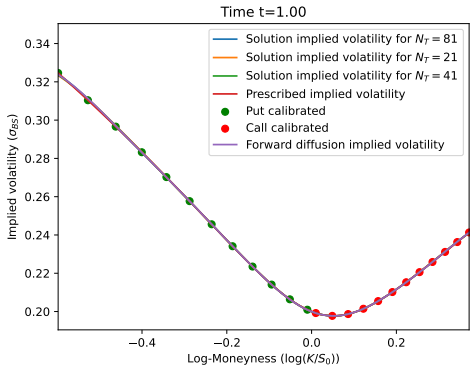
(b) Calibration at time 0.4.



(c) Calibration at time 0.6.



(d) Calibration at time 0.8.



(e) Calibration at time 1.

Figure 3: Calibration results.