Solving the FX cross-smiles problem - rate of convergence for Sinkhorn marginals, and the finite-option case

Martin Forde

23rd Nov 2025

Abstract

We adapt the arguments in [Guy20],[Nutz22] to the problem of constructing a joint law μ^* for two FX rates X and Y consistent with observed European option prices on all three cross-rates. The usual Gibbs-type exponential ansatz for μ^* leads to three coupled integral equations for the Schrödinger potentials $u(x)$, $v(y)$, and $yw(\frac{x}{y})$, and we prove $O(\frac{1}{\sqrt{n}})$ convergence for the three marginals (in the total variation metric) for the running average of the Sinkhorn iterates, assuming an admissible law exists (typically using ≈ 50 iterations in practice). The primary application here is risk-neutral pricing of e.g. Basket, Quanto, Best-of or other varietals of Rainbow options on X and Y so as to be consistent with all three smiles (or barrier options if we use the continuous-time solution in Section 2.8), and risk-neutral pricing obviates the need to estimate the $\mathbb{P}$-measure which is needed for indifference pricing[1]. We also consider the finite-option case via the usual duality for one-period markets between the (primal) entropy minimization problem over calibrated measures and the (dual) exponential utility maximization problem à la Föllmer&Schied[FS04] and discuss convergence for Newton's method in this setup. The primal and dual here can also very often be solved directly (after appropriate discretization) using CVXPY in Python with the MOSEK interior point solver[2](even with a large number of options).[3] [4]

1 Background

Regularized optimal transport methods have been a game changer for exotic option calibration problems in recent years, because it allows the inner inf in the minmax problem which arises from dualization to be computed explicitly (see e.g. pg 8 in [Guy20] or [GLOW22]). For regularized problems in the discrete-time setting, if we use an entropic penalty then we can apply the Sinkhorn fixed point scheme (with provable convergence in certain standard cases, see e.g.[Nutz22]), for which the most notable practical application has been a discrete-time solution to the SPX-VIX calibration problem in [Guy20]. In particular, [Guy20] uses the Donsker-Varadhan/Gibbs variational formula $\inf_{\mu\in\mathcal{P}}\{H(\mu|\bar{\mu})-\mathbb{E}^{\mu}(X)\}=-\log\mathbb{E}^{\bar{\mu}}(e^X)$ (see e.g. Eq 1.1 in Dupuis et al.[ACD15]) to simplify the $\sup\inf$ which arises from dualization.

In the continuous-time setting, we can also solve calibration problems using a Markov local and/or stochastic volatility models with finite tradable European and/or VIX or barrier options using more general penalty terms, by numerically solving a HJB equation which emerges from dualization (see e.g. [GLOW22] and further discussion below), although rigorous duality results with viscosity solutions are more cumbersome to establish.

In principle, a bivariate [Carr09] Local Variance Gamma (LVG) model (see also [CN17]) can also give an exact fit to three cross-smiles, but in practice the implied local correlation function $\rho_{loc}(x,y):=(\sigma_X(x)^2+\sigma_Y(y)^2-\sigma_Z(x/y)^2)/(2\sigma_X(x)\sigma_Y(y))$ that comes out from the three smiles typically falls outside $[-1,1]$ (similarly a two-maturity standard Carr LVG model can be fitted to a single SPX and VIX smile with the same maturity, but in practice the vol function for the 2nd epoch typically has a singularity, and a simpler (albeit somewhat unrealistic) model that achieves the same goal is described in Eq 6.3 in [Guy20]).

Guo et al. [GLW22],[GLOW22] (see also [HL19] and [Guy22]) show how to construct a generalized local/stochastic volatility model consistent with a finite number of European tradable options at multiple maturities by minimizing a cost function over calibrated models which penalizes deviations from a standard reference model (e.g. Black-Scholes or Heston), and then re-casting the problem via dualization as an (unconstrained) minmax problem in terms of a non-linear HJB equation (so the cost function effectively regularizes the problem). If options at multiple maturities are used in the calibration set, the HJB equation unfortunately also includes Dirac source terms (but this can be avoided using a nested PDE), and this method is extended to include VIX options in section 3.3 in [GLOW22], by re-expressing V_t for the reference model in terms of $\mathbb{E}(\int_t^T\sigma_s^2ds|\mathcal{F}_t)$ (this analysis can be simplified using that VIX_t^2 is just an affine function of V_t when the drift of V under the reference model has an affine drift). This approach is mathematically rich and exciting albeit numerically intensive since it requires numerically solving a non-linear HJB equation using e.g. fiddly implicit policy-iteration finite difference schemes, and then maximizing over the option

[1]recall that we cannot consistently estimate drift parameters for a semimartingale on a finite interval even with continuous observation.
[2]We thank Joseph Sullivan and Zhuoran Li for demonstrating this.
[3]the number of variables m for the dual is significantly smaller than the primal since m is linear in the number of constraints.
[4]We thank David Hobson and Amir Dembo for sharing their insights.

weights vector. If path-dependent options are included in the calibration set we also have the issue that the prices may admit arbitrage.

[JLO23] construct the two-maturity Bass martingale M, by considering a standard 1d Brownian motion B with initial distribution α_0 with $M_t = \mathbb{E}(F_1(B_T)|\mathcal{F}_t^B)$. If μ_t denotes the density of M_t, then from elementary arguments we can easily check that $\mu_t = (R_{1-t} * F)_{\#}(R_t * \alpha_0)$ (where R_t is the Brownian density and $\#$ denotes the pushforward). If we take μ_0 and μ_1 as given target laws and evaluate the equation for μ_t at $t=0$ and $t=1$, we obtain two coupled integral equations for the unknown α_0 and F_1. This naturally suggests an alternating Sinkhorn-type scheme where each successive iterate improves the dual objective function so we have convergence (see section 3.2 of [JLO23]) which they refer to as the *measure preserving martingale Sinkhorn* (MPMS) algorithm. On page 6 with a few more lines they verify this formulation is equivalent to the fixed-point integral equation (and associated fixed point scheme) in [CL21] for the cdf G_{α_0} of α_0 given by $G_{\alpha_0} = G_{\mu_0} \circ (R_1 * F) = G_{\mu_0} \circ (R_1 * (G_{\mu_1}^{-1} \circ (R_1 * G_{\alpha_0})))$, where $G_\nu(.)$ denotes the cdf of a probability measure ν. Acciaio et al.[AMP25] prove a contraction property for this equation (under the ∞-Wasserstein distance $W_\infty(\mu,\nu) = \lim_{p\to\infty} W_p(\mu,\nu) = \|G_\mu^{-1} - G_\nu^{-1}\|_{L^\infty(0,1)}$) and establish linear convergence for the fixed-point scheme when μ_0 and μ_1 are irreducible, and convergence of the multi-maturity Bass martingale to the Dupire local volatility model.

Recall that M in the preceding paragraph is also the solution to the martingale Benamou-Brenier problem (mBB): $\inf_{M_0\sim\mu, M_T\sim\nu, M_t = M_0 + \int_0^t \sigma_s dW_s} \mathbb{E}(\int_0^T (\sigma_t - 1)^2)dt$ ($T=1$ above), hence the sobriquet “stretched Brownian motion”, or equivalently maximizes $\mathbb{E}(\int_0^T |\sigma_t| dt)$ and $\mathbb{E}(M_T W_T)$ over the same set. Using the HJB/dualization approach we can obtain the dual representation for this problem in Eq 2 in [JLO23] using the Fenchel-Moreau theorem (see also [Loep25]) which can be re-written in a (possibly) more intuitive/familiar form as $\sup_\varphi(-\int \varphi(y)\mu_T(dy) + \int u^\varphi(x,0)\mu_0(dx))$ where u^φ satisfies the HJB equation $u_t + \inf_\sigma(\frac{1}{2}\sigma^2 u_{xx} + \frac{1}{2}(\sigma^2-1)) = 0$ (if a classical solution exists) with $u(x,T) = \varphi$, which is the same as the usual form for [GLW22],[GLOW22]-type problems with a single terminal target law μ_T except now we integrate $u^\varphi(x,0)$ over μ_0 because the two-maturity problem becomes a 1-maturity problem with a random starting point. [BST23] give an alternate more explicit formulation for the dual here as $\inf_{\psi\in L^1(\nu), \psi \text{ convex}}(\int \psi d\nu - \int (\psi^* * \gamma_1)^* d\mu)$ ($\nu = \mu_T$ above) which we can prove just using basic properties of derivatives of Legendre transforms, but the [JLO23] formula is more easily adapted to the more general mBB(H) problem of minimizing $\mathbb{E}(\int_0^T H(\sigma_t^2)dt)$ and more general diffusion processes.

Note if we solve an entropically regularized M.O.T. problem of the form $\inf_{\mu\in\mathcal{M}(\mu_1,\mu_2)}(H(\mu|\bar{\mu}) + \mathbb{E}^\mu(c(X,Y))$ and similarly for (μ_2,μ_3), with μ_1,μ_2,μ_3 in convex order, then concatenating the two conditional laws for $Y|X$ yields a (Markov) martingale measure consistent with all three target laws (see [Bour25], and related discussions in [DHL19],[EKK25]), which can then be used to price path-dependent options (of course one can also use the Carr Local Variance Gamma model to do this, cf. [Carr09],[CN17]).

[BG24] use an alternating Newton-Sinkhorn algorithm for the SPX-VIX calibration problem; specifically, for each cycle they first run a multi-dimensional Newton scheme to solve for the non-martingale portfolio weights (with $(\Delta_S(s_1,v),\Delta_L(s_1,v)$ frozen so the Hessian is much smaller); once the Newton scheme has converged to tolerance) they then use a Sinkhorn-type step where they *jointly* solve the two non-linear equations satisfied by $(\Delta_S(s_1,v),\Delta_L(s_1,v))$ at each (s_1,v) node, using the well known Levenberg-Marquardt algorithm. They then repeat this cycle. Conversely, their implied Newton scheme just uses a single global Newton scheme for the non-martingale weights but they re-solve for $(\Delta_S(s_1,v),\Delta_L(s_1,v))$ after each Newton iterate, which means the partial derivatives for the Newton scheme with respect to a^2 (i.e. the portfolio weights for the call options on S_{T_2}) are more cumbersome to evaluate than before, and have to be computed using implicit differentiation (see also [Bour25] for the simpler case when there are no VIX options, and subsection 2.6.1 below for a brief discussion on convergence for Newton's scheme applied to the problem in this article).

1.1 Outline of article

In this article, we adapt the Schrödinger bridge approach in [Guy20] to the problem of constructing a joint density for two FX rates X and Y consistent with observed European option prices on all three cross-rates. The cross-smile problem we consider in this article (as outlined in the abstract) falls under the general class of problems of the form $\inf_{\mu\in C} H(\mu|\bar{\mu})$ for a convex set C which is closed under the total variation metric (see Lemma 2.2), for which there is a general existence/uniqueness result in Theorem 1.10 in [Nutz22].

Two of the Sinkhorn equations are of the usual fixed-point type but the final equation requires numerical root finding as for the fourth and fifth equations in the first order optimality conditions for the SPX-VIX problem in [Guy20]). This can also be solved as a pure optimal transport (OT) (linear programming) problem where we compute the max or min price of a non-tradable option given tradable European options on all three cross-rates (as indeed can Guyon's SPX-VIX calibration problem[5], see also [GMN17]. However, the Schrödinger bridge solution is more realistic since the optimal coupling μ^* has a joint density unlike OT problems where the support of μ^* is typically sparse. To cite a famous example, the left-curtain coupling in [BJ16] is the solution $\mathbb{P}^*$ to $\sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu)} \mathbb{E}^{\mathbb{P}}(c(X,Y))$ if c satisfies the *Spence-Mirlees condition* $c_{xyy} > 0$ and $\mu \ll Leb$. In particular, $Y = T_u(X)$ or $T_d(X)$ if X lies in a certain range (i.e. a binomial model) and $Y = X$ otherwise, so $\text{card}(\text{supp}(\mathbb{P}^*)) \leq 2$ (see also slide 14 and the

[5] We thank Rowan Austin for demonstrating this in MOSEK using sparse flattened matrices to deal with the constraints.

heat maps on slides 16 and 38 in [dM18]), with $T_u(x) \geq x$ non-decreasing and $T_d(x) \leq x$ non-increasing. Explicit derivations of the non-linear coupled ODEs for $T_u(.)$ and $T_d(.)$ are given on page 12 in [HLT16], and the associated hedge quantities on page 18 in [HLT16], and this problem can be solved numerically using MOSEK). Variants of the left-curtain coupling are derived in the context of robust pricing/hedging for at-the-money forward straddle options using a Lagrangian argument in [HK15] ($n = 3$), and [HN12] ($n = 2$); one can also compute e.g. minimal price of an American put with finite exercise dates given two marginals, see [Norg25],[HN19],[HN25] for theoretical results in this direction, and again such problems can often be solved numerically in MOSEK.

Returning to the current article, the usual Gibbs-type exponential ansatz for the joint density leads to a system of three non-linear coupled integral equations for (u, v, w). We solve this system using a (quasi) Sinkhorn fixed point scheme, and we characterize the rate of convergence for the three marginals (which is $O(\frac{1}{n})$ in terms of relative entropy), assuming an admissible joint law exists with finite entropy.

We also discuss more realistic choices for the reference measure e.g. using the Gaussian copula, and one can also adapt the Sinkhorn scheme for the forward-starter calibration problem in the usual setup with one asset and two maturities which leads to a fourth integral equation to enforce the martingale condition $\mathbb{E}(\Delta(X)(Y - X))$ for all $\Delta \in C_b$ [6], which requires an additional root-find. In section 2.6, we focus on the finite-option case where the standard Sinkhorn algorithm is not directly applicable, using the duality between minimal entropy equivalent risk-neutral measures and the concave exponential utility maximization problem from chapter 3 in [FS04], and explain how to compute sensitivities of the value function in terms of the optimal portfolio weights which serve as the Lagrange multipliers. Again we can typically use Python (or better, Python with MOSEK or Clarabel) to solve the primal and dual problems directly here (hence no need for Sinkhorn; numerical results for the latter are given in Figure 3). We also construct a continuous-time martingale model consistent with the three marginals using conditional sampling as in [BG24].

Note if we extend the main idea here to n currencies, we have $\frac{1}{2}n(n-1)$ cross-rates, so for the finite-option problem with m-tradable options per currency we have $\frac{1}{2}n(n-1)m$ portfolio weights to solve for, plus the weights for the $n-1$ (non-redundant) forward contracts vs USD, and the integrals are now $n-1$-dimensional with respect to an $n-1$ dimensional pdf μ^* (smile data for this is available at `https://www.investing.com/currencies/eur-usd-options`), so for $n = 4$ this is the same as for the SPX-VIX problem in [Guy20].

Finally, in section 3 (under mild assumptions) we show that the upper (resp. lower) bound for options on X/Y which pay $(X - KY)^+$ in the original currency are attained by the lower (resp. upper) Fréchet-Hoeffding bound (see also Hobson et al. [HLW05],[HLW05b] for a more involved analysis with basket options where the weights are assumed to be non-negative or 1 unlike our example here).

2 Outline of problem

The BIS 2025 Triennial Survey (annex tables) shows total FX options turnover (Apr 2025, net-net) at \$634,238m/day, and for major currencies like the Euro, Dollar and Pound, European options are very actively traded on all three cross-rates. A natural question to ask is: how do we construct a model so as to be consistent with observed European option prices at multiple strikes (at a single maturity) on all three cross-rates. This problem is attempted in [Aus11] but the article does not check that the resulting joint density is non-negative.

Assuming our home currency is dollars, we let X denote the price of 1 Euro in dollars at time T in the future (known as the EUR/USD rate), and Y denote the price of 1 pound in dollars at time T (the GBP/USD rate). A European option on EUR/GBP (i.e. on the cross-rate $Z = X/Y$) pays $(Z - K)^+$ pounds at time T, or equivalently $Y(Z - K)^+$ dollars. If we assume interest rates are zero for simplicity (otherwise we work with forward rates instead), then given a risk-neutral measure $\mathbb{Q}$ associated to the US economy, the initial price of such an option (in pounds) is

$$\frac{1}{Y_0}\mathbb{E}^{\mathbb{Q}}(Y(\frac{X}{Y} - K)^+) \quad = \quad \mathbb{E}^{\tilde{\mathbb{Q}}}((Z - K)^+)$$

where $Y_0 = \mathbb{E}^{\mathbb{Q}}(Y)$, and we have re-written the expectation on the left side here using the GBP numéraire measure:

$$\tilde{\mathbb{Q}}(A) \quad := \quad \mathbb{E}^{\mathbb{Q}}(\frac{Y}{\mathbb{E}^{\mathbb{Q}}(Y)}1_A)$$

associated with using GBP as the home currency. From here, we will just use $\mathbb{E}(.)$ to denote expectations under $\mathbb{Q}$.

From the Breeden-Litzenberger formula, we can extract the law of X and Y under $\mathbb{Q}$ from European call prices at all strikes on X, and the law of X/Y under $\tilde{\mathbb{Q}}$, and in practice this would typically be done using an SVI-type parametrization to interpolate between tradable strikes with parameters chosen so as to preclude butterfly arbitrage (see the many articles by Gatheral, Jacquier, Martini, Mingone et al. on this theme).

[6] We thank Youpeng Chen and Max Chellew for demonstrating this

2.1 Formulation of the problem

We assume $\mathbb{R}_+ = (0, \infty)$ throughout. Now suppose we are given three target laws μ_X, μ_Y and μ_Z on $\mathbb{R}_+$ and (without loss of generality) we assume that

$$\int_{\mathbb{R}_+} x\mu_X(dx) \quad = \quad \int_{\mathbb{R}_+} y\mu_Y(dy) \quad = \quad \int_{\mathbb{R}_+} z\mu_Z(dz) \quad = \quad 1 \tag{1}$$

i.e. the initial rates $X_0 = Y_0 = 1$ and hence $Z_0 = X_0/Y_0 = 1$ as well. Let $\Pi(\mu, \nu)$ denote the space of joint probability measures on $\mathbb{R}_+^2$ with marginals μ and ν respectively. We wish to find a $\mu^* \in \Pi(\mu_X, \mu_Y)$ such that if X, Y have joint distribution μ^*, then $X/Y \sim \mu_Z$ under the probability measure $\tilde{\mathbb{Q}}$ defined above.

Remark 2.1 If such a μ^* exists, it can then be used to price quanto options like e.g. Best-of options which pay $(X - K_1)^+/K_1 \vee (Y - K_2)^+/K_2$, or a quanto option which pays $(\frac{X}{Y} - K)^+$ dollars (as opposed to GBP as above).

2.2 Technical results

Consider a measurable space $(\Omega, \mathcal{F})$, and let $\mathcal{P}(\Omega)$ denote the collection of probability measures on $(\Omega, \mathcal{F})$. For $\mathbb{P} \in \mathcal{P}(\Omega)$, $H(\mathbb{Q}|\mathbb{P}) := \mathbb{E}^{\mathbb{Q}}(\log \frac{d\mathbb{Q}}{d\mathbb{P}}) = \mathbb{E}^{\mathbb{P}}(\frac{d\mathbb{Q}}{d\mathbb{P}} \log \frac{d\mathbb{Q}}{d\mathbb{P}})$ denotes the *entropy* of $\mathbb{Q}$ with respect to $\mathbb{P}$. Moreover, since $f(z) = z \log z$ is convex, $H(\mathbb{Q}|\mathbb{P}) = \mathbb{E}(f(\frac{d\mathbb{Q}}{d\mathbb{P}})) \geq f(\mathbb{E}(\frac{d\mathbb{Q}}{d\mathbb{P}})) = 0$. $H(\mathbb{Q}|\mathbb{P})$ is jointly l.s.c. and convex, see e.g. Lemma 1.3 in [Nutz22].

To specialize to our problem above, we now let $\mathscr{E}$ denote the set of probability measures μ on $(\mathbb{R}_+^2, \mathcal{B}(\mathbb{R}_+^2))$ such that $\int u(x)\mu(dx, dy) = \int u d\mu_X$, $\int v(y)\mu(dx, dy) = \int v d\mu_Y$ and $\int y w(\frac{x}{y})\mu(dx, dy) = \int w d\mu_Z$ for all $u, v, w \in C_b(\mathbb{R}_+)$. Recall that the last condition is equivalent to

$$\int_{\mathbb{R}_+^2} (x - Ky)^+ \mu(dx, dy) \;=\; \int_{\mathbb{R}_+} (z - K)^+ \mu_Z(dz) \text{ for all } K > 0 \tag{2}$$

which is the formulation that we use in Appendix A. We now make the following standing assumption:

Assumption 2.1 *$\mathscr{E}$ is non-empty, and contains an element μ with $H(\mu|\bar{\mu}) < \infty$.*

Then $\mathscr{E}$ is convex, and the following property is readily verified:

Lemma 2.2 *$\mathscr{E}$ is closed in the weak topology, (and since total variation convergence implies weak convergence) $\mathscr{E}$ is also closed under the total variation metric.*

Proof. See Appendix A. ■

Then by Lemma 1.10 in [Nutz22] (or Theorem 2.1 in [Csi75]), $H(\mu|\bar{\mu})$ attains a unique minimum μ^* on $\mathscr{E}$, and (from Theorem 2.1 in [Csi75]),

$$\mu^* \quad = \quad e^g \, \bar{\mu} \tag{3}$$

where g lies in the closed subspace of $L^1(\mu^*)$ spanned by the functions of the form $u(x), v(y)$ and $y w(\frac{x}{y})$.

Unfortunately, this does not automatically imply that g is of the form $u(x) + v(y) + y w(\frac{x}{y})$. For this reason, we will not solve the problem $\inf_{\mu \in \mathscr{E}} H(\mu|\bar{\mu})$ directly or use the result in (3) (which we don't require anyway)[7], but rather we use

$$\mu^*(x, y) \quad = \quad e^{u^*(x) + v^*(y) + y w^*(\frac{x}{y})} \bar{\mu}(x, y) \tag{4}$$

as an ansatz for an admissible μ (i.e. in $\mathscr{E}$), and (assuming $\mathscr{E}$ is non-empty) we show that the marginals for the associated Sinkhorn scheme (defined below) converge under appropriate conditions on the target laws, which is all we need for our calibration problem. This circumvents the need for dualization arguments as in [Guy20], but we do discuss duality for the finite-option case below where it becomes more practically relevant.

[7]Solving $\inf_{\mu \in \mathscr{E}} H(\mu|\bar{\mu})$ also lead to issues with having to prove attainment for the dual problem to justify the exponential ansatz which (even for the much simpler problem considered in Theorem 2.1 in [Nutz22]) requires pages of intermediate lemmas to prove, see section 2.3 and the discussion below Eq 2.8 there, and the recent preprint [NW25].

2.3 Integral equations for u, v, w

From here on we assume the following:

Assumption 2.3 *μ_X, μ_Y and μ_Z have densities.*

Let $\bar{\mu}$ denote a reference probability measure, typically chosen to be the product measure of μ_X and μ_Y, so $\bar{\mu}$ already satisfies 2 of the 3 marginal constraints. Then using the ansatz in (4) the marginal constraints can be written

$$
\begin{aligned}
\mu_X(x) &= e^{u(x)} \int_0^\infty e^{v(y)+yw(\frac{x}{y})} \bar{\mu}(x,y)dy \\
\mu_Y(y) &= e^{v(y)} \int_0^\infty e^{u(x)+yw(\frac{x}{y})} \bar{\mu}(x,y)dx \\
\mu_Z(K) &= \int_0^\infty \int_0^\infty y\delta(\frac{x}{y} - K) e^{u(x)+v(y)+yw(\frac{x}{y})} \bar{\mu}(x,y)dydx \\
&= \int_0^\infty \int_0^\infty \frac{x}{z} \delta_K(dz) e^{u(x)+v(\frac{x}{z})+\frac{x}{z}w(z)} \frac{x}{z^2} \bar{\mu}(x, \frac{x}{z})dzdx \\
&= \int_0^\infty \frac{x}{K} e^{u(x)+v(\frac{x}{K})+\frac{x}{K}w(K)} \frac{x}{K^2} \bar{\mu}(x, \frac{x}{K})dx
\end{aligned} \tag{5}
$$

where we have made the transformation $z = x/y$ (i.e. $y = x/z$), so $dy = -\frac{x}{z^2}dz$ in the inner integral. We can formally obtain the same three equations using a variational argument as on pg 8 in [Guy20] (see also Remark 3.4 in [Nutz22]).

If the integrals in the expressions for $\mu_X(x)$ and $\mu_Y(y)$ are finite and non-zero, the first two equations can be re-arranged as

$$
\begin{aligned}
u(x) &= \log \mu_X(x) - \log \int_0^\infty e^{v(y)+yw(\frac{x}{y})} \bar{\mu}(x,y)dy \\
v(y) &= \log \mu_Y(y) - \log \int_0^\infty e^{u(x)+yw(\frac{x}{y})} \bar{\mu}(x,y)dx
\end{aligned}
$$

and from here on we make the usual choice that $\bar{\mu}(x,y) = \mu_X(x)\mu_Y(y)$, so these simplify further to

$$
u(x) = -\log \int_0^\infty e^{v(y)+yw(\frac{x}{y})} \mu_Y(y)dy \quad , \quad v(y) = -\log \int_0^\infty e^{u(x)+yw(\frac{x}{y})} \mu_X(x)dx
$$

and setting $K = z$, we see that

$$
\bar{H}(z; u, v, w) = \int_0^\infty e^{u(x)+v(\frac{x}{z})+\frac{x}{z}w} \frac{x^2}{z^3} \mu_X(x)\mu_Y(\frac{x}{z})dx - \mu_Z(z) = 0 \tag{6}
$$

where (with mild abuse of notation) w is a scalar here unlike above.

2.4 The Sinkhorn scheme

The Sinkhorn-type fixed point iterative scheme to solve these coupled equations is then given by

$$
u_{n+1}(x) := -\log \int_0^\infty e^{v_n(y)+yw_n(\frac{x}{y})} \mu_Y(y)dy \tag{7}
$$

$$
v_{n+1}(y) := -\log \int_0^\infty e^{u_{n+1}(x)+yw_n(\frac{x}{y})} \mu_X(x)dx \tag{8}
$$

$$
0 = \bar{H}(z; u^{n+1}, v^{n+1}, w^{n+1}) \tag{9}
$$

with $u^0 \equiv v^0 \equiv w^0 \equiv 0$. It is also instructive to re-write the first equation as

$$
\begin{aligned}
\mu_X(x) &= e^{u_{n+1}(x)} \int_0^\infty e^{v_n(y)+yw_n(\frac{x}{y})} \mu_X(x)\mu_Y(y)dy \\
&= e^{u_{n+1}(x)-u_n(x)} \int_0^\infty e^{u_n(x)+v_n(y)+yw_n(\frac{x}{y})} \mu_X(x)\mu_Y(y)dy \\
&= e^{u_{n+1}(x)-u_n(x)} \mu^X_{n,n,n}(x)
\end{aligned}
$$

and (after doing the same for the Y equation) we see that

$$
u_{n+1}(x) = u_n(x) + \log \frac{\mu_X(x)}{\mu^X_{n,n,n}(x)} \quad , \quad v_{n+1}(y) = v_n(y) + \log \frac{\mu_Y(y)}{\mu^Y_{n+1,n,n}(y)} .
$$

Note we can also re-write Eq (6) as

$$\int_0^\infty e^{u(zy)+v(y)+yw} y^2 \mu_X(yz)\mu_Y(y)dy \;-\; \mu_Z(z) \;=\; 0$$

by letting $x = yz$, so $dx = zdy$. Then if $w^*(z) = \inf\{w : \int_0^\infty e^{u(zy)+v(y)+yw} y^2 \mu_X(yz)\mu_Y(y)dy = +\infty\}$, then (from the monotone convergence theorem) we know that

$$\lim_{w \nearrow w^*(z)} \int_0^\infty e^{u(zy)+v(y)+yw} y^2 \mu_X(yz)\mu_Y(y)dy \;=\; \int_0^\infty e^{u(zy)+v(y)+yw^*(z)} y^2 \mu_X(yz)\mu_Y(y)dy\,. \tag{10}$$

The right-hand side may or may not be infinite, but we clearly need it to be $\geq \mu_Z(z)$ to be able to find a root of $\bar{H}(z; u, v, w)$. We now make an additional mild assumption to preclude any root-finding problems.

Assumption 2.4 *$0 < Y \leq \bar{Y} < \infty$, $\mu_X, \mu_Z > 0$ on $\mathbb{R}_+$ and $\mu_Y > 0$ on $(0, \bar{Y}]$.*

Remark 2.2 Note the first condition is not especially unrealistic since we are working with major FX rates not e.g. equities or crypto with fat tails, and the pound has been above \$1.054 since 1985 and below \$2.86 since 1957.

Lemma 2.5 *Under Assumption 2.4, $u_n \in L^1(\mu_X)$, $v_n \in L^1(\mu_Y)$ and $w_n \in L^1(\mu_Z)$ for all $n \in \mathbb{N}$ (cf. Lemma 6.4 in [Nutz22]).*

Proof. We proceed by induction: u_0, v_0, w_0 are zero by assumption, and we take the stated claim as our inductive hypothesis. We can re-write the first two steps of the next iterate of the Sinkhorn scheme in the form

$$\frac{d\mu_{n+1,n,n}}{d\mu_{n,n,n}}(x,y) \;=\; \frac{d\mu_X}{d\mu^X_{n,n,n}}(x) \quad , \quad \frac{d\mu_{n+1,n+1,n}}{d\mu_{n+1,n,n}}(x,y) \;=\; \frac{d\mu_Y}{d\mu^Y_{n+1,n,n}}(y) \tag{11}$$

where a superscript like μ^X refers to the X-marginal of μ (and similarly for Y, see also Algorithm 6.2 in [Nutz22]). We first remark that if $\mu_{n,n,n}$ is a probability measure, then $\mu_{n+1,n,n}$ and $\mu_{n+1,n+1,n}$ are too (and likewise $\mu_{n+1,n+1,n+1}$), because each update enforces a marginal that integrates to 1.

Moreover, any probability density of the form $\mu(x,y) = e^{u(x)+v(y)+yw(x/y)}\mu_X(x)\mu_Y(y)$ with u, v, w finite Lebesgue a.e. is equivalent to $\bar{\mu} = \mu_X \times \mu_Y$ under Assumption 2.4; in particular, the X-marginal of μ is $\chi(x)\mu_X(x)$ where $\chi(x) = e^{u(x)} \int_0^{\bar{Y}} e^{v(y)+yw(x/y)}\mu_Y(y)dy$, and hence equivalent to μ_X, because $\chi(x)$ is finite for (Lebesgue) almost all x (or else μ_X would not be a marginal, by Tonelli's theorem), and $\chi > 0$ Lebesgue-a.e.

Similarly, the Y-marginal of μ is $\int_0^\infty e^{u(x)+yw(x/y)}\mu_X(x)dx \cdot e^{v(y)}\mu_Y(y)$, and this is equivalent to μ_Y for the same reason. Hence (under the inductive hypothesis), the right-hand sides of both equations in (11) are finite Lebesgue-a.e., and $\mu_{n+1,n,n}$ and $\mu_{n+1,n+1,n}$ are both equivalent to $\bar{\mu}$.

Moreover

$$\begin{aligned} H(\mu_{n+1,n,n}|\mu_{n,n,n}) \;&=\; \int \log \frac{d\mu_{n+1,n,n}}{d\mu_{n,n,n}} \mu_{n+1,n,n}(x,y)dxdy \;=\; \int\int (u_{n+1}(x) - u_n(x))\mu_{n+1,n,n}(x,y)dxdy \\ &=\; \int (u_{n+1} - u_n)d\mu_X \\ &=: \int (u_{n+1} - u_n)^+ d\mu_X - \int (u_{n+1} - u_n)^- d\mu_X \;\in\; [0, \infty] \end{aligned}$$

since $H(\mu_{n+1,n,n}|\mu_{n,n,n}) \geq 0$ and $\mu_{n+1,n,n}$ has the correct marginal for X by construction. Since the integral is non-negative, we are not in an $\infty - \infty$ situation, so $(u_{n+1} - u_n)^- \in L^1(\mu_X)$. Then using that $u^-_{n+1} \leq u^-_n + (u_{n+1} - u_n)^-$, and $u_n \in L^1(\mu_X)$ by the inductive hypothesis, we see that $u^-_{n+1} \in L^1(\mu_X)$.

From Assumption 2.1, there exists a $\mu_* \in \mathscr{E}$ with $H(\mu_*|\bar{\mu}) < \infty$, and from the discussion above we know that $\mu_{n+1,n,n} \ll \bar{\mu}$, hence by Lemma 1.4 (a) in [Nutz22] with $Q = \mu_* \in \mathscr{E}$, $Q' = \mu_{n+1,n,n}$ and $R = \bar{\mu}$, we see that

$$(\log \frac{d\mu_{n+1,n,n}}{d\bar{\mu}})^+ \;=\; (u_{n+1} + v_n + yw_n)^+ \;\in\; L^1(\mu_*)\,.$$

Using the general identity $a^+ \leq (a+b)^+ + b^-$ with $a = u_{n+1}$ and $b = v_n + yw_n$, we know that $b^- \in L^1(\mu_*)$ from the induction hypothesis and μ_* has the correct Y and Z-marginals, so we conclude that $u^+_{n+1} \in L^1(\mu_*)$ and hence $u^+_{n+1} \in L^1(\mu_X)$. Similarly

$$H(\mu_{n+1,n+1,n}|\mu_{n+1,n,n}) \;=\; \int (v_{n+1} - v_n)d\mu_{n+1,n+1,n} \;=\; \int (v_{n+1} - v_n)d\mu_Y$$

for all $n \geq 0$, and we use an analogous argument to verify that $v_{n+1} \in L^1(\mu_Y)$.

Using that

$$\begin{aligned}\mathbb{E}^{\mu_{n+1,n+1,n}}(Y\delta(Z-z)e^{Y(w-w_n(z))}) &= \mathbb{E}^{\tilde{\mu}_{n+1,n+1,n}}(\delta(Z-z)e^{Y(w-w_n(z))}) \\ &= \mathbb{E}^{\tilde{\mu}_{n+1,n+1,n}}(e^{Y(w-w_n(z))}|Z=z)\,\tilde{\mu}^Z_{n+1,n+1,n}(z)\end{aligned}$$

(where $\frac{d\tilde{\mu}_{n+1,n+1,n}}{d\mu_{n+1,n+1,n}}=Y$ and $\tilde{\mu}^Z_{n+1,n+1,n}(z)$ is the Z-density of $\tilde{\mu}_{n+1,n+1,n}$, and $\mu_{n+1,n+1,n}$ has the correct Y-marginal so $\tilde{\mu}_{n+1,n+1,n}$ is a probability measure) the final Sinkhorn step can be re-written as finding the solution $w=w(z)$ to

$$F(z,w) \;:=\; \mathbb{E}^{\tilde{\mu}_{n+1,n+1,n}}(e^{Y(w-w_n(z))}|Z=z)-\frac{\mu_Z(z)}{\tilde{\mu}^Z_{n+1,n+1,n}(z)} \;=\; 0 \tag{12}$$

[8] for μ_Z-a.e. z. $w_n(z)$ is finite μ_Z a.e. because $w_n\in L^1(\mu_Z)$ by assumption, and (since $0<Y\leq\bar{Y}$) the conditional expectation in (12) is finite for $w\in\mathbb{R}$, and (from the monotone convergence theorem) $F(z,w)$ tends to $+\infty$ as $w\to\infty$, and (from the bounded convergence theorem) tends to $-\frac{\mu_Z(z)}{\tilde{\mu}^Z_{n+1,n+1,n}(z)}$ as $w\searrow-\infty$, and is continuous in w (from the dominated convergence theorem). Hence there exists a unique finite $w=w_{n+1}(z)$ satisfying (12), so the $\mathbb{Q}$-marginal of $\mu_{n+1,n+1,n+1}$ is μ_Z as required. Moreover, F is Borel in z, continuous and strictly increasing in w μ_Z a.e., hence (by the measurable implicit function theorem, Carathéodory's conditions), the unique root $w_{n+1}(z)$ admits a Borel version.

Then $\mu_{n+1,n+1,n+1}$ is a probability measure equivalent to $\bar{\mu}$, and

$$H(\mu_{n+1,n+1,n+1}|\mu_{n+1,n+1,n}) \;=\; \int y(w_{n+1}-w_n)d\mu_{n+1,n+1,n+1} \;=\; \int(w_{n+1}-w_n)d\mu_Z$$

so using the same argument as for u_{n+1} and v_{n+1}, we can check that $w_{n+1}\in L^1(\mu_Z)$. ■

In the next section we prove convergence of the marginals for the scheme in the total variation metric. The Sinkhorn scheme typically converges very quickly in practice, and note the third equation here requires numerical root-finding as for the SPX-VIX calibration problem discussed in [Guy20], but we have one less dimension here since we only have to compute $w(z)$ for a range of values for the one argument z not a two-variable function of the form $\Delta(s_1,v)$ as in [Guy20].

2.5 Order of convergence for the Sinkhorn scheme

Set $\mu_{i,j,k}(x,y)=e^{u_i(x)+v_j(y)+yw_k(x/y)}\bar{\mu}(x,y)$, and let $\mu^X_{i,j,k}(x)$ (resp. $\mu^Y_{i,j,k}(y)$) denote the first (resp. second) marginal of $\mu_{i,j,k}$, and let $\tilde{\mu}^Z_{i,i,i}$ denote the Z-marginal of $\tilde{\mu}_{i,i,i}$ defined via $\frac{d\tilde{\mu}_{i,i,i}}{d\mu_{i,i,i}}=Y$.

Proposition 2.6 *Under Assumptions 2.1, 2.3 and 2.4, the marginals for the Sinkhorn scheme converge; specifically we have that $\|\mu^X_{n,n,n}-\mu_X\|_{TV}\to0$ and $\|\mu^Y_{n,n,n}-\mu_Y\|_{TV}\to0$ as $n\to\infty$ (where $\|.\|_{TV}$ denotes the total variation distance, see [Nutz22]) and $\tilde{\mu}^Z_{n,n,n}=\mu_Z$ for all $n\geq1$. Moreover there exists a constant A such that $\|(\frac{1}{n}\sum_{i=0}^{n-1}\mu_{i,i,i})^X-\mu_X\|_{TV}\leq\frac{A}{\sqrt{n}}$ and $\|(\frac{1}{n}\sum_{i=0}^{n-1}\mu_{i,i,i})^Y-\mu_Y\|_{TV}\leq\frac{A}{\sqrt{n}}$ i.e. the marginals of the average of the Sinkhorn iterates have $O(\frac{1}{\sqrt{n}})$ convergence.*

Proof. Following Lemma 6.4 in [Nutz22], we see that

$$\begin{aligned}H(\mu_{n+1,n,n}|\mu_{n,n,n}) &= \int\log\frac{d\mu_{n+1,n,n}}{d\mu_{n,n,n}}\mu_{n+1,n,n}(x,y)dxdy \;=\; \int\int(u_{n+1}(x)-u_n(x))\mu_{n+1,n,n}(x,y)dxdy\\ &= \int(u_{n+1}-u_n)d\mu_X \end{aligned}\tag{13}$$

since $\mu_{n+1,n,n}$ has the correct marginal for X by construction (see also Eq 6.5 in [Nutz22]), and similarly

$$\begin{aligned}H(\mu_{n+1,n+1,n}|\mu_{n+1,n,n}) &= \int(v_{n+1}-v_n)d\mu_{n+1,n+1,n} \;=\; \int(v_{n+1}-v_n)d\mu_Y\\ H(\mu_{n+1,n+1,n+1}|\mu_{n+1,n+1,n}) &= \int y(w_{n+1}-w_n)d\mu_{n+1,n+1,n+1} \;=\; \int(w_{n+1}-w_n)d\mu_Z\end{aligned}\tag{14}$$

for all $n\geq0$. From Assumption 2.1, there exist a $\mu_*\in\mathscr{E}$ with $H(\mu_*|\bar{\mu})<\infty$, so (by Lemma 1.4b) in [Nutz22]) we see that

$$H(\mu_*|\bar{\mu})-H(\mu_*|\mu_{n,n,n}) \;=\; \mathbb{E}^{\mu_*}(\log\frac{d\mu_{n,n,n}}{d\bar{\mu}}) \tag{15}$$

[8] Note that $\tilde{\mu}^Z_{n+1,n+1,n}(z)>0$ for μ_Z-a.e. z by Assumption 2.4 and the equivalence of $\mu_{n+1,n+1,n}$ and $\bar{\mu}$, so the denominator is well-defined.

which (from the definition of the scheme) we can re-write as

$$\mathbb{E}^{\mu_*}(\log \frac{d\mu_{n,n,n}}{d\bar{\mu}}) \quad = \quad \mathbb{E}^{\mu_*}(u_n(X)+v_n(Y)+Yw_n(X/Y)) \quad = \quad \mathbb{E}^{\mu_X}(u_n(X))+\mathbb{E}^{\mu_Y}(v_n(Y))+\mathbb{E}^{\mu_Z}(w_n(Z)). \tag{16}$$

where the integrability from Lemma 2.5 is used to obtain the final equality, and we have used that $\mu_* \in \mathscr{E}$ has marginals μ_X, μ_Y, μ_Z in the sense of the defining constraints). Re-writing the right-hand side here as a telescoping sum using the three equations at the beginning of the proof, we see that

$$H(\mu_*|\mu_{n,n,n}) \quad = \quad H(\mu_*|\bar{\mu}) - \sum_{i=0}^{n-1}(H(\mu_{i+1,i,i}|\mu_{i,i,i})+H(\mu_{i+1,i+1,i}|\mu_{i+1,i,i})+H(\mu_{i+1,i+1,i+1}|\mu_{i+1,i+1,i})) \tag{17}$$

so $0 \leq H(\mu_*|\mu_{n,n,n}) \leq H(\mu_*|\bar{\mu})$, and we see that $H(\mu_*|\mu_{n,n,n})$ is non-increasing[9] (recall $H(\mu_*|\bar{\mu})$ is finite by assumption). The sum on the right in (17) is bounded by $H(\mu_*|\bar{\mu})$, so the summand tends to zero. Then from the Data Processing Inequality (Example 1.7 in [Nutz22]), for $i \geq 1$, the third term in the summand in (17) dominates $H(\mu^Y_{i+1,i+1,i+1}|\mu_Y)$ and the first term dominates $H(\mu_X|\mu^X_{i,i,i})$ (note the ordering is now the other way round, but this won't matter when we apply Pinsker's inequality below) so these two quantities also tend to zero as $i \to \infty$ (so clearly $H(\mu^Y_{i,i,i}|\mu_Y) \to 0$ as well), and $\tilde{\mu}^Z_{i,i,i} = \mu_Z$ by construction for all $i \geq 1$. Then from Pinsker's inequality (see Lemma 1.2 in [Nutz22] with the 2 typo corrected),

$$\begin{aligned} \|\mu^X_{i,i,i} - \mu_X\|^2_{TV} \quad &\leq \quad \frac{1}{2}H(\mu_X|\mu^X_{i,i,i}) \quad \to \quad 0 \\ \|\mu^Y_{i,i,i} - \mu_Y\|^2_{TV} \quad &\leq \quad \frac{1}{2}H(\mu^Y_{i,i,i}|\mu_Y) \quad \to \quad 0 \end{aligned}$$

so we have convergence for the three marginals as claimed. Finally, to get the order of convergence, from the joint convexity of $H(.,.)$ and the Data processing inequality,

$$\begin{aligned} H(\mu_X|(\frac{1}{n}\sum_{i=0}^{n-1}\mu_{i,i,i})^X) \quad &= \quad H(\mu_X|\frac{1}{n}\sum_{i=0}^{n-1}\mu^X_{i,i,i}) \quad \leq \quad \frac{1}{n}\sum_{i=0}^{n-1}H(\mu_X|\mu^X_{i,i,i}) \quad \leq \quad \frac{1}{n}\sum_{i=0}^{n-1}H(\mu_{i+1,i,i}|\mu_{i,i,i}) \quad \leq \quad \frac{H(\mu_*|\bar{\mu})}{n} \\ H((\frac{1}{n}\sum_{i=1}^{n}\mu_{i,i,i})^Y|\mu_Y) \quad &= \quad H(\frac{1}{n}\sum_{i=1}^{n}\mu^Y_{i,i,i}|\mu_Y) \quad \leq \quad \frac{1}{n}\sum_{i=1}^{n}H(\mu^Y_{i,i,i}|\mu_Y) \quad \leq \quad \frac{1}{n}\sum_{i=1}^{n}H(\mu_{i,i,i}|\mu_{i,i,i-1}) \quad \leq \quad \frac{H(\mu_*|\bar{\mu})}{n}. \end{aligned}$$

Hence (from Pinsker's inequality) $\|(\frac{1}{n}\sum_{i=0}^{n-1}\mu_{i,i,i})^X - \mu_X\|_{TV}$ and $\|(\frac{1}{n}\sum_{i=1}^{n}\mu_{i,i,i})^Y - \mu_Y\|_{TV}$ are $O(\frac{1}{\sqrt{n}})$ and again we know $\tilde{\mu}^Z_{i+1,i+1,i+1} = \mu_Z$ by construction, and (from the definition of the total variation metric the contribution from $\frac{1}{n}\mu_{0,0,0}$ is irrelevant in the limit). ■

Remark 2.3 See also Section 6.2 in [Nutz22] and [GN22] for further discussion on the rate of convergence for Sinkhorn schemes.

2.6 The finite-option case

In this section, we mimic the setup in sections 1.1 (and section 3.1) for a one-period market model in [FS04] (see also Example 1.18 in [Nutz22]).

Assume we now only have $d+1$ tradable assets with (dollar) prices S^i at time 1 of the form $g_i(X,Y)$ for some measurable non-negative functions g_i with $S^0 = 1$ (i.e. cash), and $S^1 = X$ and $S^2 = Y$ (i.e. forwards on X and Y respectively), where X and Y are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}^2_+, \mathcal{B}(\mathbb{R}^2_+), \bar{\mu})$. As before, X and Y represent the EUR/USD and GBP/USD exchange rates at time 1, and we assume these $d+1$ assets have (dollar) market prices π^i at time 0. In practice we are typically interested in the case when the assets for $i > 2$ comprise European calls (resp. puts) on X and Y and cross-rate options which pay $(X-KY)^+$ (resp. $(KY-X)^+$) in dollars.

We set $\tilde{Y}^i = S^i - \pi^i$ for $i = 1, \ldots, d$, and assume interest rates are zero W.L.O.G., and assume the prices π^i are arbitrage-free (see Eq 3.5 in [FS04] for definition), and the payoffs S^i are non-redundant (see Definition 3.3 in [FS04]), so in particular we do not include a forward on the cross-rate $Z = X/Y$ since this is worth $Y \cdot \frac{X}{Y} = X = S^1$ dollars at time 1.

Then under the (somewhat stringent) moment condition $\mathbb{E}^{\bar{\mu}}(e^{\lambda \cdot \tilde{Y}}) < \infty$ for all $\lambda \in \mathbb{R}^d$, from Corollary 3.27 and Thm 3.30 in [FS04] (with $m = 0$), there exists a unique equivalent probability measure $\mu^* \sim \bar{\mu}$ and a λ^* such that

$$H(\mu^*|\bar{\mu}) \quad = \quad \inf_{\mu \sim \bar{\mu}\,:\,\mathbb{E}^{\mu}(\tilde{Y}^i)=0,\forall i} H(\mu|\bar{\mu}) \quad = \quad \sup_{\lambda \in \mathbb{R}^d} -\log \mathbb{E}^{\bar{\mu}}(e^{\lambda \cdot \tilde{Y}}) \quad = \quad -\log \mathbb{E}^{\bar{\mu}}(e^{\lambda^* \cdot \tilde{Y}}) \tag{18}$$

[9] From (15) and (16) we also see that $\mathbb{E}^{\mu_*}(\log \frac{d\mu_{n,n,n}}{d\bar{\mu}}) = \mathbb{E}^{\mu_X}(u_n(X)) + \mathbb{E}^{\mu_Y}(v_n(Y)) + \mathbb{E}^{\mu_Z}(w_n(Z)) \leq H(\mu_*|\bar{\mu}) < \infty$ i.e. we have a uniform bound.

with $\frac{d\mu^*}{d\bar{\mu}} = \frac{e^{\lambda^*\cdot\tilde{Y}}}{\mathbb{E}^{\bar{\mu}}(e^{\lambda^*\cdot\tilde{Y}})}$, i.e. we minimize the log mgf of $\tilde{Y}$ under $\bar{\mu}$ which is strictly convex and smooth (see e.g. Lemma 3.20 and Theorem 3.5 in [FS04]), which is tantamount to maximizing exponential utility using all available assets under $\bar{\mu}$. The constraints $\mathbb{E}^{\mu}(\tilde{Y}) = 0$ correspond to the first-order optimality conditions (FOC): $\frac{\partial}{\partial\lambda}\mathbb{E}^{\bar{\mu}}(e^{\lambda\cdot\tilde{Y}})|_{\lambda=\lambda^*} = \mathbb{E}^{\bar{\mu}}(\tilde{Y}e^{\lambda^*\cdot\tilde{Y}}) = 0$ (see Proposition 3.9 in [FS04]).

This is essentially the FTAP (in the direction no-arbitrage $\Rightarrow$ existence of risk-neutral measure) for a one-period market (see Remark 3.13 in [FS04]), and (18) is the finite-dimensional analog of the concave maximization problem in Eq 5.5 in [Guy20] (appropriately adapted for our problem). Figure 3 displays numerical results for this maximization scheme (using CVXPY with MOSEK) applied to the FX option (and forward) prices in Table 1.

Remark 2.4 From the envelope theorem, the Lagrange multipliers λ_i^* yield the sensitivities of the certainty equivalent:

$$-\frac{\partial}{\partial\pi^i}\log\mathbb{E}^{\bar{\mu}}(e^{\lambda^*(\pi)\cdot(S-\pi)}) \quad = \quad \lambda_i^*$$

(recall λ^* also depends on π, but the contribution from $\frac{\partial\lambda^*}{\partial\pi^i}$ doesn't show up in the final answer because it vanishes due to the FOC).

Remark 2.5 Note we can either choose $\bar{\mu}$ to be the real world market model (i.e. the $\mathbb{P}$-measure), in which case the dual (maximization) problem here is the classical problem of maximizing exp utility from the available assets, or (if e.g. we cannot easily estimate $\mathbb{P}$ from historical data), we can just choose $\bar{\mu}$ as a fictitious reference measure which we can specify to be consistent with the tradable European options prices on X and Y to give a warm starting point for the concave maximization scheme (see also Remark 13 in [Guy24]).

2.6.1 Newton's method

If we let $V(\lambda) = \log\mathbb{E}^{\bar{\mu}}(e^{\lambda\cdot\tilde{Y}})$, then the Hessian of V is

$$H_{ij} \quad := \quad \partial^2_{\lambda_i,\lambda_j}V \quad = \quad \mathrm{Cov}_{\mathbb{P}_\lambda}(\tilde{Y}^i,\tilde{Y}^j) \tag{19}$$

(cf. Lemma 3.20 and Definition 3.16 in [FS04] for the definition of $\mathbb{P}_\lambda$), and H is positive-definite because $a^\top\mathrm{Cov}_{\mathbb{P}}(\tilde{Y})a = \mathrm{Var}_{\mathbb{P}}(a\cdot\tilde{Y}) \geq 0$, with equality if and only if $a = 0$ (from the non-redundancy condition above), and $\mathbb{P}_\lambda$ is equivalent to $\mathbb{P}$ so this property carries over to $\mathbb{P}_\lambda$.

Newton's method for the dual exponential utility maximization problem in the final term of (18) is $\lambda_{n+1} = g(\lambda_n)$, where $g(\lambda) := \lambda - H(\lambda)^{-1}\nabla V(\lambda)$. Hence (from the Banach fixed point theorem) a sufficient condition for convergence is that the initial guess λ_0 lies in the neighborhood of λ^* for which g is a contraction, i.e. where the Jacobian

$$(\nabla g(\lambda))_{im} \quad := \quad \sum_{k,\ell=1}^{d} H^{-1}_{ik}\,\partial^3_{k\ell m}V(\lambda)\,(H^{-1}\nabla V(\lambda))_\ell$$

is strictly less than 1 in the operator 2 norm $\|A\|_2 := \sup_{x\in\mathbb{R}^d:\|x\|_2=1}\|Ax\|_2$ (recall that $\nabla g = 0$ at λ^*, and V is C^∞ (Lemma 3.20 in [FS04]) and $\nabla^3 V$ is a rank-3 ($d\times d\times d$) tensor here). Convergence for the Newton scheme is quadratic in this neighborhood (although the size of this neighborhood is not explicit without further bounds (e.g. a Lipschitz constant for H near λ^*), but in principle we have to numerically compute a double integral for each component of the lower triangular part of the Hessian here (or triple integrals for the scheme in [BG24]). [10]

2.7 Using a copula for $\bar{\mu}$

A more flexible choice for $\bar{\mu}$ is to use e.g. use the Gaussian copula, and fit the correlation parameter ρ to e.g. historical data or use the *Margrabe formula*: $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y$ (used for pricing EUR/GBP options under a bivariate Black-Scholes model) to back out ρ using implied volatilities.

To compute $\bar{\mu}$ for the Gaussian copula, we set $X = F_X^{-1}(\Phi(Z_1))$ and $Y = F_Y^{-1}(\Phi(Z_2))$ where $Z_1, Z_2 \sim N(0,1)$ with $\mathbb{E}(Z_1Z_2) = \rho$, then $X \sim \mu_X$ and $Y \sim \mu_Y$ as required, and

$$\mathbb{P}(X\leq x, Y\leq y) \quad = \quad \mathbb{P}(F_X^{-1}(\Phi(Z_1))\leq x, F_Y^{-1}(\Phi(Z_2))\leq y) \quad = \quad \mathbb{P}(Z_1\leq\Phi^{-1}(F_X(x)), Z_2\leq\Phi^{-1}(F_Y(y)))\,.$$

$\bar{\mu}$ is then given by

$$\bar{\mu}(x,y) \quad = \quad \frac{\partial^2}{\partial x\partial y}\mathbb{P}(X\leq x, Y\leq y) \quad = \quad \bar{F}(x)\bar{G}(y)f(\Phi^{-1}(F_X(x)),\Phi^{-1}(F_Y(y)))$$

where f is the joint density of Z_1 and Z_2, and $\bar{F}(x) = \frac{d}{dx}(\Phi^{-1}(F_X(x)))$ and $\bar{G}(y) = \frac{d}{dy}(\Phi^{-1}(F_Y(y)))$ (see numerical results in Figure 3).

[10] in practice this could be done as a single aggregated computation for the whole Hessian, and other speed-ups can also be applied

Remark 2.6 Using the Margrabe formula with all combinations of the 5×5×5 (mid) implied volatilities in the table below for the EUR-USD-GBP triangle leads to a range for the implied correlation for EUR/USD and GBP/USD of $[0.7445, 0.8156]$.

EUR/USD	1.0567	1.0680	1.0798	1.0950	1.1025
Implied vol	5.69/6.315%	5.621/5.966%	5.54/5.815%	5.509/5.854%	5.45/6.075%
a, b, σ, ρ, m	-0.0005100	0.009510	0.08579	0.30719	0.03433
GBP/USD	1.2331	1.2480	1.2632	1.2718	1.2919
Implied vol	6.148/7.239%	6.125/6.727%	5.985/6.465%	5.873/6.475%	5.681/6.772%
a, b, σ, ρ, m	0.0002773	0.002254	0.01867	-0.3272	0.003880
EUR/GBP	0.84261	0.84852	0.85478	0.86142	0.8681
Implied vol	3.414/4.584%	3.59/4.232%	3.66/4.17%	3.735/4.373%	3.681/4.841%
a, b, σ, ρ, m	0.0001047	0.001959	0.01139	0.2032	-0.001484

Table 1: Table of (bid-ask) implied volatilities with 1-month maturity (with strikes going horizontally in the 1st, 3rd and 5th rows) on 11th Feb 2024, and SVI parameters fitted to the mid-implied vols. The forward prices here are 1.0796 for EUR/USD, 1.2630 for GBP/USD and .85483 for EUR/GBP (data obtained from Bloomberg), and these SVI parameters are used for the Sinkhorn scheme.

2.8 A continuous martingale consistent with (μ_X, μ_Y, μ_Z)

We can construct a Markov functional-type continuous martingale model $(X_t, Y_t)_{t\geq 0}$ consistent with the three marginals $(\mu_X,\, \mu_Y,\, \mu_Z)$ using conditional sampling as in [BG24]. Specifically, we let

$$\begin{aligned} X_t &= \mathbb{E}(F_X^{-1}(\Phi(\frac{W_T}{\sqrt{T}}))|\mathcal{F}_t^W) &= \mathbb{E}(f(W_T)|W_t) \\ Y_t &= \mathbb{E}(F_{Y|X}^{-1}(\Phi(\frac{B_T}{\sqrt{T}}), X_T)|\mathcal{F}_t^{W,B}) &= \mathbb{E}(g(B_T, W_T)|W_t, B_t) \end{aligned}$$

for $t \in [0, T]$ (setting $X = X_T$ and $Y = Y_T$), where W, B are two independent Brownian motions, Φ is the standard Normal cdf, F_X is the distribution function of μ_X, and $F_{Y|X}(., x)$ is the conditional distribution function of Y given X (which comes from the solution $\mu^*(x, y)$ to the Sinkhorn equations). f and $g(., .)$ are shorthand for the functions which appear inside the expectations in the middle eqs in each line but with the second function re-expressed in terms of W_T. To see that Y has the correct conditional law given X at time zero, we note that

$$Y_T = F_{Y|X}^{-1}(\Phi(\frac{B_T}{\sqrt{T}}), X) = F_{Y|X}^{-1}(V, X)$$

and B and W are independent, so the $U[0, 1]$ random variables $U = \Phi(\frac{W_T}{\sqrt{T}})$ and $V = \Phi(\frac{B_T}{\sqrt{T}})$ are independent, so we are sampling $Y|X$ correctly.

Note this approach is somewhat asymmetric in so far as the Y process is more complicated that the X process; a more symmetric approach would be to use the true Bass martingale where $(X, Y) = (V_x(W_T, B_T), V_y(W_T, B_T))$ for some convex function V characterized in terms of the standard stretched Brownian motions s^2BM (see [BBHK20],[BST23],[AMP25 see e.g. Theorem 1.4 in [BST23].

3 Extremal prices for the cross-rate options given full marginals μ_X and μ_Y and Kantorovich duality

In this section we compute robust price bounds (and the corresponding sub/super hedge) for cross-rate options which pay $(X - KY)^+$ dollars, i.e. we solve the (non-entropic) optimal transport problem.

To this end, we first note the triangle-type inequality:

$$(x - y)^+ \leq (x - x_1)^+ + (x_1 - y)^+ \tag{20}$$

for all $x, y, x_1 \in \mathbb{R}$ (we just have to check all cases to verify this identity). Now let $y = p(x) := F_{\mu_Y}^{-1}(1 - F_{\mu_X}(x))$ where $F_\mu(x) := \mu([0, x])$ so $Y = p(X) \sim \mu_Y$ if $X \sim \mu_X$ i.e. the Fréchet-Hoeffding countermonotone coupling, and note that p is strictly decreasing if we assume μ_X and μ_Y have strictly positive densities. We further assume there is a unique root x_* of $p(x) = x$, with $p(x) > x_*$ for $x < x_*$ and vice versa (see first plot in Figure 4 below). Then setting $x_1 = x_*$ and $y = p(x)$ and assuming $x > x_*$, (20) becomes

$$x - p(x) \leq x - x_* + x_* - p(x)$$

(since $p(x) < x$) i.e. an equality. Conversely if $y = p(x)$ and $x < x_*$, $p(x) > x_* > x$ so both sides of (20) are zero. Hence $(x - x_*)^+ + (x_* - y)^+$ (i.e. a call on X plus a put on Y, both with strike x_*) is a superhedge for

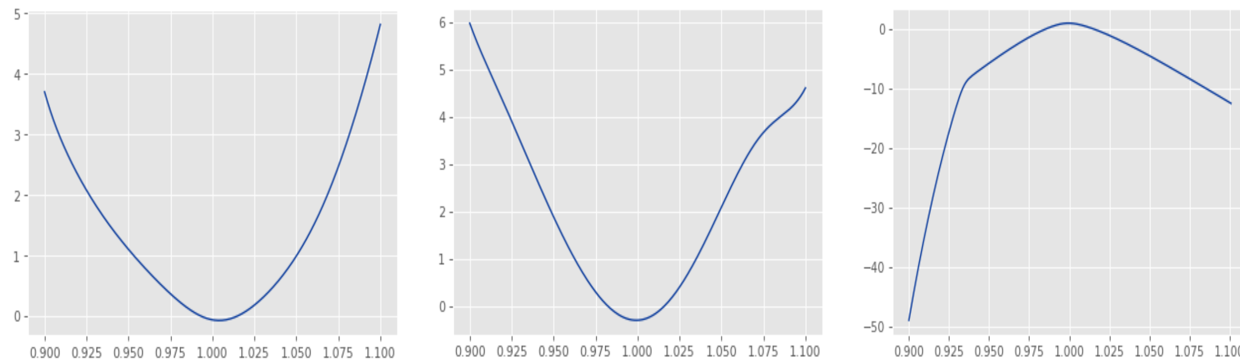

Figure 1: Here we have plotted the maximizing u, v and w respectively after 40 iterations of the Sinkhorn algorithm, using $[.9, 1.1]^2$ as the domain for numerical integration for X, Y.

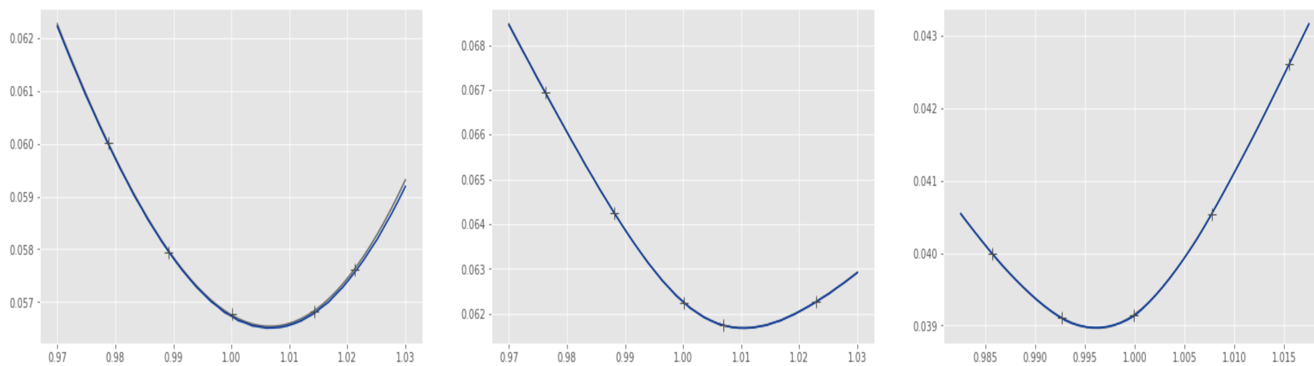

Figure 2: Here we have plotted the Sinkhorn smiles (blue) as a function of moneyness K/F_0 associated with the u, v and w in the plot above verses the original (mid)-market smiles (grey crosses) from Table 1 and the SVI interpolated smiles (grey line, which can barely be seen as it is very close to the blue curve) for EUR/USD, GBP/USD and EUR/GBP on 11th Feb 2024. Five options were used for each cross-rate (At-the-money, and .10, .25, .75, and .90 Delta calls, as is customary in FX options markets), using the standard SVI parametrization to interpolate between them and 400 point Gaussian quadrature for the single and double integrals which appear in the Sinkhorn equations, and the bisection method for the root-finding.

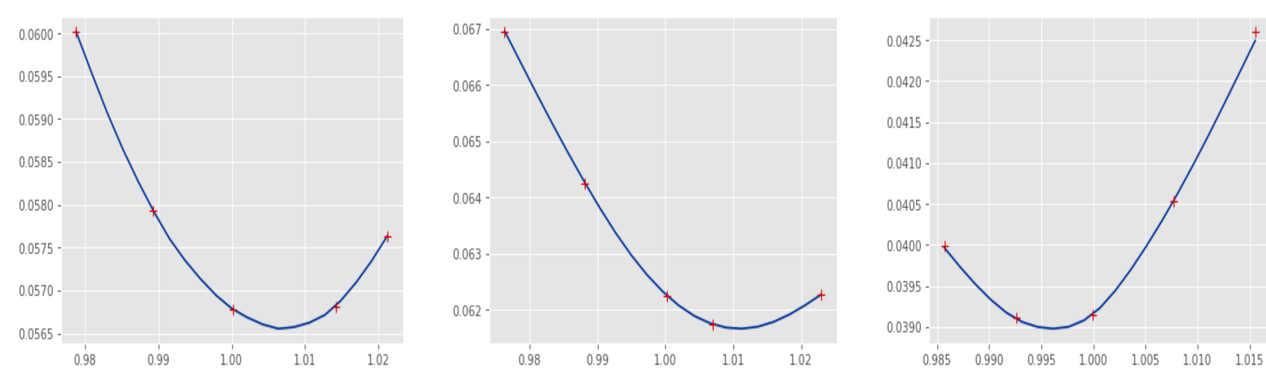

Figure 3: Here we have plotted the EUR/USD, GBP/USD and EUR/GBP smiles below) by solving the finite-dimensional concave maximization problem on the right-hand side of (18) for the same option data in Table 1, using a Gaussian copula for $\bar{\mu}$ with correlation $\rho = .6$.

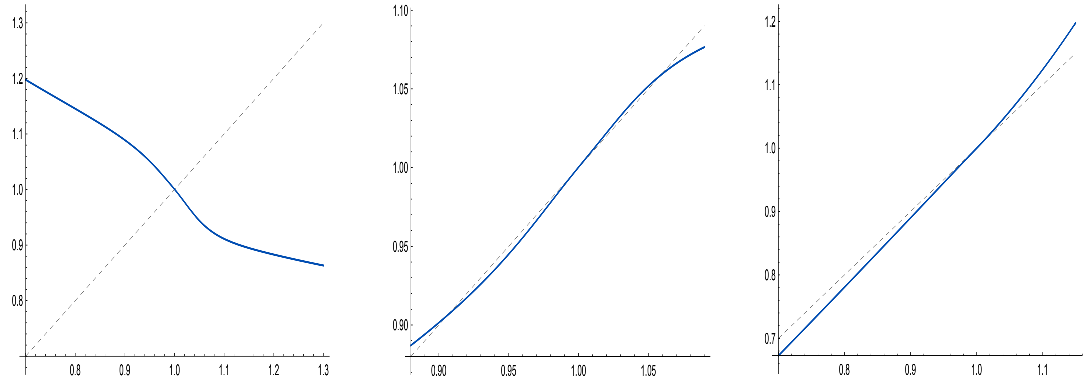

Figure 4: On the left we see the optimal transport map $y = p(x)$ (in blue) from Section 3 using the countermonotone Frechet-Hoeffding bound versus the $y = x$ line (dashed). In the 2nd plot we see the $p(x)$ map for the minimal price for EUR/GBP options which comes from the comonotone Frechet-Hoeffding bound (and we see there are three points of equality which is a problem case), and the final plot shows the same $p(x)$ function for EUR/USD options (given EUR/JPY and USD/JPY smiles on 3rd Mar 2024) which only has 1 point of equality (non-problem case) as assumed in the proof in subsection 3.

$c(x, y) = (x - y)^+$, and equality is attained by the coupling $Y = p(X)$, so this coupling is optimal for the max problem, i.e.

$$\int_0^1 (F_{\mu_X}^{-1}(u) - F_{\mu_Y}^{-1}(1-u))^+ du \quad = \quad \sup_{\mu \in \Pi(\mu_X, \mu_Y)} \mathbb{E}^\mu((X - Y)^+)$$

(see also [HLW05],[HLW05b] who look at this problem in the context of basket options). For the general case when $c(x, y) = (x - Ky)^+$, we just regard KY as the new Y variable, then x_* is now the root of $p(x) = Ky$ in the proof above, which is otherwise unchanged except now x_* depends on K.

For the lower bound, we first note that

$$(x - y)^+ \quad \geq \quad -(x_1 - x)^+ + (x_1 - y)^+ \tag{21}$$

for all $x, y, x_1 \in \mathbb{R}$. Now let $y = p(x) = F_{\mu_Y}^{-1}(F_{\mu_X}(x))$ so $Y = p(X) \sim \mu_Y$ if $X \sim \mu_X$ i.e. the Fréchet-Hoeffding comonotone bound coupling, and note that p is strictly increasing if we assume μ_X and μ_Y have strictly positive densities. We again assume there is a unique root x^* of $p(x) = x$, but now with $p(x) > x$ for $x > x^*$ and vice versa (see 3rd plot in Figure 4). Then again setting $x_1 = x^*$ and $y = p(x)$ and assuming $x \leq x^*$, (21) becomes

$$x - p(x) \quad \geq \quad -(x^* - x) + (x^* - p(x)) \quad = \quad x - p(x)$$

so we have equality. Conversely, if $x > x^*$, both sides of (21) vanish. Hence (repeating the same arguments as above), the coupling with $Y = p(X)$ is optimal for the min problem, i.e.

$$\int_0^1 (F_{\mu_X}^{-1}(u) - F_{\mu_Y}^{-1}(u))^+ du \quad = \quad \inf_{\mu \in \Pi(\mu_X, \mu_Y)} \mathbb{E}^\mu((X - Y)^+)$$

and again we can extend this argument to the general case when $c(x, y) = (x - Ky)^+$. Note our assumption that $p(x) = x$ has a unique root fails for our EUR-USD-GBP triangle example (see second plot in Figure 4) where $p(x) = x$ at three distinct x-values (see also Corollary 1.2 in [BJ16], although their result requires strict convexity condition on $c(x, y) = h(y - x)$).

Remark 3.1 For a single cross-rate option paying $(x - Ky)^+$ with market price P, we can construct a "model" consistent with P and the given marginals for X and Y if and only if P lies within the upper and lower price bounds computed above, in which case the model can just be chosen as a weighted coin toss between the two extremal models with the weight chosen to match P. It is an interesting open problem as to whether we can construct a consistent model if we have prices for two (or n) such cross-rate options when the market prices for each these options lie between their respective upper and lower price bounds [11]

References

[ACD15] Atar, R., K.Chowdhary, and P.Dupuis, "Robust bounds on risk-sensitive functionals via Rényi divergence", *SIAM Journal on Uncertainty Quantification*, 3(1):1–30, 2015.

[11]We thank David Hobson for pointing this out.

[AMP25] Acciaio, B., A.Marini, G.Pammer, "Calibration of the Bass Local Volatility model", *SIAM J. Finan. Math.*, 16/3, 803-833, 2025.

[Aus11] Austing, P., "Repricing the Cross Smile", `risk.net`, 2011

[BBHK20] Backhoff-Veraguas, J., M.Beiglböck, M.Huesmann and S.Källblad, "Martingale Benamou–Brenier: A Probabilistic Perspective", *The Annals of Probability*, Vol. 48, No. 5, 2258–2289, 2020

[BST23] Backhoff-Veraguas, J., W.Schachermayer, and B.Tschiderer, "The Bass functional of martingale transport", preprint, 2023

[BJ16] Beiglböck, M. and N.Juillet, "On a problem of optimal transport under marginal martingale constraints", *The Annals of Probability*, 2016, Vol. 44, No. 1, 42–106

[Bert15] Bertsekas, D., "Convex Optimization Algorithms", Athena Scientific, 2015.

[BG24] Bourgey, F. and J.Guyon, "Fast Exact Joint S&P 500/VIX Smile Calibration in Discrete and Continuous Time". `risk.net`, Feb 2024

[Bour25] Bourgey, F. "Fast Smile Calibration in Discrete and Continuous Time Using Sinkhorn Algorithms", SIAM conference talk, Miami, July 2025.

[Carr09] Carr, P., "Local Variance Gamma Option Pricing Model", presentation Bloomberg/Courant Institute, April 28, 2009.

[CN17] Carr, P. and S.Nadtochiy, "Local Variance Gamma and Explicit Calibration to Option Prices", *Mathematical Finance*, 27(1):151–193, 2017

[CL21] Conze, A. and P.Henry-Labordère, "Bass Construction with Multi-Marginals: Lightspeed Computation in a New Local Volatility Model", preprint, 2021.

[Csi75] Csiszár, I. "*I*-divergence geometry of probability distributions and minimization problems", *The Annals of Probability*, 3(1):146–158, 1975.

[DHL19] De March, H. and Henry-Labordere, P., "Building arbitrage-free implied volatility: Sinkhorn's algorithm and variants", preprint.

[dMHL15] de Marco, S. and P.Henry-Labordère, "Linking Vanillas and VIX options: A constrained martingale optimal transport problem", *SIAM J. Finan. Math.*, 4(54), pp. 1171–1194, 2015

[DZ10] Dembo, A. and O.Zeitouni, "Large deviations techniques and applications", Jones and Bartlet publishers, Boston, second edition, 2010.

[EKK25] Engström, L., S.Källblad, and J.Karlsson "Computation of Robust Option Prices via Structured Multimarginal Martingale Optimal Transport",*SIAM J. Finan. Math.*, 16(3), 2025

[F24] Forde, M., "Rough Bergomi revisited - exact and minimal-variance hedging for VIX and European options, and exact calibration to multiple smiles", preprint, 2024.

[FS04] Föllmer, H. and A.Schied, "Stochastic Finance: An Introduction in Discrete Time", W. de Gruyter, Berlin, 2nd edition, 2004.

[GN22] Ghosal, P. and M.Nutz, "On the Convergence Rate of Sinkhorn's Algorithm", preprint

[GLOW22] Guo, I., G.Loeper, J.Obłój and S.Wang, "Optimal transport for model calibration", `risk.net`, Jan 2022

[GLW22] Guo, I., G.Loeper, S.Wang, "Calibration of Local-Stochastic Volatility Models by Optimal Transport", *Mathematical Finance*, 32(1), 46-77, 2022.

[Guy13] Guyon, J., "A New Class of Local Correlation Models", Quantitative Research, Bloomberg LP, 2013

[Guy20] Guyon, J., "The Joint S&P 500/VIX Smile Calibration Puzzle Solved", `risk.net`, April 2020

[Guy22] Guyon, J., "Dispersion-Constrained Martingale Schrödinger Bridges: Joint Entropic Calibration of Stochastic Volatility Models to S&P 500 and VIX Smiles", preprint, 2022

[Guy24] Guyon, J., "Dispersion-Constrained Martingale Schrödinger Problems and the Exact Joint S&P 500/VIX Smile Calibration Puzzle", *Finance and Stochastics*, 28(1):27-79, 2024

[Guy24b] Guyon, J., "Fast exact joint S&P 500/VIX smile calibration in discrete and continuous time", London Mathematical Finance Seminar, Feb 2024

[GMN17] Guyon, J., R.Menegaux and M.Nutz, “Bounds for VIX Futures given S&P 500 Smiles”, *Finance&Stochastics*, Vol. 21, No. 3, pp. 593-630, 2017

[dM18] De March, H., “Multi-dimensional Martingale Optimal Transport: Local structure and Numerics”, presentation at Oxford University, 2018

[HN12] Hobson, D. and A.Neuberger, “Robust bounds for forward start options”, *Mathematical Finance*, 22, 31-56, 2012

[HK15] Hobson, D. and M.Klimmek, “Robust price bounds for the forward starting straddle”, *Finance and Stochastics*, Vol 19, Issue 1, 189-214, 2015

[HL19] Henry-Labordère, P., “From (Martingale) Schrödinger bridges to a new class of Stochastic Volatility Models”, preprint, 2019.

[HLW05] Hobson, D., P.Laurence, T-H.Wang, “Static-arbitrage upper bounds for the prices of basket options”, *Quantitative Finance*, 5 (4) 329-342, 2005

[HLW05b] Hobson, D., P.Laurence, T-H.Wang, “Static-arbitrage optimal subreplicating strategies for basket options.” *Insurance: Mathematics and Economics*, 37(3):553–572, 2005.

[HLT16] Henry-Labordere, P. and N.Touzi, “An explicit martingale version of the one-dimensional Brenier theorem”, *Finance and Stochastics*, 20(3):635–668, 2016 (version from 10th April 2015 cited here)

[HN19] Norgilas, N., D.Hobson,“Robust bounds for the American put”, *Finance and Stochastics*, Volume 23, pages 359–395, 2019

[HN25] Norgilas, N., D.Hobson,“Model-independent upper bounds for the prices of Bermudan options with convex payoffs”, preprint.

[Loep25] Loeper, G., “Sinkhorn algorithms for Schrödinger and Bass problems: a PDE point of view”, VCMF conference, Vienna, July 2025

[JLO23] Joseph, B., G.Loeper and Jan Obłój, “The Measure Preserving Martingale Sinkhorn Algorithm”, preprint

[Norg25] Norgilas, N., “Model-free price bounds of Bermudan contracts”, SIAM conference talk, Miami, July 2025.

[Nutz22] Nutz, M., “Introduction to Entropic Optimal Transport”, Lecture Notes, 2022

[Nutz24] Nutz, M., “Quadratically Regularized Optimal Transport: Existence and Multiplicity of Potentials”, preprint

[NW25] Nutz, M. and J.Wiesel, “On The Martingale Schrödinger Bridge Between Two Distributions”, preprint, 2025

A Proof of Lemma 2.2

Let $\mu_n \in \mathscr{E}$ with $\mu_n \xrightarrow{w} \mu$. For any $u \in C_b(\mathbb{R}_+)$, from weak convergence $\int u d\mu_n \to \int u d\mu$, but we also know that $\int u d\mu_n = \int u(x)\mu^X(dx)$ so $\int u d\mu = \int u(x)\mu^X(dx)$ as required, so μ has the correct X-marginal (and we use a similar argument for the Y-marginal). For the final part of the proof, we use the formulation of $\mathscr{E}$ in (2): let $f_K(x,y) := (x-Ky)^+$ and $f_{K,m} := (x-Ky)^+ \wedge m$. Then

$$0 \quad \le \quad f_K - f_{K,m} \quad = \quad ((x-Ky)^+ - m)^+ \quad \le \quad x 1_{x>m}.$$

Fix $\varepsilon > 0$. Since $\int x\, d\mu_X = 1 < \infty$, we can choose m sufficiently large so that $\int x 1_{\{x>m\}} d\mu_X \le \varepsilon$, and hence

$$\sup_n \int (f_K - f_{K,m})\, d\mu_n \quad \le \quad \varepsilon, \qquad \int (f_K - f_{K,m})\, d\mu \quad \le \quad \int x 1_{\{x>m\}}\, d\mu_X \quad \le \quad \varepsilon$$

since $\mu^X = \mu^X_n = \mu_X$ (from the first part of the proof). Then from the triangle inequality

$$|\int f_K d\mu_n - \int f_K d\mu| \quad \le \quad \int (f_K - f_{K,m}) d\mu_n + |\int f_{K,m} d\mu_n - \int f_{K,m}\, d\mu| + \int (f_K - f_{K,m}) d\mu\,.$$

Using weak convergence for the middle term, we see that

$$\limsup_{n\to\infty} |\int f_K\, d\mu_n - \int f_K\, d\mu| \quad \le \quad 2\varepsilon\,.$$

Then since $\varepsilon > 0$ is arbitrary, the claim follows.