



**HAL**  
open science

# Optimal Portfolio Choice with Cross-Impact Propagators

Eduardo Abi Jaber, Eyal Neuman, Sturmius Tuschmann

► **To cite this version:**

Eduardo Abi Jaber, Eyal Neuman, Sturmius Tuschmann. Optimal Portfolio Choice with Cross-Impact Propagators. 2024. hal-04567802

**HAL Id: hal-04567802**

**<https://hal.science/hal-04567802v1>**

Preprint submitted on 3 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Optimal Portfolio Choice with Cross-Impact Propagators

Eduardo Abi Jaber<sup>\*1</sup>, Eyal Neuman<sup>2</sup>, and Sturmius Tuschmann<sup>†2</sup>

<sup>1</sup>Ecole Polytechnique, CMAP

<sup>2</sup>Department of Mathematics, Imperial College London

March 15, 2024

## Abstract

We consider a class of optimal portfolio choice problems in continuous time where the agent's transactions create both transient cross-impact driven by a matrix-valued Volterra propagator, as well as temporary price impact. We formulate this problem as the maximization of a revenue-risk functional, where the agent also exploits available information on a progressively measurable price predicting signal. We solve the maximization problem explicitly in terms of operator resolvents, by reducing the corresponding first order condition to a coupled system of stochastic Fredholm equations of the second kind and deriving its solution. We then give sufficient conditions on the matrix-valued propagator so that the model does not permit price manipulation. We also provide an implementation of the solutions to the optimal portfolio choice problem and to the associated optimal execution problem. Our solutions yield financial insights on the influence of cross-impact on the optimal strategies and its interplay with alpha decays.

**Mathematics Subject Classification (2010):** 93E20, 60H30, 91G80

**JEL Classification:** C02, C61, G11

**Keywords:** optimal portfolio choice, price impact, cross-impact, propagator models, predictive signals, Volterra stochastic control

---

<sup>\*</sup>EAJ is grateful for the financial support from the Chaires FiME-FDD, Financial Risks, Deep Finance & Statistics and Machine Learning and systematic methods in finance at Ecole Polytechnique.

<sup>†</sup>ST is supported by the EPSRC Centre for Doctoral Training in Mathematics of Random Systems: Analysis, Modelling and Simulation (EP/S023925/1).

# 1 Introduction

Optimal portfolio choice has been a core problem in quantitative finance. In this class of problems investors dynamically select their portfolio while taking into account expected returns, risk factors, transaction and price impact costs, and trading signals. The seminal papers of Gârleanu and Pedersen [13, 14] provided a tractable framework for formulating and explicitly solving these problems under the assumptions of quadratic trading costs and exponential decay of the price impact, among others. Their work was further generalized in various directions (see e.g. [12, 18, 28]). Despite its phenomenal impact, the main drawback of Gârleanu and Pedersen’s framework is that it relies on the assumption that the price impact and the cross-impact decay at an exponential rate.

Price impact refers to the empirical fact that the execution of a large order affects a risky asset’s price in an adverse and persistent manner leading to less favourable prices. Propagator models are a central tool in describing these phenomena mathematically. They express price moves in terms of the influence of past trades, and therefore capture the decay of the price impact after each trade [6, 15]. Cross-impact models provide an additional explanation for the price dynamics of a risky asset in terms of the influence of past trades of other assets in the market. The price distortion  $D_t$  due to both effects is quantified by the  $N$ -dimensional process,

$$D_t = \int_0^t G(t, s) dX_s, \quad t \geq 0, \quad (1.1)$$

where  $X_t = (X_t^1, \dots, X_t^N)$  describes the amount of shares in each of the  $N$  assets in the portfolio at time  $t$ . Here  $G(t, s)$  is a matrix of Volterra kernel functions, so that self-impact is captured by its diagonal entries and cross-impact is captured by its off-diagonal entries. The components of  $G$  are often referred to as propagators (see e.g. [5, 6, 20, 23, 25, 26]). Throughout this paper we will adopt the convention introduced in Section 14.5.3 of [6] and refer to price impact (or equivalently price distortion) as the aggregated effects of self-impact and cross-impact.

Gârleanu and Pedersen along with follow-up papers assumed that the price distortion  $D_t$  decays exponentially with time. On the other hand, Bouchaud et al. [6] (see Chapter 14.5.3) report on two main empirical observations regarding price impact: (i) diagonal and off-diagonal elements of the propagator matrix  $G$  decay as a power-law of the lag, i.e.  $G_{ij}(t, s) \approx (t - s)^{-\beta_{ij}}$ , with  $0 < \beta_{ij} < 1$ . (ii) most of the cross-correlations between price moves ( $\approx 60 - 90\%$ , depending on the time scale), are mediated by trades themselves, i.e. through a cross-impact mechanism, rather than through the cross-correlation of noise terms, which are not directly related to trading. See also [5] for further details. Mastromatteo et al. [23] show that neglecting cross-impact effects leads to an incorrect estimation of the liquidity and to suboptimal execution strategies. In particular, they demonstrate the importance of synchronizing

the trades of correlated assets and comment on the characteristics of optimal portfolio trading strategies. Le Coz et al. [20] perform an empirical analysis using data of 500 US assets from 2017 to 2022. They conclude that price formation occurs endogenously within highly liquid assets first, and then trades in these assets influence the prices of their less liquid correlated products, with an impact speed constrained by their minimum trading frequency.

While power-law behaviour of cross-impact has strong evidence coming from market data, results on the corresponding optimal portfolio choice problem have been scarce for two main reasons: (i) power-law type or any more general propagator matrices which do not have an exponential decay turn the optimal portfolio choice problem into non-Markovian and often time inconsistent, hence standard tools of stochastic control such as dynamic programming or FBSDEs do not apply in this case. (ii) cross-impact may cause various problems from the financial point of view, such as round trips or transaction-triggered price manipulations (see [3, 19, 25] for some pathological examples). Some recent progress has been made in [1, 2] for the single asset version of this problem in the context of optimal execution with signals. However the solution for the single asset problem clearly does not take into account the aforementioned cross-impact effects.

The main theoretical contribution in this non-Markovian portfolio choice setting was derived by Alfonsi et al. [3], who proposed a discrete-time model for portfolio liquidation in presence of linear transient cross-impact using convolution type decay kernels  $G(t)$ . They characterized conditions on  $G$  which guarantee that the model admits well-behaved optimal execution strategies by ensuring that price manipulation in the sense of Huberman and Stanzl [19] is excluded. This means that a portfolio starting and terminating with zero inventory ( $X_0 = X_T = 0$ ) cannot create any profit, or equivalently negative trading costs, due to the induced price distortion  $D_t$  in (1.1). This condition translates into the following inequality,

$$\mathbb{E} \left[ \int_0^T D_t dX_t \right] \geq 0, \quad (1.2)$$

where the time grid is discrete in [3], so that the integral is written as a sum. Alfonsi et al. [3] concluded that  $G$  must be a matrix-valued nonnegative definite function in order for the model to satisfy condition (1.2) and thus preclude price manipulation. A sufficient condition for (1.2) in which  $G$  is nonincreasing, nonnegative, convex, symmetric and commuting was given in [3], as well as a characterization of matrix-valued nonnegative definite functions. Alfonsi et al. also derived a first order condition for the optimal strategy of the execution problem and obtained an explicit solution for the case where the propagator matrix is given by  $G(t) = \exp(-tC)$  for a symmetric nonnegative definite matrix  $C \in \mathbb{R}^{N \times N}$ . Note that in the framework of [3], not only the model is discrete in time, but also the alpha signals and the risk terms, which are central features of portfolio choice problems, are not incorporated. These simplifying

assumptions turn the resulting stochastic control problem into a completely deterministic one. Moreover, explicit solutions in [3] are only given for the special case of an exponential propagator matrix.

**Our contribution:** In this work we extend the results of Gârleanu and Pedersen [14] and Alfonsi et al. [3] in a few crucial directions which are of relevance to recent empirical studies on cross-impact and also of theoretical interest. Our results provide financial insights on the influence of cross-impact on optimal trading strategies and the interplay between cross-impact and alpha decays.

- (i) Derivation of explicit solutions: we formulate and solve the optimal portfolio choice problem in continuous time, allowing for a general Volterra propagator matrix  $G : [0, T]^2 \rightarrow \mathbb{R}^{N \times N}$  of nonnegative definite type as well as for general progressively measurable signals. Our optimal strategy is derived explicitly in Theorem 2.8 in terms of resolvents of the operators involved, and it is implemented in Section 3. Theorem 2.8 extends the results of [14], as it allows to solve the problem for a general propagator matrix, which includes the exponential decay kernels used in [14], as well as the power-law kernels from [5, 23] as specific examples. We also allow for general progressively measurable signals, which substantially generalize the mean-reverting signals used in [14] and the diffusion signals used in follow-up papers. Theorem 2.8 also generalizes the results of [3] in various directions. First we solve the problem in continuous time, which is compatible with the high-frequency trading timescale. We also include stochastic signals and risk factors, which are crucial for general portfolio choice problems, and turn them from being matrix-valued and deterministic as in [3], into being operator-valued and stochastic. Moreover, we allow for a general Volterra propagator matrix  $G(t, s)$  instead of a convolution propagator matrix  $G(t)$  as used in [3, 14], and provide explicit solutions in the general case, in contrast to the special case with an exponential propagator matrix, which was solved in [3]. As mentioned earlier, due to the generality of the propagator matrix, the portfolio choice problem that we solve is non-Markovian and time inconsistent, hence we introduce new tools, in the form of stochastic forward-backward systems of Fredholm equations of the second kind (see Section 4), in order to solve the corresponding first order conditions.
- (ii) Preventing price manipulation: in Theorem 2.14 we give sufficient conditions for a convolution propagator matrix  $G(t)$  to be nonnegative definite, which means that the expected costs caused by transient price impact are nonnegative for any trading strategy, so that in particular price manipulation in the sense of Huberman and Stanzl [19] is prevented (see (1.2)). Specifically, we show that if  $G$  is nonincreasing, convex, nonnegative and symmetric, then  $G$  is nonnegative definite, and price manipulations are excluded. Our result generalizes Theorem

2 of [3] from discrete to continuous time domains, and is of independent interest to the theory of nonnegative definite Volterra kernels. In Corollary 2.17 we address a popular example from the econophysics literature (see e.g. [23]). Namely, we prove that if the convolution propagator matrix is factorized as  $G(t) = C\phi(t)$ , where  $C$  is a symmetric nonnegative definite matrix and the function  $\phi : [0, T] \rightarrow \mathbb{R}$  is nonnegative, nonincreasing and convex on  $(0, T)$ , then  $G$  is nonnegative definite, hence price manipulations are excluded.

(iii) Insights on the influence of cross-impact: our numerical study in Section 3 explores the influence of transient cross-impact on the optimal liquidation of Asset 1, with an initial inventory of 0 in Asset 2. The conclusions can be summarized as follows:

- (a) In the absence of signals, cross-impact induces a ‘transaction-triggered’ round trip in Asset 2, prompting more aggressive trading in both assets. The strategy is particularly aggressive for the exponential propagator matrix compared to the more persistent fractional one (see Figure 2).
- (b) In the presence of positive alpha signals on both assets, cross-impact leverages the different alpha decays. The optimal strategy may involve selling or shorting the asset with the fastest alpha decay to leverage the cross-impact effect that decreases the price of the other asset with the slowest alpha decay. This allows profiting from both the more persistent signal on the second asset as well as the cross-impact effect (see Figures 3 and 4).

**Organization of the paper:** In Section 2 we present our main results regarding the solution to the portfolio choice problem and the prevention of price manipulation. Section 3 is dedicated to numerical illustrations of our main results. In Section 4 we derive explicit solutions to a class of systems of stochastic Fredholm equations arising from the first order condition in the proof of Theorem 2.8. Finally, Sections 5–7 are dedicated to the proofs of our main results.

## 2 Model Setup and Main Results

### 2.1 Model setup

Motivated by [14], we introduce in the following a variant of the optimal portfolio choice problem with transient price impact driven by a Volterra propagator matrix and in the presence of general alpha signals.

Let  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \leq t \leq T}, \mathbb{P})$  be a filtered probability space satisfying the usual conditions of right-continuity and completeness and let  $T > 0$  be a deterministic finite time horizon. We consider  $N \in \mathbb{N}$  risky assets whose unaffected price is given by an

$N$ -dimensional semimartingale  $P = (P_t)_{0 \leq t \leq T} = (P_t^1, \dots, P_t^N)_{0 \leq t \leq T}^\top$  with a canonical decomposition

$$P = A + M. \quad (2.1)$$

Here  $A = (A_t^1, \dots, A_t^N)_{0 \leq t \leq T}^\top$  is a predictable finite-variation process with

$$\mathbb{E} \left[ \int_0^T \|A_t\|^2 dt \right] < \infty,$$

and  $M = (M_t^1, \dots, M_t^N)_{0 \leq t \leq T}^\top$  is a continuous martingale such that

$$d[M^i, M^j]_t = \Sigma_{ij} dt, \quad \text{for } i, j = 1, \dots, N, \quad (2.2)$$

where  $\Sigma = (\Sigma_{ij}) \in \mathbb{R}^{N \times N}$  is a symmetric nonnegative definite covariance matrix and  $\|\cdot\|$  denotes the Euclidean norm.

We consider an investor whose initial portfolio is given by  $X_0 \in \mathbb{R}^N$ , where the amounts of shares held in each of the  $N$  risky assets are given by

$$X_t^u = X_0 + \int_0^t u_s ds, \quad 0 \leq t \leq T. \quad (2.3)$$

Here  $(u_s)_{0 \leq s \leq T} = (u_s^1, \dots, u_s^N)_{0 \leq s \leq T}^\top$  denotes the trading speed chosen by the investor from the set of admissible strategies

$$\mathcal{U} := \left\{ u : \Omega \times [0, T] \rightarrow \mathbb{R}^N \mid u \text{ progressively measurable with } \mathbb{E} \left[ \int_0^T \|u_t\|^2 dt \right] < \infty \right\}. \quad (2.4)$$

We assume that the investor's trading activity causes linear temporary impact on the assets' execution prices given by

$$\frac{1}{2} \Lambda u_t, \quad 0 \leq t \leq T, \quad (2.5)$$

where  $\Lambda \in \mathbb{R}^{N \times N}$  is a positive definite (not necessarily symmetric) matrix, i.e.  $x^\top \Lambda x > 0$  for all  $x \in \mathbb{R}^N$ . Note that the diagonal entries of  $\Lambda$  capture the temporary self-impact while the off-diagonal entries  $\Lambda_{ij}$  introduce temporary cross-impact on the price of asset  $i$  caused by trading one share of asset  $j$  per unit of time.

**Remark 2.1.** *Note that empirical findings in Capponi and Cont [8], Le Coz et al. [20] and Cont et al. [9] suggest that there is no significant temporary cross-impact of assets, i.e. that the off-diagonal entries of  $\Lambda$  are zero. Incorporating these findings into our model would turn  $\Lambda$  into a diagonal matrix with positive entries that capture the temporary self-impact of the assets à la Almgren and Chriss [4].*

The investor's trading activity also creates a price distortion which is given by

$$D_t^u := \int_0^t G(t, s) u_s ds, \quad 0 \leq t \leq T, \quad (2.6)$$

where  $G : [0, T]^2 \rightarrow \mathbb{R}^{N \times N}$  is a matrix of Volterra kernels, i.e. each entry in the matrix satisfies  $G_{ij}(t, s) = 0$  for  $t < s$ . The so-called propagator matrix  $G$  captures linear transient self-impact in the diagonal terms along with linear cross-impact in the off-diagonal terms. That is, the value  $G_{ij}(t, s)$  describes the impact at time  $t$  on the  $i$ -th asset's price caused by trading at time  $s$  one share of the  $j$ -th asset per unit of time. We further say that  $G$  is nonnegative definite, if it holds for every  $f \in L^2([0, T], \mathbb{R}^N)$  that

$$\int_0^T \int_0^T f(t)^\top G(t, s) f(s) ds dt \geq 0. \quad (2.7)$$

We define  $L^2([0, T]^2, \mathbb{R}^{N \times N})$  to be the space of Borel-measurable matrix-valued kernels  $G : [0, T]^2 \rightarrow \mathbb{R}^{N \times N}$  satisfying

$$\int_0^T \int_0^T \|G(t, s)\|^2 ds dt < \infty. \quad (2.8)$$

Here,  $\|\cdot\|$  denotes the Frobenius norm in consistency with our notation  $\|\cdot\|$  for the Euclidean norm. The set of admissible kernels  $\mathcal{G}$  is given by,

$$\mathcal{G} := \left\{ G \in L^2([0, T]^2, \mathbb{R}^{N \times N}) \mid G \text{ is nonnegative definite with } G(t, s) = 0 \text{ for } t < s \right\}. \quad (2.9)$$

**Example 2.2.** We present some typical examples for propagator matrices which arise from applications and can be incorporated into our model.

- (i) *Exponential decay:* Let  $C$  be a symmetric nonnegative definite matrix in  $\mathbb{R}^{N \times N}$ . Motivated by Obizhaeva and Wang [24], Alfonsi et al. [3] investigate the matrix exponential kernel

$$G(t, s) = \mathbf{1}_{\{t \geq s\}} \exp(-(t - s)C).$$

In [14], Gârleanu and Pedersen employ the factorized exponential kernel

$$G(t, s) = \mathbf{1}_{\{t \geq s\}} e^{-R(t-s)} C$$

with scalar price resiliency  $R > 0$ .

- (ii) *Power-law decay:* Benzaquen et al. [5] and Mastromatteo et al. [23] use for their empirical studies the factorized power-law kernel

$$G(t, s) = \mathbf{1}_{\{t \geq s\}} \left(1 + \frac{t - s}{t_0}\right)^{-\beta} C,$$



where  $C \in \mathbb{R}^{N \times N}$  is symmetric nonnegative definite,  $\beta \in (0, 1)$  and  $t_0 > 0$ . If  $\beta < \frac{1}{2}$ , our model also allows the factorized kernel

$$G(t, s) = \mathbf{1}_{\{t > s\}}(t - s)^{-\beta} C,$$

with singular power-law decay as introduced by Gatheral [15].

(iii) *Permanent impact: The constant kernel*

$$G(t, s) = \mathbf{1}_{\{t \geq s\}} C,$$

for a symmetric nonnegative definite matrix  $C \in \mathbb{R}^{N \times N}$  incorporates permanent self- and cross-impact as discussed in Section 5 of Huberman and Stanzl [19], Section 3 of Schneider and Lillo [25] and Example 1 of Alfonsi et al. [3]. If  $C$  is a diagonal matrix with nonnegative entries,  $G$  represents permanent impact à la Almgren and Chriss [4].

(iv) *Constructed decay kernels: Motivated by Section 3.1 of [3], the general kernel*

$$G(t, s) = Q^T \mathbf{1}_{\{t \geq s\}} \text{diag}(g_1(t - s), \dots, g_N(t - s)) Q$$

for an invertible matrix  $Q \in \mathbb{R}^{N \times N}$  and nonnegative, nonincreasing, convex kernel functions  $g_1, \dots, g_N : [0, T] \rightarrow \mathbb{R}$ , is nonnegative definite.

(v) *Interest rate derivatives: The following non-convolution kernel is a straight-forward generalization of the price impact for bonds trading model proposed in Section 3.1 of Brigo et al. [7], to a portfolio of bonds,*

$$G(t, s) = \alpha(T - t) \mathbf{1}_{\{t > s\}} H(t - s) C.$$

Here  $H : [0, T] \rightarrow \mathbb{R}$  is a nonnegative, nonincreasing, convex function taking values in  $\mathbb{R}$ , such as the exponential or power-law kernels in the above examples and  $C$  is a symmetric nonnegative definite matrix  $C \in \mathbb{R}^{N \times N}$ . The factor  $\alpha(T - t)$  for  $\alpha \in (0, \infty)$  is added in order to enforce a terminal condition on the bond price regarding its expiration at time  $T$ .

**Lemma 2.3.** *All propagator matrices introduced in Example 2.2 belong to the class of admissible kernels  $\mathcal{G}$ .*

Lemma 2.3 is proved in Section 7.

The investor's objective is the maximization of the following portfolio performance criterion over the time period  $[0, T]$ ,

$$J(u) := \mathbb{E} \left[ \int_0^T -u_t^\top (P_t + D_t^u + \frac{1}{2} \Lambda u_t) dt + (X_T^u)^\top P_T - \frac{\gamma}{2} \int_0^T (X_t^u)^\top \Sigma X_t^u dt \right] \quad (2.10)$$

over all admissible trading strategies  $u \in \mathcal{U}$ . The first two terms in the right-hand side of (2.10) describe the investor's terminal wealth, in terms of her final cash position resulting from trading the assets in the presence of price impact as prescribed above, as well as her remaining final portfolio's book value. The last term in (2.10) represents the portfolio's risk in the sense of Markowitz [22] (see also Section 1 of [14]), where  $\Sigma$  was defined in (2.2) and  $\gamma \geq 0$  is a risk aversion parameter.

**Remark 2.4.** *One may consider to implement in (2.10) the final portfolio's distorted value  $(X_T^u)^\top (P_T + D_T^u)$  instead of its final book value  $(X_T^u)^\top P_T$ . Recall the decomposition  $P = M + A$ . Hence in this case if  $A_t$  and  $D_t^u$  are both differentiable with respect to  $t$ , an application of integration by parts yields that maximizing  $J(u)$  is equivalent to maximizing the following cost functional,*

$$\tilde{J}(u) := \mathbb{E} \left[ \int_0^T \left( X_t^\top (\dot{A}_t + \dot{D}_t^u) - \frac{1}{2} u_t^\top \Lambda u_t dt - \frac{\gamma}{2} (X_t^u)^\top \Sigma X_t^u \right) dt \right],$$

which is a finite time horizon version of the objective functional in Gârleanu and Pedersen [14]. Note however that in this case  $\tilde{J}(\cdot)$  might not be concave for general Volterra kernels  $G$ , unlike for the exponential kernel case studied in [14].

## 2.2 Solution of the portfolio choice problem

Before we present our results regarding the solution to the portfolio choice problem we introduce some essential definitions and notation.

We denote the inner product on  $L^2([0, T], \mathbb{R}^N)$  by  $\langle \cdot, \cdot \rangle_{L^2}$ , i.e.

$$\langle f, g \rangle_{L^2} := \int_0^T f(t)^\top g(t) dt, \quad f, g \in L^2([0, T], \mathbb{R}^N),$$

and its induced norm by  $\| \cdot \|_{L^2}$ .

For any  $G \in L^2([0, T]^2, \mathbb{R}^{N \times N})$  (recall (2.8)) define the linear integral operator  $\mathbf{G}$  on  $L^2([0, T], \mathbb{R}^N)$  induced by  $G$  as

$$(\mathbf{G}f)(t) := \int_0^T G(t, s) f(s) ds, \quad 0 \leq t \leq T, \quad f \in L^2([0, T], \mathbb{R}^N).$$

Then  $\mathbf{G}$  is a bounded linear operator from  $L^2([0, T], \mathbb{R}^N)$  into itself (see Theorem 9.2.4 and Proposition 9.2.7 (iii) in [17]). Moreover, we denote by  $G^*$  the adjoint kernel of  $G$  with respect to  $\langle \cdot, \cdot \rangle_{L^2}$  given by

$$G^*(t, s) := G(s, t)^\top, \quad (s, t) \in [0, T]^2,$$

and by  $\mathbf{G}^*$  the induced adjoint bounded linear operator on  $L^2([0, T], \mathbb{R}^N)$ .

**Remark 2.5.** Note that if  $G \in L^2([0, T]^2, \mathbb{R}^{N \times N})$  then it holds that

$$\langle f, \mathbf{G}f \rangle_{L^2} = \langle f, \mathbf{G}^*f \rangle_{L^2} = \frac{1}{2} \langle f, (\mathbf{G} + \mathbf{G}^*)f \rangle_{L^2}, \quad \text{for all } f \in L^2([0, T], \mathbb{R}^N).$$

Moreover, recalling (2.7), the following three statements are equivalent:

- (i) The integral operator  $\mathbf{G}$  is nonnegative definite,
- (ii) The integral operator  $\mathbf{G}^*$  is nonnegative definite,
- (iii) The integral operator  $\mathbf{G} + \mathbf{G}^*$  is nonnegative definite,

**Definition 2.6.** Let  $h \in L^2([0, T], \mathbb{R}^{N \times N})$  denote a matrix-valued function in one variable and  $\mathbf{A}$  be a linear operator from  $L^2([0, T], \mathbb{R}^N)$  into itself. Then we define the column-wise application of  $\mathbf{A}$  to  $h$  as

$$\mathbf{A} \diamond h : [0, T] \rightarrow \mathbb{R}^{N \times N}, \quad (\mathbf{A} \diamond h)(r) := ((\mathbf{A}h_{\bullet 1})(r), \dots, (\mathbf{A}h_{\bullet N})(r)),$$

where

$$h_{\bullet i} : [0, T] \rightarrow \mathbb{R}^N, \quad h_{\bullet i}(r) := h(r)e_i,$$

is given by the  $i$ -th column of  $h$  for  $1 \leq i \leq N$ .

**Definition 2.7.** For a Volterra kernel  $G \in L^2([0, T]^2, \mathbb{R}^{N \times N})$  and a fixed  $t \in [0, T]$  define

$$G_t(s, r) := \mathbf{1}_{\{r \geq t\}} G(s, r),$$

and let  $\mathbf{G}_t$  denote the induced integral operator on  $L^2([0, T], \mathbb{R}^N)$ .

**Notation.** Recall that  $\Lambda$  was defined in (2.5). We denote by

$$\bar{\Lambda} := \frac{1}{2}(\Lambda + \Lambda^\top) \in \mathbb{R}^{N \times N} \tag{2.11}$$

the symmetric part of  $\Lambda$ , which is a positive definite matrix again.

Recall that  $\Sigma$  and  $\gamma$  were introduced in (2.10). For a Volterra kernel  $G$  as before, we introduce the following Volterra kernels  $F, K : [0, T]^2 \rightarrow \mathbb{R}^{N \times N}$ :

$$F(t, s) := \gamma \Sigma \mathbf{1}_{\{t > s\}}(T - t), \quad K(t, s) := G(t, s) + F(t, s). \tag{2.12}$$

We denote by  $\mathbb{E}_t[\cdot]$  the conditional expectation with respect to  $\mathcal{F}_t$ .

Finally, we define the following linear operator,

$$\mathbf{D} := \mathbf{G} + \mathbf{G}^* + \mathbf{F} + \mathbf{F}^* + \bar{\Lambda} \mathbf{I}, \tag{2.13}$$

where  $\mathbf{I}$  is the identity operator on  $L^2([0, T], \mathbb{R}^N)$ , and the stochastic process

$$g_t := \mathbb{E}_t[P_T - P_t] - \gamma(T - t)\Sigma X_0, \quad 0 \leq t \leq T, \tag{2.14}$$

where  $X_0$  is given in (2.3).

Recall that the class of admissible propagator matrices  $\mathcal{G}$  was defined in (2.9). In the following theorem we derive the unique maximizer of the portfolio's revenue-risk functional (2.10) in terms of resolvents. Recall that the resolvent of a Volterra operator  $\mathbf{B}$  is given by  $\mathbf{R}^B = \mathbf{I} - (\mathbf{I} + \mathbf{B})^{-1}$ .

**Theorem 2.8.** *Let  $G \in \mathcal{G}$ . Then the unique maximizer  $u^* \in \mathcal{U}$  of the objective functional  $J(u)$  in (2.10) is given by*

$$u_t^* = ((\mathbf{I} + \mathbf{B})^{-1}a)(t), \quad 0 \leq t \leq T,$$

where

$$a_t := \bar{\Lambda}^{-1} \left( g_t - \int_t^T K(r, t)^\top (\mathbf{D}_t^{-1} \mathbf{1}_{\{t \leq \cdot\}} \mathbb{E}_t[g]) (r) dr \right),$$

$$B(t, s) := -\bar{\Lambda}^{-1} \left( \mathbf{1}_{\{s \leq t\}} \int_t^T K(r, t)^\top (\mathbf{D}_t^{-1} \diamond \mathbf{1}_{\{t \leq \cdot\}} K(\cdot, s)) (r) dr - K(t, s) \right),$$

where  $K$ ,  $\mathbf{D}$  and  $g$  are defined in (2.12), (2.13) and (2.14) respectively, and  $\mathbf{B}$  is the integral operator induced by the kernel  $B$ .

In the following corollary we derive the maximizer of (2.10) for the case of a deterministic signal  $A$  in (2.1). This assumption considerably simplifies the optimal portfolio choice and is of relevance for practical applications.

**Corollary 2.9.** *Let  $G \in \mathcal{G}$  and assume that  $A$  in (2.1) is deterministic. Then, the unique maximizer  $u^* \in \mathcal{U}$  of the objective functional  $J(u)$  in (2.10) is given by*

$$u_t^* = (\mathbf{D}^{-1}g)(t), \quad 0 \leq t \leq T,$$

where  $\mathbf{D}$  is defined in (2.13) and  $g$  in (2.14) becomes the deterministic function in  $L^2([0, T], \mathbb{R}^N)$  given by

$$g_t = A_T - A_t - \gamma(T - t)\Sigma X_0.$$

The proofs of Theorem 2.8 and Corollary 2.9 are given in Section 5.

In the following remarks we give additional details regarding the contribution and implementation of our results.

**Remark 2.10.** *If the investor's objective is to optimally liquidate the portfolio, an additional terminal penalty on the remaining inventory can be added to (2.10). Specifically, we define the revenue-risk functional as follows:*

$$\bar{J}(u) = \mathbb{E} \left[ \int_0^T -u_t^\top (P_t + D_t^u + \frac{1}{2} \Lambda u_t) dt + (X_T^u)^\top P_T - \frac{\gamma}{2} \int_0^T (X_t^u)^\top \Sigma X_t^u dt - \frac{\rho}{2} (X_T^u)^\top \Pi X_T^u \right],$$

where in the last term of  $\bar{J}(u)$ ,  $\varrho$  is a positive constant and  $\Pi \in \mathbb{R}^{N \times N}$  is a symmetric nonnegative definite matrix. In this case the result of Theorem 2.8 applies with a slight modification by setting:

$$F(t, s) := \mathbf{1}_{t>s}(\gamma(T-t)\Sigma + \varrho\Pi),$$

and

$$g_t := \mathbb{E}_t[P_T - P_t] - (\gamma(T-t)\Sigma + \varrho\Pi)X_0. \quad (2.15)$$

**Remark 2.11.** In Section 3 we provide a numerical implementation for the optimal strategy which is derived in Theorem 2.8 and Corollary 2.9. Our implementation includes examples of optimal portfolio choice and of optimal liquidation for various choices of signals and propagator matrices.

**Remark 2.12.** Theorem 2.8 extends the results of Proposition 2 in [14], as it allows to solve the problem for general propagator matrices as in (2.9), which include the exponential decay kernels used in [14], as well as the power-law kernels from [5, 23] and other specific examples outlined earlier in this section. We also allow for general semimartingale signals, which substantially generalize the integrated Ornstein-Uhlenbeck signals used in [14]. Theorem 2.8 also generalizes the results of [3] in the following directions: First, we solve the problem in continuous time, which is compatible with the high-frequency trading timescale. We also include stochastic signals and risk factors in the objective functional (2.10), which are crucial for portfolio choice problems. This turns the portfolio choice problem from being matrix-valued and deterministic as in [3] into an operator-valued stochastic control problem. We also allow for a general Volterra propagator matrix  $G(t, s)$  instead of the convolution propagator matrix  $G(t)$  in [3], and provide explicit solutions, in contrast to the first order condition derived in Theorem 3 of [3], which is solely solved for exponential propagator matrices (see Example 2 therein). In particular, in the proof of Theorem 2.8, we characterize the first order condition corresponding to the objective functional (2.10) in terms of a system of coupled stochastic Fredholm equations of the second kind with both forward and backward components. In Section 4 we develop a method for solving this system explicitly, which is a central ingredient for the proof of Theorem 2.8.

**Remark 2.13.** In the single asset case where  $N = 1$ , Theorem 2.8 generalizes the main results of both [1] and [2] to a larger class of admissible kernels  $\mathcal{G}$ . Indeed, the proof of Theorem 2.8 relies heavily on the theory of kernels of type  $L^2$  and their resolvents. See in particular the proofs of Proposition 4.1 and Lemma 4.8. The method of the proof allows us to extend the class of admissible kernels  $\mathcal{G}$  such that only square-integrability of the kernels is needed (see (2.9) and (2.8)), instead of the more restrictive assumption that

$$\sup_{t \leq T} \int_0^T |G(t, s)|^2 ds + \sup_{s \leq T} \int_0^T |G(t, s)|^2 dt < \infty,$$

which was assumed by Abi Jaber and Neuman [1] and Abi Jaber et al. [2] who studied the single asset case. In particular our new approach can be applied to prove Proposition 4.5 and Lemma 7.1 in [1] as well as Proposition 5.1 and Lemma 5.5 in [2] under the more general integrability assumption on the admissible kernels  $G$  in  $\mathcal{G}$  given by

$$\int_0^T \int_0^T |G(t, s)|^2 ds dt < \infty.$$

## 2.3 Results on price manipulations

In this section we address the following question:

*How to choose the propagator matrix  $G$  in order to prevent price manipulation?*

We recall that for a kernel  $G \in L^2([0, T]^2, \mathbb{R}^{N \times N})$  the expected costs caused by transient impact induced by a strategy  $u \in \mathcal{U}$  are given by

$$C(u) := \mathbb{E} \left[ \int_0^T \int_0^T u(t)^\top G(t, s) u(s) ds dt \right]. \quad (2.16)$$

Note that (2.16) can be obtained by considering strategies with fuel constraints (i.e.  $X_T = 0$ ), setting the risk aversion term  $\gamma$  and temporary price impact matrix  $\Lambda$  in (2.10) to 0 and by choosing a martingale unaffected price process  $P = M$  which prevents arbitrage opportunities (see Section 2.1 of [21] for the derivation). Hence the condition that a propagator matrix  $G \in \mathcal{G}$  satisfies (2.7) rules out the possibility of price manipulations in the sense of [19]. This means that a portfolio starting and terminating with zero inventory ( $X_0 = X_T = 0$ ) cannot create any profit, or equivalently negative trading costs, so we must have

$$C(u) \geq 0.$$

Put differently, a price manipulation in this context is a round trip strategy with positive expected profit, which in the absence of trading signals, is only possible if (2.16) is negative. We refer to Section 2 of Alfonsi et al. [3] for the price impact manipulation condition in discrete time portfolio choice problems. Interestingly enough, assumption (2.7) is also needed in order to prove that the objective functional (2.10) is concave (see Lemma 5.1), hence it is essential for the proof of Theorem 2.8.

Note that condition (2.7) is not straightforward to verify. The main result of this section derives sufficient conditions for a large class of convolution kernels to satisfy (2.7). We recall that in the single asset case, convolution kernels of the form  $G(t, s) := \mathbb{1}_{\{t \geq s\}} H(t-s)$  are nonnegative definite kernels whenever the real function  $H$  on  $[0, T]$  is nonnegative, nonincreasing and convex (see Example 2.7 in [16]). However in the multi-asset case deriving a characterization for nonnegative definite kernels is substantially more involved. Recall that the class of admissible Volterra kernels  $\mathcal{G}$  was defined in (2.9).

**Theorem 2.14.** Let  $H : [0, T] \rightarrow \mathbb{R}^{N \times N}$  be a convolution kernel so that the associated Volterra kernel  $G(t, s) := \mathbf{1}_{\{t \geq s\}} H(t - s)$  is in  $L^2([0, T]^2, \mathbb{R}^{N \times N})$ . Assume that  $H$  satisfies the following assumptions:

- nonincreasing, i.e. the function  $t \mapsto x^\top H(t)x$  is nonincreasing on  $(0, T)$  for every  $x \in \mathbb{R}^N$ ,
- convex, i.e. the function  $t \mapsto x^\top H(t)x$  is convex on  $(0, T)$  for every  $x \in \mathbb{R}^N$ ,
- nonnegative, i.e. the matrix  $H(t)$  is nonnegative definite for any  $t \in (0, T)$ ,
- symmetric, i.e. the matrix  $H(t)$  is symmetric for any  $t \in (0, T)$ .

Then  $H$  is nonnegative definite and  $G$  is in the class of admissible kernels  $\mathcal{G}$ , i.e.

$$\int_0^T \int_0^t f(t)^\top H(t - s) f(s) ds dt = \langle f, \mathbf{G}f \rangle_{L^2} \geq 0, \quad \text{for all } f \in L^2([0, T], \mathbb{R}^N).$$

The proof of Theorem 2.14 is given in Section 6.

**Remark 2.15.** The convexity of  $H$  implies that the function  $t \mapsto x^\top H(t)x$  is continuous on  $(0, T)$  for any  $x \in \mathbb{R}^N$ . Thus, it follows from the symmetry of  $H$  that it is continuous on  $(0, T)$ . Nevertheless, Theorem 2.14 permits kernels  $H$  which have a singularity at  $t = 0$  such as the power-law propagator in Example 2.2(ii).

**Remark 2.16.** Theorem 2.14 generalizes the result of Theorem 2 of [3] from discrete time grids to a continuous domain, which is of independent interest to the theory of nonnegative definite Volterra kernels. One of the main ingredients of the proof is Proposition 6.3, which states that a matrix-valued Volterra kernel which is nonnegative definite in the discrete sense (see (6.1)) is also nonnegative definite on a continuous domain, i.e. it satisfies assumption (2.7).

The following corollary rules out the possibility of price manipulation in the sense of [19] for a specific class of factorized convolution propagator matrices introduced by Benzaquen et al. [5], Mastromatteo et al. [23].

**Corollary 2.17.** Let  $C \in \mathbb{R}^{N \times N}$  be a symmetric nonnegative definite matrix and  $\phi \in L^2([0, T], \mathbb{R})$  be nonnegative, nonincreasing and convex on  $(0, T)$ . Then the convolution kernel

$$H : [0, T] \rightarrow \mathbb{R}^{N \times N}, \quad H(t) := C\phi(t) \tag{2.17}$$

is nonnegative definite and thus its associated Volterra kernel is in  $\mathcal{G}$ .

The proof of Corollary 2.17 is postponed to Section 6.

**Remark 2.18.** From Corollary 2.17 it follows that our model accommodates the factorized model introduced in [5, 23], where empirical evidence for the explanatory relevance of cross-impact for the cross-correlation of US stocks was given. The authors found that as a good and efficient approximation the price impact kernel can be written as  $G(t, s) = C\phi(t - s)$  for a symmetric nonnegative definite matrix  $C \in \mathbb{R}^{N \times N}$  and a power-law function

$$\phi(t) = (1 + t/t_0)^{-\beta} \mathbf{1}_{\{t \geq 0\}},$$

where  $\beta \in (0, 1)$  and  $t_0 > 0$  are estimated model parameters. The main conclusions of their papers are also briefly summarized in Section 14.5.3 of [6].

### 3 Numerical Illustrations

#### 3.1 Optimal liquidation: what is the influence of cross-impact?

In this section, we explore the influence of transient cross-impact for two assets in a liquidation problem with time horizon  $T = 10$ , where Asset 1 has 10 shares initial inventory and Asset 2 starts from zero inventory, i.e.  $X_0 = (10, 0)^\top$ . Our objective is to elucidate the influence of cross-impact on optimal trading speeds and inventories and its interplay with trading signals (also called alphas). To isolate these effects, we consider a penalty on terminal inventory using the parameters  $(\varrho, \Pi)$  of the form

$$\varrho = 4, \quad \Pi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and we set  $\gamma = 0$ , removing the risk component in the objective functional  $\bar{J}$  in Remark 2.10 originating from the covariance matrix between the assets.

We consider temporary self-impact but no temporary cross-impact through the  $2 \times 2$ -matrix  $\Lambda$ :

$$\Lambda = \begin{pmatrix} 0.03 & 0 \\ 0 & 0.03 \end{pmatrix}. \tag{3.1}$$

The transient price impact is given by the parsimonious propagator from [5, 23] (see Corollary 2.17) of the form

$$G(t, s) = C\phi(t - s)\mathbf{1}_{\{s < t\}}, \tag{3.2}$$

with a deterministic symmetric  $2 \times 2$ -matrix  $C$  that will be taken to be either diagonal or non-diagonal to study the impact of cross-impact (see (3.3)-(3.4) below), and a scalar convolution kernel  $\phi$  which will be either set to  $\phi^{\text{zero}}(t) = 0$ , exponential  $\phi^{\text{exp}}(t) = e^{-\rho t}$  with  $\rho = 0.5$  or fractional  $\phi^{\text{frac}}(t) = t^{-\alpha}$  with  $\alpha = 0.25$ .

We analyze four figures to illustrate our findings:



1. **Figure 1:** No trading signal, no cross-impact with

$$C = \begin{pmatrix} 0.06 & 0 \\ 0 & 0.06 \end{pmatrix}. \quad (3.3)$$

2. **Figures 2:** No trading signal, with cross-impact with

$$C = \begin{pmatrix} 0.06 & 0.05 \\ 0.05 & 0.06 \end{pmatrix}. \quad (3.4)$$

3. **Figures 3-4:** With ‘noisy’ trading signals of different decay  $\beta$  as follows:

$$dP_t = I_t dt + dM_t, \quad (3.5)$$

with  $M$  a martingale and

$$dI_t = -\beta I_t + dW_t, \quad (3.6)$$

where  $W$  is a Brownian motion and

$$I_0 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix}. \quad (3.7)$$

We consider the cases without and with cross-impact given in (3.3) and (3.4) (respectively).

Our conclusions are summarized as follows:

1. In the absence of trading signals (see Figures 1 and 2), cross-impact induces:
  - a ‘transaction-triggered’ round trip in Asset 2 on the bottom panels of Figure 2 which was absent from Figure 1. Initiating fast liquidation of Asset 1 at time  $t = 0^+$  triggers a drop in Asset 2’s price due to  $C_{12} > 0$ , prompting a ‘transaction-triggered’ shorting signal’ on Asset 2. Therefore, the optimal strategy starts by shorting Asset 2. After a while, as the selling speed of Asset 1 slows down, the strategy starts buying Asset 2 at a steady rate. This strategic move aims at increasing the price of Asset 1 undergoing liquidation, using the positive cross-impact term  $C_{21} > 0$ . This steady buying even results in a long position in Asset 2 which is then liquidated quickly near the maturity. The strategy is more aggressive for the exponential kernel compared to the fractional one.
  - a more aggressive trading in Asset 1 even in the presence of transient impact. Inventories with transient impact align closer to those without transient impact on the top panels of Figure 2 than on those of Figure 1.

2. In the presence of positive trading signals on both assets with different alpha decays (see Figures 3 and 4):

- **Without cross-impact (solid lines):** in both figures, the positive signal for Asset 1 prompts from the start a purchasing of Asset 1 even in the presence of transient impact. The ‘buy’ strategy is more aggressive in the absence of transient impact. Similarly, the positive signal of Asset 2 triggers a ‘buy round trip’ on Asset 2.
- **With cross-impact (dashed lines):** in Figure 3 with  $(\beta_1, \beta_2) = (0.9, 0.3)$  in (3.7) we observe that in contrary to the no cross-impact case, there is almost no buying of Asset 1 at  $t = 0^+$ , even in the presence of the positive signal for Asset 1. Indeed, since the ‘buy’ signal for Asset 1 decays more rapidly compared to that of Asset 2, recall the mean-reversion coefficients  $\beta$  in (3.7), the optimal strategy starts selling Asset 1 quickly (and even shorts it) in order to leverage the cross-impact that will decrease the price of Asset 2 that has a more persistent positive signal. Hence, this strategy allows profiting from both the persistent signal on Asset 2 and from the cross-impact effect. More of Asset 2 is bought compared with the case without cross-impact.

In Figure 4 with  $(\beta_1, \beta_2) = (0.3, 0.9)$  in (3.7) we observe the opposite effect, since now Asset 1 has the more persistent signal.

Without Cross Impact, Without Signal

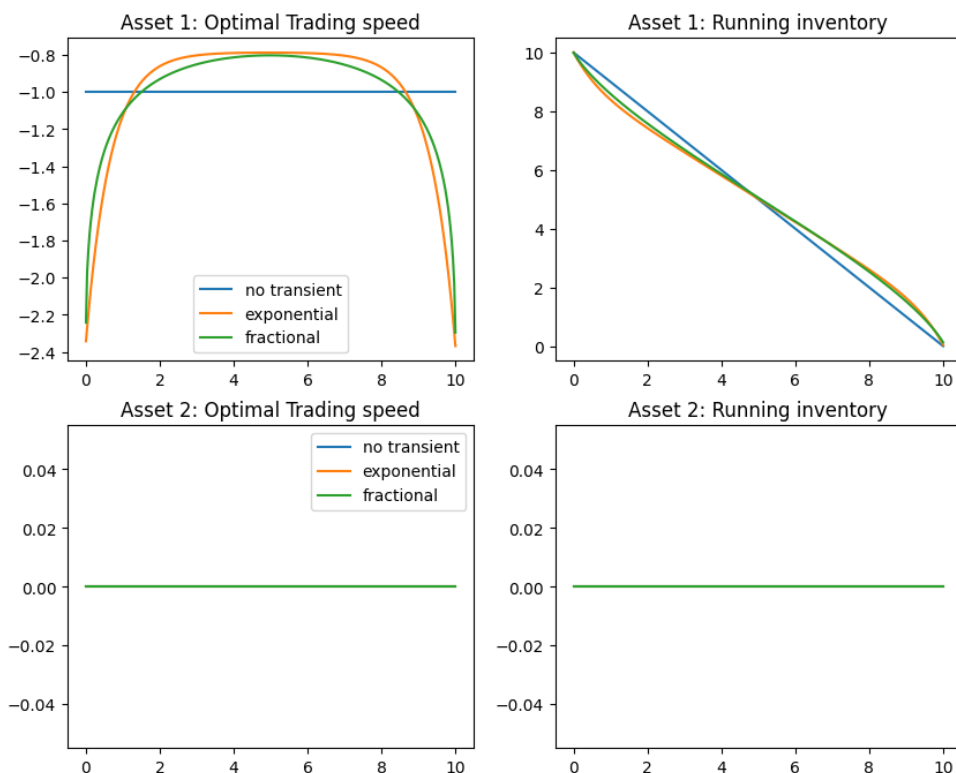


Figure 1: Impact of different kernels on the optimal trading speed and inventory of two assets in the absence of a signal **without cross-impact**  $C = \begin{pmatrix} 0.06 & 0 \\ 0 & 0.06 \end{pmatrix}$  with the three impact kernels:  $\phi^{\text{zero}}$  (blue),  $\phi^{\text{exp}}$  (yellow) and  $\phi^{\text{frac}}$  (green).

### 3.2 Optimal trading with frictions

In this section, we illustrate the influence of transient impact on an portfolio choice example with two independent assets. Here we set  $\varrho = 0$  as in the objective functional (2.10) (i.e. no penalty on terminal inventory) and we set the risk aversion coefficient to  $\gamma = 5$ . We look at the time evolution of holdings in each asset in comparison with the Markowitz portfolio, i.e. the zero temporary and transient price impact case. This experiment has already appeared in [14] for the exponential kernel.

We set  $T = 10$ ,  $X_0 = (7.5, -7.5)^\top$  with two independent assets with a  $2 \times 2$ -covariance matrix  $\Sigma$  given by

$$\Sigma = \begin{pmatrix} 0.04 & 0 \\ 0 & 0.05 \end{pmatrix},$$

With Cross Impact, Without Signal

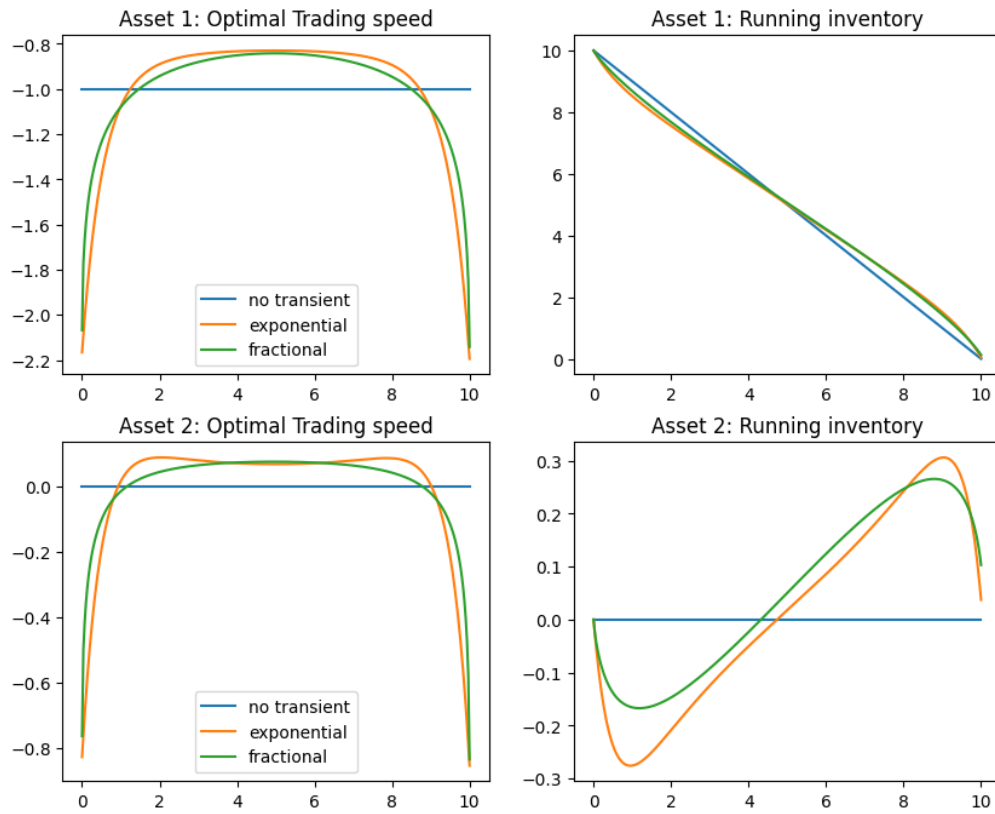


Figure 2: Impact of different kernels on the optimal trading speed and inventory of two assets in the absence of a signal **with cross-impact**  $C = \begin{pmatrix} 0.06 & 0.05 \\ 0.05 & 0.06 \end{pmatrix}$  with the three impact kernels:  $\phi^{\text{zero}}$  (blue),  $\phi^{\text{exp}}$  (yellow) and  $\phi^{\text{frac}}$  (green).

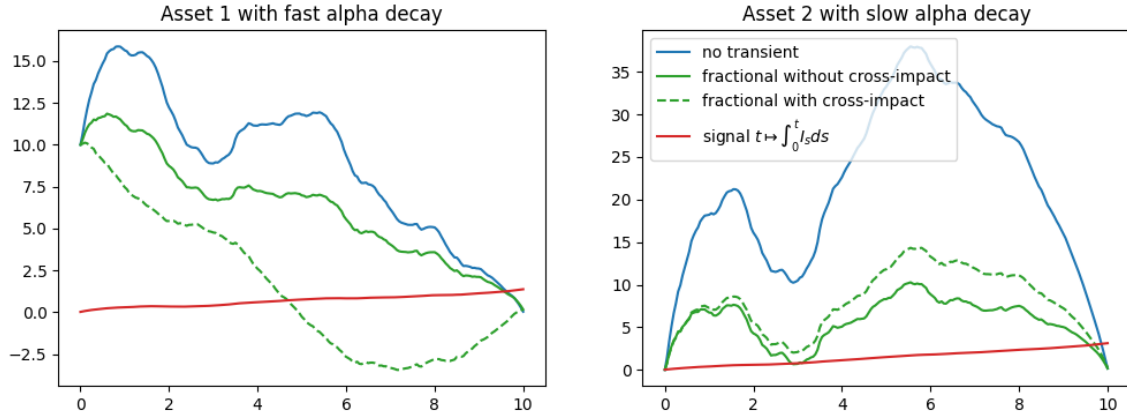


Figure 3: Impact of the fractional transient cross-impact on the optimal inventory of two assets in the presence of buy trading signals with different alpha decay on each asset given by  $(\beta_1, \beta_2) = (0.9, 0.3)$  in (3.7) **without (solid) and with (dashed) cross-impact** with the impact kernels:  $\phi^{\text{zero}}$  (blue) and  $\phi^{\text{frac}}$  (green).

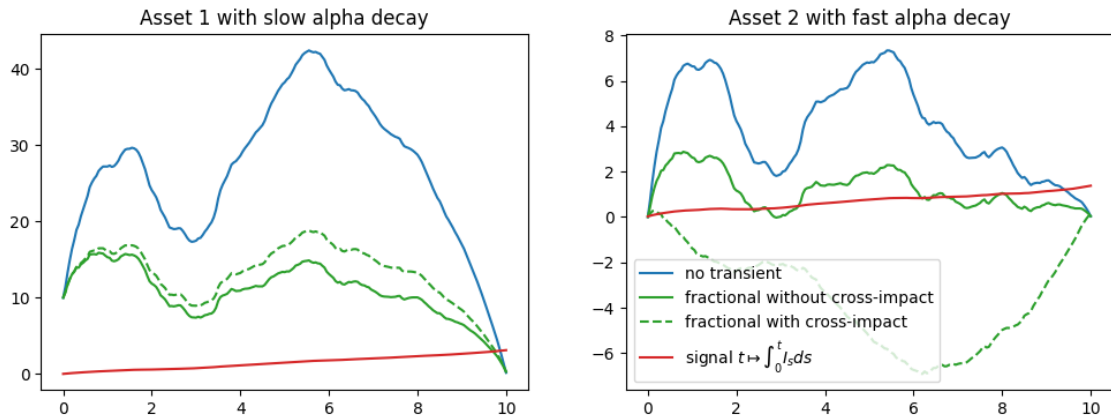


Figure 4: Impact of the fractional transient cross-impact on the optimal inventory of two assets in the presence of buy trading signals with different alpha decay on each asset given by  $(\beta_1, \beta_2) = (0.3, 0.9)$  in (3.7) **without (solid) and with (dashed) cross-impact** with the impact kernels:  $\phi^{\text{zero}}$  (blue) and  $\phi^{\text{frac}}$  (green).

and a stochastic trading signal of the form (3.5)-(3.6) with

$$I_0 = \begin{pmatrix} 0.01 \\ -0.01 \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} 0.05 & 0 \\ 0 & 0.3 \end{pmatrix}.$$

The temporary impact  $\Lambda$  is given by (3.1) and the transient impact is given by (3.2), where we do not consider cross-impact, i.e.  $C$  is given by (3.4) and  $\phi$  is chosen as in the previous section.

We make precise the notion of the Markowitz portfolio in this context.

**Remark 3.1.** *In the absence of any market frictions, i.e.  $\Lambda = G = 0$ , for an unaffected price of the form*

$$dP_t = I_t dt + dM_t,$$

*with  $M$  a martingale, the optimal position is given by the celebrated Markowitz portfolio*

$$X_t^{\text{Markowitz}} = \frac{\Sigma^{-1} I_t}{\gamma}. \quad (3.8)$$

*To see this it suffices to set  $\Lambda = G = 0$  in the functional  $J$  in (2.10) and apply an integration by parts to get:*

$$\mathbb{E} \left[ P_T^\top X_T - \int_0^T P_t^\top u_t dt \right] = X_0^\top P_0 + \mathbb{E} \left[ \int_0^T I_t^\top X_t dt \right],$$

*which would then yield the functional*

$$X_0^\top P_0 + \mathbb{E} \left[ \int_0^T \left( I_t^\top X_t - \frac{\gamma}{2} X_t^\top \Sigma X_t \right) dt \right]$$

*and the optimality the Markowitz portfolio (3.8).*

The optimal positions in Assets 1 and 2 without cross-impact are illustrated in Figure 5. The Markowitz portfolio is just a reference portfolio that describes the weights of the optimal portfolio derived from the optimal trade-off between risk  $\Sigma$  and expected excess return  $I_t$  scaled by  $\gamma$ . We make the following observations:

- Similarly to Gârleanu and Pedersen [14], we observe that if one chooses any initial starting portfolio  $X_0$ , all optimal positions in the presence of frictions try to follow and get closer to the Markowitz portfolio in a smooth way.
- The ‘speed of convergence’ towards the Markowitz portfolio seems to depend on the type of the frictions, being faster for the no-transient kernel than for the exponential kernel, and slowest for the more persistent fractional case.

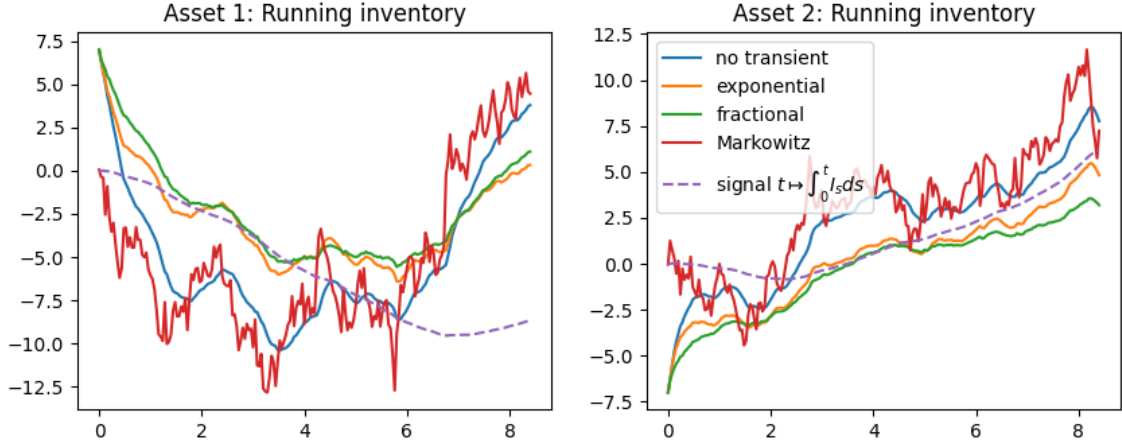


Figure 5: Optimal trading without cross-impact with two independent assets and stochastic signals with the three impact kernels:  $\phi^{\text{zero}}$  (blue),  $\phi^{\text{exp}}$  (yellow) and  $\phi^{\text{frac}}$  (green) and the Markowitz portfolio (red) in the absence of any friction as in (3.8) with a stochastic signal (purple).

### 3.3 Numerical scheme

A major advantage of the operator formulas appearing in Theorem 2.8 is their ease of implementation. One possibility is the use of the so-called Nyström method to discretize the operators, and the method described in Section 5.1 of [1] for the single asset case can be readily adapted to our multi-asset setting.

For completeness, we briefly describe the numerical implementation for deterministic signals as in Corollary 2.9.

Fix  $N$  assets,  $n \in \mathbb{N}$  and a partition  $0 = t_0 < t_1 < t_2 < \dots < t_n = T$  of  $[0, T]$  as well as a propagator matrix in the form (3.2). We set  $\Delta t := T/n$ .

**Step 1.** Specify the signal  $P = \int_0^\cdot I_s ds + M$  for a deterministic function  $I$  and compute the  $N(n+1)$ -vector,

$$g_n := (g(t_0)^\top, g(t_1)^\top, \dots, g(t_n)^\top)^\top,$$

using the expression of  $g$  in (2.15). (Note that  $g(t_i) \in \mathbb{R}^N$ .) Second, specify the scalar convolution kernel  $\phi : [0, T] \rightarrow \mathbb{R}$  that appears in (3.2) and set  $F_1(t, s) = \mathbb{1}_{\{t>s\}}(T-t)$  and  $F_2(t, s) = \mathbb{1}_{\{t>s\}}$ . Define the following lower and upper triangular  $(n+1) \times (n+1)$ -

matrices  $L_1, L_2, L$  and  $U_1, U_2, U$  where the nonzero elements are given by:

$$\begin{aligned}
L_1^{kj} &= \int_{t_j}^{t_{j+1}} F_1(t_k, s) ds = (T - t_k)\Delta t, \quad k = 0, \dots, n, \quad j = 0, \dots, (k - 1), \\
U_1^{kj} &= \int_{t_j}^{t_{j+1}} F_1(s, t_k) ds \approx (T - t_j)\Delta t, \quad k = 0, \dots, n, \quad j = k, \dots, (n - 1), \\
L_2^{kj} &= \int_{t_j}^{t_{j+1}} F_2(t_k, s) ds = \Delta t, \quad k = 0, \dots, n, \quad j = 0, \dots, (k - 1), \\
U_2^{kj} &= \int_{t_j}^{t_{j+1}} F_2(s, t_k) ds = \Delta t, \quad k = 0, \dots, n, \quad j = k, \dots, (n - 1), \\
L^{kj} &= \int_{t_j}^{t_{j+1}} \phi(t_k - s) ds, \quad k = 0, \dots, n, \quad j = 0, \dots, (k - 1), \\
U^{kj} &= \int_{t_j}^{t_{j+1}} \phi(s - t_k) ds, \quad k = 0, \dots, n, \quad j = k, \dots, (n - 1).
\end{aligned}$$

For instance, for the three particular kernels in the previous subsection,  $L$  and  $U$  admit explicit expressions collected in Table 1.

	$\phi(t)$	$L^{kj}$ for $0 \leq j \leq k - 1$	$U^{kj}$ for $k \leq j \leq n - 1$
No-transient	0	0	0
Exponential	$ce^{-\rho t}$	$c \frac{e^{\rho \Delta t} - 1}{\rho} e^{-\rho(k-j)\Delta t}$	$c \frac{1 - e^{-\rho \Delta t}}{\rho} e^{-\rho(j-k)\Delta t}$
Fractional	$ct^{-\alpha}$	$\frac{c(\Delta t)^{1-\alpha}}{1-\alpha} ((k-j)^{1-\alpha} - (k-j-1)^{1-\alpha})$	$\frac{c(\Delta t)^{1-\alpha}}{1-\alpha} ((j+1-k)^{1-\alpha} - (j-k)^{1-\alpha})$

Table 1: Some kernels  $\phi$  and the corresponding explicit nonzero elements of the matrices  $L$  and  $U$ .

**Step 2.** Construct the  $N(n+1) \times N(n+1)$ -matrix  $D_n$  using

$$D_n = I_{(n+1) \times (n+1)} \otimes \bar{\Lambda} + L \otimes C + U \otimes C + \gamma L_1 \otimes \Sigma + \gamma U_1 \otimes \Sigma + \rho L_2 \otimes \Pi + \rho U_2 \otimes \Pi,$$

where  $\otimes$  is the Kronecker product.

**Step 3.** Recover the  $N(n+1)$ -vector for the optimal control path

$$u_n = (D_n)^{-1} g_n$$

and note that the first  $N$  components are the first trading values for all assets, the next  $N$  components are the second trading values for all assets etc.



## 4 Systems of Stochastic Fredholm Equations

In this section we derive solutions to a class of coupled stochastic Fredholm equations of the second kind with both forward and backward components, which arise from the first order condition in the proof of Theorem 2.8 in Section 5. The main result of this section is given in the following proposition. Recall that the class of admissible propagator matrices  $\mathcal{G}$  was defined in (2.9) and that the  $\diamond$  operation was defined in Definition 2.6. We also recall for any kernel  $G \in \mathcal{G}$ , the kernel  $G_t$  is the truncation of  $G$  from Definition 2.7 and that  $\mathbf{G}_t$  is the operator induced by the kernel  $G_t$ .

**Proposition 4.1.** *Let  $K, L$  be Volterra kernels in  $\mathcal{G}$ , let  $\bar{\Lambda} \in \mathbb{R}^{N \times N}$  be a symmetric positive definite matrix, and let  $f = (f_t)_{0 \leq t \leq T}$  be an  $N$ -dimensional progressively measurable process with  $\int_0^T \mathbb{E}[\|f_t\|^2] dt < \infty$ . Then the following coupled system of stochastic Fredholm equations*

$$\bar{\Lambda} u_t = f_t - \int_0^t K(t, r) u_r dr - \int_t^T L(r, t)^\top \mathbb{E}_t[u_r] dr, \quad t \in [0, T], \quad (4.1)$$

admits a unique progressively measurable solution  $u^* = (u_t^*)_{0 \leq t \leq T}$  in  $\mathcal{U}$  given by

$$u_t^* = ((\mathbf{I} + \mathbf{B})^{-1} a)(t), \quad t \in [0, T],$$

where

$$\begin{aligned} a_t &= \bar{\Lambda}^{-1} \left( f_t - \int_t^T L(r, t)^\top (\mathbf{D}_t^{-1} \mathbb{1}_{\{t \leq \cdot\}} \mathbb{E}_t[f \cdot])(r) dr \right), \\ B(t, s) &= -\bar{\Lambda}^{-1} \left( \mathbb{1}_{\{s \leq t\}} \int_t^T L(r, t)^\top (\mathbf{D}_t^{-1} \diamond \mathbb{1}_{\{t \leq \cdot\}} K(\cdot, s))(r) dr - K(t, s) \right), \\ \mathbf{D}_t &= \bar{\Lambda} \mathbf{I} + \mathbf{K}_t + \mathbf{L}_t^*. \end{aligned}$$

Here  $\mathbf{I}$  is the identity operator on  $L^2([0, T], \mathbb{R}^N)$  and  $\mathbf{B}$  is the integral operator induced by the kernel  $B$ .

**Remark 4.2.** *Proposition 4.1 is a generalization of Proposition 5.1 in [2] from the class of one-dimensional stochastic Fredholm equations to  $N$ -dimensional coupled systems. Note that in the one-dimensional case  $\bar{\Lambda}$  is just a positive scalar, so that the integral operators  $\bar{\Lambda}^{-1} \mathbf{K}_t$  and  $\bar{\Lambda}^{-1} \mathbf{L}_t^*$  are nonnegative definite and therefore the operator  $\mathbf{I} + \bar{\Lambda}^{-1} (\mathbf{K}_t + \mathbf{L}_t^*)$  invertible, which is an important ingredient in the derivation of the solution. However, since the composition of two general nonnegative definite linear operators on a Hilbert space is not necessarily nonnegative definite, even if they are self-adjoint, this argument does not go through in the  $N$ -dimensional case. Thus, the positive definite operator  $\mathbf{D}_t = \bar{\Lambda} \mathbf{I} + \mathbf{K}_t + \mathbf{L}_t^*$  is inverted as part of the solution, and  $\bar{\Lambda}^{-1}$  appears as part of the terms  $a$  and  $B$ .*

As a preparation for the the proof, we introduce the concept of kernels of type  $L^2$  following the terminology of Definition 9.2.2 in [17].

**Definition 4.3.** A Borel-measurable function  $G : [0, T]^2 \rightarrow \mathbb{R}^{N \times N}$  is called a kernel of type  $L^2$  if it satisfies

$$\|G\|_{L^2} := \sup_{\substack{f: \|f\|_{L^2} \leq 1 \\ g: \|g\|_{L^2} \leq 1}} \int_0^T \int_0^T \|g(t)G(t, s)f(s)\| ds dt < \infty,$$

where the supremum is taken over all real-valued functions  $f, g \in L^2([0, T], \mathbb{R})$  with norm bounded by 1 and  $\|\cdot\|$  denotes the Frobenius norm (following the convention introduced after (2.8)).

Every kernel in  $L^2([0, T]^2, \mathbb{R}^{N \times N})$  is also a kernel of type  $L^2$  (see Proposition 9.2.7 (iii) in [17]). However, the converse does not hold in general as the following counterexample shows.

**Example 4.4.** Define the kernel  $G : [0, T]^2 \rightarrow \mathbb{R}$  by  $G(t, s) = \mathbf{1}_{\{t > s\}} \varphi(t - s)$ , where

$$\varphi : (0, T] \rightarrow \mathbb{R}, \quad \varphi(u) := u^{-0.5}.$$

Then  $\int_0^T G(t, s)^2 dt = \infty$  for every  $s < T$  and therefore  $G \notin L^2([0, T]^2, \mathbb{R})$ . Now let  $f, g \in L^2([0, T], \mathbb{R})$  such that  $\|f\|_{L^2} = \|g\|_{L^2} = 1$ . Then applications of Hölder's inequality and Young's convolution inequality imply,

$$\begin{aligned} \int_0^T \int_0^T |g(t)G(t, s)f(s)| ds dt &= \int_0^T |g(t)| \int_0^t \varphi(t - s) |f(s)| ds dt \\ &= \int_0^T |g(t)| (\varphi \star |f|)(t) dt \\ &\leq \|g\|_{L^2} \|\varphi \star |f|\|_{L^2} \\ &\leq \|g\|_{L^2} \|\varphi\|_{L^1} \|f\|_{L^2}, \end{aligned}$$

and therefore that  $\|G\|_{L^2} \leq \|\varphi\|_{L^1} < \infty$ . Here,  $\varphi \star |f|$  denotes the convolution of  $\varphi$  and  $|f|$ . This example can be generalized to the case where  $G$  is  $\mathbb{R}^{N \times N}$ -valued.

The following auxiliary lemmas are essential ingredients for the proof of Proposition 4.1.

**Lemma 4.5.** For  $K, L \in \mathcal{G}$  any fixed  $t \in [0, T]$ , the integral operator  $\mathbf{K}_t + \mathbf{L}_t^*$  is nonnegative definite.

*Proof.* Let  $f \in L^2([0, T], \mathbb{R}^N)$ . From Definition 2.7 we have,

$$\begin{aligned}
& \langle f, (\mathbf{K}_t + \mathbf{L}_t^*)f \rangle_{L^2} \\
&= \int_0^T \int_0^T f(s)^\top \left( \mathbf{1}_{\{r \geq t\}} K(s, r) + \mathbf{1}_{\{s \geq t\}} L(r, s)^\top \right) f(r) dr ds \\
&= \int_0^T \int_0^T f(s)^\top \left( \mathbf{1}_{\{r \geq t\}} \mathbf{1}_{\{s \geq r\}} \mathbf{1}_{\{s \geq t\}} K(s, r) + \mathbf{1}_{\{s \geq t\}} \mathbf{1}_{\{r \geq s\}} \mathbf{1}_{\{r \geq t\}} L(r, s)^\top \right) f(r) dr ds \\
&= \int_0^T \int_0^T \mathbf{1}_{\{s \geq t\}} f(s)^\top \left( K(s, r) + L(r, s)^\top \right) \mathbf{1}_{\{r \geq t\}} f(r) dr ds \\
&= \langle f_t, (\mathbf{K} + \mathbf{L}^*)f_t \rangle_{L^2} \geq 0,
\end{aligned}$$

where  $f_t(s) := \mathbf{1}_{\{s \geq t\}} f(s)$  and we have used the fact that  $K$  and  $L$  are Volterra kernels in  $\mathcal{G}$  (see (2.9)).  $\square$

**Lemma 4.6.** *Let  $\bar{\Lambda} \in \mathbb{R}^{N \times N}$  be a symmetric positive definite matrix. Then, the linear operator  $\bar{\Lambda} \mathbf{I}$  on  $L^2([0, T], \mathbb{R}^N)$  is self-adjoint and satisfies*

$$\langle \bar{\Lambda} \mathbf{I} f, f \rangle_{L^2} = \langle \bar{\Lambda} f, f \rangle_{L^2} \geq \lambda_{\min} \langle f, f \rangle_{L^2}, \quad \text{for all } f \in L^2([0, T], \mathbb{R}^N)$$

where  $\lambda_{\min} > 0$  denotes the smallest eigenvalue of  $\bar{\Lambda}$ . In particular,  $\bar{\Lambda} \mathbf{I}$  is positive definite.

*Proof.* Since  $\bar{\Lambda}$  is symmetric positive definite, it can be decomposed as

$$\bar{\Lambda} = Q \Delta Q^\top,$$

where  $Q$  is an orthogonal matrix whose columns are orthonormal eigenvectors of  $\bar{\Lambda}$  and  $\Delta$  is a diagonal matrix whose entries  $(\lambda_i)_{i=1, \dots, N}$  are the positive eigenvalues of  $\bar{\Lambda}$ . Let  $x \in \mathbb{R}^N$  and define  $y = Q^\top x$ , which in particular implies that  $\|y\| = \|x\|$  as  $Q$  is orthogonal. We get,

$$x^\top \bar{\Lambda} x = x^\top Q \Delta Q^\top x = \|\Delta^{\frac{1}{2}} Q^\top x\|^2 = \sum_{i=1}^N \lambda_i y_i^2 \geq \sum_{i=1}^N \lambda_{\min} y_i^2 = \lambda_{\min} \|x\|^2. \quad (4.2)$$

Let  $f \in L^2([0, T], \mathbb{R}^N)$ . Then from (4.2) it follows that

$$\langle f, \bar{\Lambda} f \rangle_{L^2} = \int_0^T f(t)^\top \bar{\Lambda} f(t) dt \geq \int_0^T \lambda_{\min} \|f(t)\|^2 dt = \lambda_{\min} \langle f, f \rangle_{L^2}.$$

$\square$

**Notation.** Let  $E : V \rightarrow W$  be a linear operator between two normed real vector spaces  $(V, \|\cdot\|_V)$  and  $(W, \|\cdot\|_W)$ . Then the operator norm of  $E$  is denoted as,

$$\|E\|_{\text{op}} := \sup \{ \|E(x)\|_W : x \in V \text{ with } \|x\|_V \leq 1 \}. \quad (4.3)$$

**Lemma 4.7.** Let  $E$  be a bounded linear operator from a real Hilbert space  $V$  into itself. Suppose that there exists a constant  $c > 0$  such that  $\langle Ex, x \rangle \geq c\langle x, x \rangle$  for all  $x \in V$ . Then  $E$  is invertible and  $\|E^{-1}\|_{\text{op}} \leq c^{-1}$ .

*Proof.* See [27], Chapter 10, Problem 165. □

Now we are ready to prove Proposition 4.1.

*Proof of Proposition 4.1.* For every  $0 \leq t \leq T$  define the auxiliary process

$$m_t(s) = \mathbf{1}_{\{t \leq s\}} \mathbb{E}_t[u_s], \quad 0 \leq s \leq T.$$

Taking the conditional expectation  $\mathbb{E}_t[\cdot]$  on both sides of (4.1), then multiplying by  $\mathbf{1}_{\{t \leq s\}}$  and using the tower property we get

$$\begin{aligned} \bar{\Lambda} m_t(s) &= \mathbf{1}_{\{t \leq s\}} \mathbb{E}_t[f_s] - \mathbf{1}_{\{t \leq s\}} \int_0^t K(s, r) u_r dr \\ &\quad - \mathbf{1}_{\{t \leq s\}} \int_t^s K(s, r) \mathbb{E}_t[u_r] dr - \mathbf{1}_{\{t \leq s\}} \int_s^T L(r, s)^\top \mathbb{E}_t[u_r] dr \\ &= f_t^u(s) - \int_t^s K_t(s, r) m_t(r) dr - \int_s^T L_t(r, s)^\top m_t(r) dr \\ &= f_t^u(s) - ((\mathbf{K}_t + \mathbf{L}_t^*) m_t)(s), \quad \text{for all } 0 \leq s \leq T, \end{aligned} \quad (4.4)$$

where

$$f_t^u(s) := \mathbf{1}_{\{t \leq s\}} \mathbb{E}_t[f_s] - \mathbf{1}_{\{t \leq s\}} \int_0^t K(s, r) u_r dr. \quad (4.5)$$

The bounded linear operator  $\mathbf{K}_t + \mathbf{L}_t^*$  is nonnegative definite by Lemma 4.5. Thus, due to Lemma 4.6 and Lemma 4.7, the operator  $\mathbf{D}_t = \bar{\Lambda} \mathbf{I} + \mathbf{K}_t + \mathbf{L}_t^*$  is invertible with,

$$\|\mathbf{D}_t^{-1}\|_{\text{op}} \leq \lambda_{\min}^{-1}, \quad \text{for all } 0 \leq t \leq T, \quad (4.6)$$

where  $\|\cdot\|_{\text{op}}$  denotes the operator norm as introduced in (4.3) and  $\lambda_{\min} > 0$  denotes the smallest eigenvalue of  $\bar{\Lambda}$ . It follows from (4.4) that

$$m_t(s) = (\mathbf{D}_t^{-1} f_t^u)(s), \quad \text{for all } 0 \leq s \leq T. \quad (4.7)$$

By plugging in (4.7) into (4.1) we get

$$\bar{\Lambda} u_t = f_t - \int_0^t K(t, r) u_r dr - \int_t^T L(r, t)^\top (\mathbf{D}_t^{-1} f_t^u)(r) dr. \quad (4.8)$$

Using (4.5) we rewrite the third term on the right-hand side of (4.8) as follows:

$$\begin{aligned}
& \int_t^T L(r, t)^\top (\mathbf{D}_t^{-1} f_t^u)(r) dr \\
&= \int_t^T L(r, t)^\top (\mathbf{D}_t^{-1} \mathbb{1}_{\{t \leq \cdot\}} \mathbb{E}_t[f \cdot])(r) dr - \int_t^T L(r, t)^\top (\mathbf{D}_t^{-1} \mathbb{1}_{\{t \leq \cdot\}} \int_0^t K(\cdot, s) u_s ds)(r) dr \\
&= \int_t^T L(r, t)^\top (\mathbf{D}_t^{-1} \mathbb{1}_{\{t \leq \cdot\}} \mathbb{E}_t[f \cdot])(r) dr \\
&\quad - \int_0^t \int_t^T L(r, t)^\top (\mathbf{D}_t^{-1} \diamond \mathbb{1}_{\{t \leq \cdot\}} K(\cdot, s))(r) dr u_s ds,
\end{aligned} \tag{4.9}$$

where  $\mathbf{D}_t^{-1} \diamond \mathbb{1}_{\{t \leq \cdot\}} K(\cdot, s)$  denotes the column-wise application of the operator  $\mathbf{D}_t^{-1}$  to the function  $\mathbb{1}_{\{t \leq r\}} K(r, s)$  for fixed  $s \in [0, T]$  (see Definition 2.6). The above equality holds because of Fubini's theorem and the fact that the operator  $\mathbf{D}_t^{-1}$ , multiplication by  $\mathbb{1}_{\{t \leq r\}} K(r, s)$ , and integration are all linear operations. From (4.8) and (4.9) it follows that

$$u_t = a_t - \int_0^T B(t, s) u_s ds, \quad 0 \leq t \leq T, \tag{4.10}$$

where  $a(\cdot)$  and  $B(\cdot, \cdot)$  are given in the statement of Proposition 4.1.

We introduce the following auxiliary lemma, which will be proved at the end of this section.

**Lemma 4.8.** *Let  $a, B$  be defined as in Proposition 4.1 and let  $u$  be a solution of (4.10). Then the following hold:*

- (i)  $\mathbb{E}[\int_0^T \|a_t\|^2 dt] < \infty$ ,
- (ii)  $\int_0^T \int_0^T \|B(t, s)\|^2 ds dt < \infty$ ,
- (iii)  $\mathbb{E}[\int_0^T \|u_t\|^2 dt] < \infty$ .

Note that by Corollary 9.3.16 of [17] and Lemma 4.8(ii), the deterministic Volterra kernel  $B$  has a resolvent  $R^B : [0, T]^2 \rightarrow \mathbb{R}^{N \times N}$  of type  $L^2$  (recall Definition 4.3), which is the kernel corresponding to this operator,

$$\mathbf{R}^B = \mathbf{I} - (\mathbf{I} + \mathbf{B})^{-1}.$$

Therefore, by Theorem 9.3.6 of [17] the integral equation (4.10) admits a unique solution  $u^*(\omega)$  for any fixed  $\omega \in \Omega$  for which  $a(\omega) \in L^2([0, T], \mathbb{R}^N)$ , i.e. for  $\mathbb{P}$ -a.e.  $\omega$  by Lemma 4.8(i). This solution  $u^*$  is given by a variation of constants formula in terms of the resolvent  $R^B$ , namely

$$u_t^* = a_t - \int_0^T R^B(t, s) a_s ds, \quad 0 \leq t \leq T. \tag{4.11}$$

It follows that any solution  $u^* \in \mathcal{U}$  to the coupled system of stochastic Fredholm equations must be of the form

$$u_t^* = ((\mathbf{I} + \mathbf{B})^{-1}a)(t), \quad 0 \leq t \leq T, \quad \mathbb{P} - \text{a.s.}, \quad (4.12)$$

which yields the uniqueness of  $u^*$  up to modifications. For the existence, let  $u^*$  be defined as in (4.12). It follows that  $u^*$  satisfies (4.11) and thus (4.10). Therefore, it also satisfies (4.8) and consequently (4.1) due to (4.7), as desired. Thanks to Lemma 4.8(iii) it is ensured that  $u^* \in \mathcal{U}$ .  $\square$

*Proof of Lemma 4.8.* (i) Recall that

$$a_t = \bar{\Lambda}^{-1} \left( f_t - \int_t^T L(r, t)^\top (\mathbf{D}_t^{-1} \mathbf{1}_{\{t \leq \cdot\}} \mathbb{E}_t[f \cdot])(r) dr \right).$$

Since  $C_1 := \mathbb{E} \left[ \int_0^T \|f_t\|^2 dt \right] < \infty$  by the hypothesis of the lemma, it suffices to show that

$$\mathbb{E} \left[ \int_0^T \left\| \int_t^T L(r, t)^\top (\mathbf{D}_t^{-1} \mathbf{1}_{\{t \leq \cdot\}} \mathbb{E}_t[f \cdot])(r) dr \right\|^2 dt \right] < \infty. \quad (4.13)$$

To see this, note that from (4.6), from the conditional Jensen inequality and the tower property we get

$$\begin{aligned} \mathbb{E} \left[ \int_0^T \left\| (\mathbf{D}_t^{-1} \mathbf{1}_{\{t \leq \cdot\}} \mathbb{E}_t[f \cdot])(r) \right\|^2 dr \right] &\leq \lambda_{\min}^{-2} \mathbb{E} \left[ \int_0^T \left\| \mathbf{1}_{\{t \leq r\}} \mathbb{E}_t[f_r] \right\|^2 dr \right] \\ &\leq \lambda_{\min}^{-2} \int_0^T \mathbb{E} \left[ \mathbb{E}_t[\|f_r\|^2] \right] dr \\ &\leq \lambda_{\min}^{-2} C_1, \quad \text{for all } 0 \leq t \leq T. \end{aligned} \quad (4.14)$$

Moreover, recall that  $L \in \mathcal{G}$  by assumption and thus

$$C_2 := \int_0^T \int_0^T \|L(r, t)\|^2 dr dt < \infty,$$

see (2.9). The submultiplicativity of the Frobenius norm, Hölder's inequality and (4.14) yield

$$\begin{aligned} &\mathbb{E} \left[ \int_0^T \left\| \int_t^T L(r, t)^\top (\mathbf{D}_t^{-1} \mathbf{1}_{\{t \leq \cdot\}} \mathbb{E}_t[f \cdot])(r) dr \right\|^2 dt \right] \\ &\leq \mathbb{E} \left[ \int_0^T \left( \int_0^T \|L(r, t)\| \left\| (\mathbf{D}_t^{-1} \mathbf{1}_{\{t \leq \cdot\}} \mathbb{E}_t[f \cdot])(r) \right\| dr \right)^2 dt \right] \\ &= \int_0^T \int_0^T \|L(r, t)\|^2 dr \mathbb{E} \left[ \int_0^T \left\| (\mathbf{D}_t^{-1} \mathbf{1}_{\{t \leq \cdot\}} \mathbb{E}_t[f \cdot])(r) \right\|^2 dr \right] dt \\ &\leq C_2 C_1, \end{aligned}$$

which verifies (4.13).

(ii) Recall that

$$B(t, s) = -\bar{\Lambda}^{-1} \left( \mathbf{1}_{\{s \leq t\}} \int_t^T L(r, t)^\top (\mathbf{D}_t^{-1} \diamond \mathbf{1}_{\{t \leq \cdot\}} K(\cdot, s))(r) dr - K(t, s) \right). \quad (4.15)$$

Since  $K \in \mathcal{G}$ , it holds that  $C_3 := \int_0^T \int_0^T \|K(r, s)\|^2 ds dr < \infty$ . Moreover, the submultiplicativity of the Frobenius norm, Hölder's inequality and (4.6) imply that

$$\begin{aligned} & \int_0^T \int_0^T \left\| \mathbf{1}_{\{s \leq t\}} \int_t^T L(r, t)^\top (\mathbf{D}_t^{-1} \diamond \mathbf{1}_{\{t \leq \cdot\}} K(\cdot, s))(r) dr \right\|^2 ds dt \\ & \leq \int_0^T \int_0^T \left( \int_0^T \|L(r, t)\| \|(\mathbf{D}_t^{-1} \diamond \mathbf{1}_{\{t \leq \cdot\}} K(\cdot, s))(r)\| dr \right)^2 ds dt \\ & \leq \int_0^T \int_0^T \int_0^T \|L(r, t)\|^2 dr \int_0^T \|(\mathbf{D}_t^{-1} \diamond \mathbf{1}_{\{t \leq \cdot\}} K(\cdot, s))(r)\|^2 dr ds dt \\ & \leq \int_0^T \int_0^T \int_0^T \|L(r, t)\|^2 dr \lambda_{\min}^{-2} \int_0^T \|K(r, s)\|^2 dr ds dt \\ & \leq \lambda_{\min}^{-2} C_2 C_3. \end{aligned}$$

Together with (4.15) we get the result.

(iii) From (i) it holds that  $C_4 := \mathbb{E}[\int_0^T \|a_t\|^2 dt] < \infty$ . Furthermore, since  $R^B$  is of type  $L^2$ , we have that  $C_5 := \|R^B\|_{L^2}^2 < \infty$ . Now it follows from an  $L^p$ -inequality for kernels (see [17], Theorem 9.2.4) that

$$\begin{aligned} \mathbb{E} \left[ \int_0^T \left\| \int_0^T R^B(t, s) a_s ds \right\|^2 dt \right] &= \mathbb{E} \left[ \left\| \int_0^T R^B(t, s) a_s ds \right\|_{L^2}^2 \right] \\ &\leq \mathbb{E} \left[ \|R^B\|_{L^2}^2 \|a\|_{L^2}^2 \right] \\ &= C_5 C_4. \end{aligned}$$

Together with (2.4) and (4.11) it follows that  $u^* \in \mathcal{U}$ . □

## 5 Proofs of Theorem 2.8 and Corollary 2.9

In order to prove Theorem 2.8 and Corollary 2.9, we introduce the following essential lemmas. We first prove that the objective functional  $J(u)$  in (2.10) is strictly concave in  $u \in \mathcal{U}$ .

**Lemma 5.1.** *For any fixed  $G \in \mathcal{G}$  the map  $u \mapsto J(u)$  is strictly concave in  $u \in \mathcal{U}$ .*

*Proof.* We prove the result by showing that each of the components of  $J(u)$  is concave and at least one of them is strictly concave. We first notice that the term

$$-\int_0^T u_t^\top P_t dt + (X_T^u)^\top P_T,$$

is affine in  $u$ .

Recall that  $\Lambda$  is a positive definite matrix, hence the quadratic form  $x \mapsto x^\top \Lambda x$  is strictly convex in  $x \in \mathbb{R}^N$ , which implies that

$$-\int_0^T u_t^\top \Lambda u_t dt$$

is strictly concave in  $u$  due to the linearity and monotonicity of the integral. Thus by (2.10),  $J(u)$  is strictly concave if the term

$$-\int_0^T u_t^\top D_t^u dt - \frac{\gamma}{2} \int_0^T (X_t^u)^\top \Sigma X_t^u dt =: C_1(u) + C_2(u), \quad (5.1)$$

is concave in  $u$ . Without loss of generality, we assume that  $X_0 = 0$ , as terms in  $C_2(\cdot)$  involving  $X_0$  are affine in  $u$ . In order to show the concavity of (5.1), we adopt the approach from Gatheral et al. (see Proposition 2.9 in [16]). Since  $G \in \mathcal{G}$  and  $\Sigma$  is nonnegative definite, it follows from (2.6) and (2.7) that  $C_1(u), C_2(u) \leq 0$  for all  $u \in \mathcal{U}$ . Next, define the cross functionals

$$\tilde{C}_1(u, v) := -\int_0^T u_t^\top D_t^v dt \quad \text{and} \quad \tilde{C}_2(u, v) := -\frac{\gamma}{2} \int_0^T (X_t^u)^\top \Sigma X_t^v dt, \quad u, v \in \mathcal{U}. \quad (5.2)$$

From (5.1) and (5.2) it follows that

$$C_i(u - v) = C_i(u) + C_i(v) - \tilde{C}_i(u, v) - \tilde{C}_i(v, u), \quad i = 1, 2$$

and since  $C_i(u - v) \leq 0$ ,  $i = 1, 2$  we get that

$$C_i\left(\frac{1}{2}u + \frac{1}{2}v\right) = \frac{1}{4}C_i(u) + \frac{1}{4}C_i(v) + \frac{1}{4}\tilde{C}_i(u, v) + \frac{1}{4}\tilde{C}_i(v, u) \geq \frac{1}{2}C_i(u) + \frac{1}{2}C_i(v). \quad (5.3)$$

The concavity of  $C_i$  for  $i = 1, 2$ , follows from (5.3) and the fact that the map

$$\lambda \mapsto C_i(\lambda u + (1 - \lambda)v) = \lambda^2 C_i(u) + (1 - \lambda)^2 C_i(v) + \lambda(1 - \lambda)(\tilde{C}_i(u, v) + \tilde{C}_i(v, u))$$

is continuous in  $\lambda$ . We therefore obtain the concavity of the left-hand side of (5.1) and hence of  $J(\cdot)$ .  $\square$



From Lemma 5.1 it follows that the objective functional  $J$  in (2.10) admits a unique maximizer characterized by the critical point  $u \in \mathcal{U}$  at which its Gâteaux derivative

$$\langle J'(u), v \rangle := \lim_{\varepsilon \rightarrow 0} \frac{J(u + \varepsilon v) - J(u)}{\varepsilon}$$

vanishes for any  $v \in \mathcal{U}$  (see Propositions 1.2 and 2.1 in Chapter 2 of [11]). In the following lemma we calculate  $\langle J'(u), v \rangle$ .

**Lemma 5.2.** *For any  $u, v \in \mathcal{U}$ , the Gâteaux derivative  $\langle J'(u), v \rangle$  of  $J$  is given by*

$$\mathbb{E} \left[ \int_0^T v_t^\top \left( -P_t - ((\mathbf{G} + \mathbf{G}^*)u)(t) - \bar{\Lambda}u_t + P_T - \gamma \int_t^T \Sigma X_s^u ds \right) dt \right].$$

*Proof.* Let  $u, v \in \mathcal{U}$  and let  $\varepsilon > 0$ . Recall that the objective functional is given by

$$J(u) = \mathbb{E} \left[ \int_0^T -u_t^\top (P_t + D_t^u + \frac{1}{2} \Lambda u_t) dt + (X_T^u)^\top P_T - \frac{\gamma}{2} \int_0^T (X_t^u)^\top \Sigma X_t^u dt \right].$$

A direct computation using (2.3) and (2.6) yields

$$\begin{aligned} & J(u + \varepsilon v) - J(u) \\ &= \varepsilon \cdot \mathbb{E} \left[ \int_0^T \left( -v_t^\top (P_t + D_t^u + \frac{1}{2} \Lambda u_t) - u_t^\top (D_t^v + \frac{1}{2} \Lambda v_t) \right) dt \right. \\ & \quad \left. + \int_0^T v_t^\top P_T dt - \frac{\gamma}{2} \int_0^T \left( \int_0^t v_s^\top ds \right) \Sigma X_t^u - (X_t^u)^\top \Sigma \int_0^t v_s ds dt \right] + O(\varepsilon^2) \\ &= \varepsilon \cdot \mathbb{E} \left[ \int_0^T v_t^\top \left( -P_t - ((\mathbf{G} + \mathbf{G}^*)u)(t) - \frac{1}{2} (\Lambda + \Lambda^\top) u_t + P_T \right. \right. \\ & \quad \left. \left. - \gamma \int_t^T \Sigma X_s^u ds \right) dt \right] + O(\varepsilon^2), \end{aligned}$$

where we used the symmetry of  $\Sigma$  and applied Fubini's theorem in the last equality. Recalling (2.11), dividing by  $\varepsilon$  and taking the limit as  $\varepsilon \downarrow 0$  yields the result.  $\square$

Next, we derive from Lemma 5.2 a system of stochastic Fredholm equations which is satisfied by the unique maximizer of the objective functional  $J(u)$  in (2.10).

**Lemma 5.3.** *For fixed  $G \in \mathcal{G}$ , the unique maximizer of the objective functional  $J(u)$  in (2.10) satisfies the coupled system of stochastic Fredholm equations of the second kind given by*

$$\bar{\Lambda}u_t = g_t - \int_0^t K(t, s)u_s ds - \int_t^T K(s, t)^\top \mathbb{E}_t[u_s] ds, \quad 0 \leq t \leq T, \quad (5.9)$$

where  $\bar{\Lambda}$ ,  $K$  and  $g$  are defined in (2.11), (2.12) and (2.14), respectively.

*Proof.* The unique maximizer of  $J(u)$  is characterized by the critical point  $u \in \mathcal{U}$  at which the Gâteaux derivative  $\langle J'(u), v \rangle$  is equal to 0 for any  $v \in \mathcal{U}$ . It follows from Lemma 5.2 that this is equivalent to

$$\mathbb{E} \left[ \int_0^T v_t^\top \left( -P_t - ((\mathbf{G} + \mathbf{G}^*)u)(t) - \bar{\Lambda}u_t + P_T - \gamma \int_t^T \Sigma X_s^u ds \right) dt \right] = 0, \quad \forall v \in \mathcal{U}. \quad (5.10)$$

Next, by an application of Fubini's theorem and recalling that  $F$  was defined in (2.12),

$$\begin{aligned} \gamma \int_t^T \Sigma X_s^u ds &= \gamma \Sigma X_0(T-t) + \gamma \Sigma \int_t^T \int_0^s u_r dr ds \\ &= \gamma \Sigma X_0(T-t) + \gamma \Sigma \int_0^T (T - (r \vee t)) u_r dr \\ &= \gamma \Sigma X_0(T-t) + ((\mathbf{F} + \mathbf{F}^*)u)(t), \end{aligned} \quad (5.11)$$

where  $r \vee t := \max\{r, t\}$ . Plugging (5.11) into (5.10), conditioning on  $\mathcal{F}_t$  and using the tower property of conditional expectation we get the following first order condition,

$$\begin{aligned} & -P_t - ((\mathbf{G} + \mathbf{G}^*)(\mathbb{E}_t u))(t) - \bar{\Lambda}u_t + \mathbb{E}_t[P_T] - \gamma \Sigma X_0(T-t) - ((\mathbf{F} + \mathbf{F}^*)(\mathbb{E}_t u))(t) \\ &= 0, \quad d\mathbb{P} \otimes dt\text{-a.e. on } \Omega \times [0, T]. \end{aligned} \quad (5.12)$$

Recalling the definition of  $K$  in (2.12) and  $g$  in (2.14), equation (5.12) simplifies to

$$\bar{\Lambda}u_t = g_t - ((\mathbf{K} + \mathbf{K}^*)(\mathbb{E}_t u))(t), \quad d\mathbb{P} \otimes dt\text{-a.e. on } \Omega \times [0, T],$$

which is equivalent to (5.9), since  $K$  is a Volterra kernel.  $\square$

In order to prove Theorem 2.8 and Corollary 2.9, the following lemma is needed, which in combination with Remark 2.5 shows that the Volterra kernel  $F$  defined in (2.12) is in  $\mathcal{G}$ .

**Lemma 5.4.** *The following inequality holds,*

$$\int_0^T \int_0^T f(t)^\top \gamma \Sigma (T - (s \vee t)) f(s) ds dt \geq 0, \quad \text{for all } f \in L^2([0, T], \mathbb{R}^N).$$

*Proof.* First, observe that for all  $(s, t) \in [0, T]^2$  we have

$$T - (s \vee t) = (T - s) \wedge (T - t) = \int_0^T \mathbf{1}_{\{r \leq T-s\}} \mathbf{1}_{\{r \leq T-t\}} dr = \int_0^T \mathbf{1}_{\{s \leq T-r\}} \mathbf{1}_{\{t \leq T-r\}} dr.$$

Thus for any  $f \in L^2([0, T], \mathbb{R}^N)$  we get that

$$\begin{aligned} & \int_0^T \int_0^T f(t)^\top \gamma \Sigma (T - (s \vee t)) f(s) ds dt \\ &= \gamma \int_0^T \int_0^{T-r} \int_0^{T-r} f(t)^\top \Sigma f(s) dt ds dr \\ &= \gamma \int_0^T \left( \int_0^{T-r} f(t) dt \right)^\top \Sigma \left( \int_0^{T-r} f(t) dt \right) dr \\ &\geq 0, \end{aligned}$$

since  $\Sigma$  is nonnegative definite and  $\gamma \geq 0$ .  $\square$

Now we are ready to prove Theorem 2.8 and Corollary 2.9.

*Proof of Theorem 2.8.* Let  $G \in \mathcal{G}$ . Recall that  $F$ ,  $K$  and  $g$  are defined in (2.11), (2.12) and (2.14), respectively. It follows from Lemma 5.4 and Remark 2.5 that  $F \in \mathcal{G}$  and thus  $K \in \mathcal{G}$  as well. Moreover,  $g$  is progressively measurable and satisfies

$$\int_0^T \mathbb{E}[\|g_t\|^2] dt < \infty,$$

because  $P$  is progressively measurable and satisfies  $\int_0^T \mathbb{E}[\|P_t\|^2] dt < \infty$  by assumption.

Now by Lemmas 5.1 and 5.3, the objective functional  $J(u)$  in (2.10) admits a unique maximizer in  $\mathcal{U}$ , and this maximizer satisfies (5.9). An application of Proposition 4.1 with  $f = g$  and  $L = K$  yields that (5.9) admits a unique progressively measurable solution  $u^* \in \mathcal{U}$  given by

$$u_t^* = ((\mathbf{I} + \mathbf{B})^{-1}a)(t), \quad 0 \leq t \leq T, \quad (5.17)$$

with  $a$  and  $\mathbf{B}$  defined as in Theorem 2.8. It follows that  $u^*$  in (5.17) must be the unique maximizer of  $J(u)$ , which finishes the proof.  $\square$

*Proof of Corollary 2.9.* Let  $G \in \mathcal{G}$  and assume that  $A$  in (2.1) is deterministic. Then,  $g$  in (2.14) is a deterministic function in  $L^2([0, T], \mathbb{R}^N)$  given by

$$g_t = A_T - A_t - \gamma(T - t)\Sigma X_0, \quad 0 \leq t \leq T.$$

It follows that the unique maximizer  $u^*$  of the objective functional  $J(u)$  from Theorem 2.8, which satisfies (5.9), is deterministic and satisfies the simplified equation

$$\bar{\Lambda}u_t = g_t - ((\mathbf{K} + \mathbf{K}^*)u)(t), \quad 0 \leq t \leq T. \quad (5.18)$$

Now (5.18) admits the deterministic solution

$$u_t^* = (\mathbf{D}^{-1}g)(t), \quad 0 \leq t \leq T, \quad (5.19)$$

where the bounded linear operator

$$\mathbf{D} = (\mathbf{K} + \mathbf{K}^* + \bar{\mathbf{I}}) = (\mathbf{G} + \mathbf{G}^* + \mathbf{F} + \mathbf{F}^* + \bar{\mathbf{I}})$$

defined in (2.12) and (2.13) is invertible because of the following. Since  $G \in \mathcal{G}$  by assumption,  $F \in \mathcal{G}$  due to Lemma 5.4 and Remark 2.5, and there exists a real number  $c > 0$  such that  $\bar{\mathbf{I}}$  satisfies

$$\langle \bar{\mathbf{I}}f, f \rangle_{L^2} \geq c \langle f, f \rangle_{L^2}, \quad \text{for all } f \in L^2([0, T], \mathbb{R}^N)$$

due to Lemma 4.6. It follows that the operator  $\mathbf{D}$  is positive definite with

$$\langle \mathbf{D}f, f \rangle_{L^2} \geq c \langle f, f \rangle_{L^2}, \quad \text{for all } f \in L^2([0, T], \mathbb{R}^N).$$

Therefore, Lemma 4.7 implies that  $\mathbf{D}$  is invertible.

It follows that the deterministic  $u^*$  in (5.19) solves (5.9) in Lemma 5.3. This solution is unique due to Proposition 4.1. Thus,  $u^*$  in (5.19) is the unique maximizer of  $J(u)$ .  $\square$

## 6 Proofs of Theorem 2.14 and Corollary 2.17

This section is dedicated to the proofs of of Theorem 2.14 and Corollary 2.17, which concern convolution kernels. However, one of the main ingredients for these proofs is Proposition 6.3, which gives a new sufficient condition for the nonnegative definite property for a more general class of Volterra kernels. Since this result could be of independent interest for the theory of Volterra equations we assume in the first part of this section that the kernel  $G$  is a Volterra kernel.

### 6.1 Volterra kernels

**Definition 6.1.** *Let  $G \in L^2([0, T]^2, \mathbb{R}^{N \times N})$  be a Volterra kernel. Then the associated mirrored kernel  $\tilde{G} \in L^2([0, T]^2, \mathbb{R}^{N \times N})$  is defined as*

$$\tilde{G}(t, s) := \begin{cases} G(t, s) & \text{for } t > s, \\ \frac{1}{2}(G(t, t) + G(t, t)^\top) & \text{for } t = s, \\ G(s, t)^\top & \text{for } t < s. \end{cases}$$

**Remark 6.2.** *It holds for any Volterra kernel  $G \in L^2([0, T]^2, \mathbb{R}^{N \times N})$  that*

$$\tilde{G}(t, s) = G(t, s) + G(s, t)^\top \quad dt \otimes ds\text{-a.e. on } [0, T]^2,$$

*and therefore that  $\langle f, \tilde{\mathbf{G}}f \rangle_{L^2} = \langle f, (\mathbf{G} + \mathbf{G}^*)f \rangle_{L^2}$  for every  $f \in L^2([0, T], \mathbb{R}^N)$ . This observation will be crucial for the proof of Proposition 6.3.*

The following proposition provides a sufficient condition (6.1) for the nonnegative definite property of Volterra kernels. Note that (6.1) below is similar to Definition 2 in [3], which was used in order to define nonnegative definite convolution kernels.

**Proposition 6.3.** *Let  $G \in L^2([0, T]^2, \mathbb{R}^{N \times N})$  be a matrix-valued Volterra kernel such that its mirrored kernel  $\tilde{G}$  is continuous and satisfies*

$$\sum_{k,l=1}^n x_k^\top \tilde{G}(t_k, t_l) x_l \geq 0, \quad (6.1)$$

for any  $n \in \mathbb{N}$ ,  $t_1, \dots, t_n \in [0, T]$  and  $x_1, \dots, x_n \in \mathbb{R}^N$ . Then for any function  $f \in L^2([0, T], \mathbb{R}^N)$  the following holds,

$$\langle f, \mathbf{G}f \rangle_{L^2} = \frac{1}{2} \langle f, (\mathbf{G} + \mathbf{G}^*)f \rangle_{L^2} = \frac{1}{2} \langle f, \tilde{\mathbf{G}}f \rangle_{L^2} \geq 0.$$

In order to prove Proposition 6.3 we will need the following auxiliary lemma.

**Lemma 6.4.** *Let  $G : [0, T] \times [0, T] \rightarrow \mathbb{R}^{N \times N}$  be a map satisfying  $G(t, s) = G(s, t)^\top$  for all  $t, s \in [0, T]$ . Then it holds that*

$$\sum_{k,l=1}^n x_k^\top G(t_k, t_l) x_l \geq 0 \text{ for any } n \in \mathbb{N}, t_1, \dots, t_n \in [0, T], x_1, \dots, x_n \in \mathbb{R}^N \quad (6.2)$$

if and only if

$$\sum_{k,l=1}^n \bar{z}_k^\top G(t_k, t_l) z_l \geq 0 \text{ for any } n \in \mathbb{N}, t_1, \dots, t_n \in [0, T], z_1, \dots, z_n \in \mathbb{C}^N. \quad (6.3)$$

*Proof.* Let  $n \in \mathbb{N}$ ,  $t_1, \dots, t_n \in [0, T]$ ,  $z_1, \dots, z_n \in \mathbb{C}^N$ , and assume (6.2). Using (6.2) we get

$$\begin{aligned} & \sum_{k,l=1}^n \bar{z}_k^\top G(t_k, t_l) z_l \\ &= \sum_{k,l=1}^n (\operatorname{Re}(z_k) - i \operatorname{Im}(z_k))^\top G(t_k, t_l) (\operatorname{Re}(z_l) + i \operatorname{Im}(z_l)) \\ &\geq -i \sum_{k,l=1}^n \operatorname{Im}(z_k)^\top G(t_k, t_l) \operatorname{Re}(z_l) + i \sum_{k,l=1}^n \operatorname{Re}(z_k)^\top G(t_k, t_l) \operatorname{Im}(z_l) \\ &= -i \sum_{k,l=1}^n \operatorname{Im}(z_k)^\top G(t_k, t_l) \operatorname{Re}(z_l) + i \sum_{k,l=1}^n \operatorname{Im}(z_l)^\top G(t_k, t_l)^\top \operatorname{Re}(z_k) \\ &= 0, \end{aligned}$$

since  $G(t_k, t_l)^\top = G(t_l, t_k)$  by assumption this proves (6.3). The reverse direction is trivial.  $\square$

*Proof of Proposition 6.3.* The main idea of the proof is to apply a generalized version of Mercer's theorem for matrix-valued kernels (see [10]). In order to apply this result, first notice that  $[0, T]$  is a separable metric space with respect to the Euclidean distance and that the Lebesgue measure defined on  $\mathcal{B}([0, T])$  is finite. Next, by assumption, the mirrored kernel  $\tilde{G} : [0, T]^2 \rightarrow \mathbb{R}^{N \times N}$  associated with  $G$  is continuous and satisfies condition (6.1). Therefore by Lemma 6.4 it also satisfies (6.3). Moreover, since  $[0, T]^2$  is compact and  $\tilde{G}$  is continuous, it holds that

$$\int_0^T \text{Tr} \tilde{G}(t, t) dt \leq TN \sup_{1 \leq i \leq N} \sup_{0 \leq t \leq T} \tilde{G}_{ii}(t, t) < \infty, \quad (6.9)$$

where  $\text{Tr}$  denotes the trace of a matrix in  $\mathbb{R}^{N \times N}$ . Note that  $G$  must be nonnegative on the diagonal, this can be seen for example from the expansion (6.10) used later. Next, define the linear integral operator

$$\tilde{\mathbf{G}}_{\mathbb{C}} : L^2([0, T], \mathbb{C}^N) \rightarrow L^2([0, T], \mathbb{C}^N), \quad (\tilde{\mathbf{G}}_{\mathbb{C}} f)(t) = \int_0^T \tilde{G}(t, s) f(s) ds,$$

which is well-defined and bounded (see Theorem 9.2.4 in [17]), and denote by  $\ker(\tilde{\mathbf{G}}_{\mathbb{C}})$  its kernel. Since (6.3) and (6.9) hold, we can apply a generalized version of Mercer's theorem for matrix-valued kernels (see Theorem 4.1 in [10]), which, in combination with the auxiliary Theorem A.1 in [10], implies that there exists a countable orthonormal basis  $\{\varphi_m\}_{m \in I}$  of  $\ker(\tilde{\mathbf{G}}_{\mathbb{C}})^\perp \subset L^2([0, T], \mathbb{C}^N)$  of continuous eigenfunctions of  $\tilde{\mathbf{G}}_{\mathbb{C}}$  with a corresponding family  $\{\sigma_m\}_{m \in I} \subset (0, \infty)$  of positive eigenvalues such that

$$\tilde{G}_{ij}(t, s) = \sum_{m \in I} \sigma_m \overline{\varphi_m^i(t)} \varphi_m^j(s), \quad \text{for all } (s, t) \in [0, T]^2, \quad i, j \in \{1, \dots, N\}, \quad (6.10)$$

where the series converges uniformly on  $[0, T]^2$ . Let  $f \in L^2([0, T], \mathbb{R}^N)$ . From (6.10) it follows that

$$\begin{aligned} \langle f, \tilde{\mathbf{G}} f \rangle_{L^2} &= \int_0^T \int_0^T \sum_{i,j=1}^N f^i(t) \left( \sum_{m \in I} \sigma_m \overline{\varphi_m^i(t)} \varphi_m^j(s) \right) f^j(s) ds dt \\ &= \int_0^T \int_0^T \sum_{m \in I} \sigma_m \sum_{i,j=1}^N f^i(t) \overline{\varphi_m^i(t)} \varphi_m^j(s) f^j(s) ds dt \\ &= \int_0^T \int_0^T \int_I \sigma_m \sum_{i,j=1}^N f^i(t) \overline{\varphi_m^i(t)} \varphi_m^j(s) f^j(s) d\mu(m) ds dt, \end{aligned} \quad (6.11)$$

where  $\mu$  is the counting measure on the countable index set of the eigenfunctions  $I$ . In particular,  $\mu$  is  $\sigma$ -finite.

Next, we would like to switch the order of integration on the right-hand side of (6.11) using Fubini's theorem. In order to do that, we show that the following integral

is finite, by using Young's inequality, (6.9) and (6.10),

$$\begin{aligned}
& \int_0^T \int_0^T \int_I \left| \sigma_m \sum_{i,j=1}^N f^i(t) \overline{\varphi_m^i(t)} \varphi_m^j(s) f^j(s) \right| d\mu(m) ds dt \\
& \leq \int_0^T \int_0^T \sum_{i,j=1}^N |f^i(t)| \int_I \sigma_m (|\varphi_m^i(t)|^2 + |\varphi_m^j(s)|^2) d\mu(m) |f^j(s)| ds dt \\
& = \int_0^T \int_0^T \sum_{i,j=1}^N |f^i(t)| (\tilde{G}_{ii}(t, t) + \tilde{G}_{jj}(s, s)) |f^j(s)| ds dt \\
& \leq 2 \left( \sup_{1 \leq i \leq N} \sup_{0 \leq t \leq T} \tilde{G}_{ii}(t, t) \right) \int_0^T \int_0^T \sum_{i,j=1}^N |f^i(t)| |f^j(s)| ds dt \\
& < \infty.
\end{aligned}$$

Let  $\mathbf{1}_N$  denote the  $N \times N$  matrix of ones in all entries, which is nonnegative definite as its eigenvalues are  $N$  with multiplicity 1 and 0 with multiplicity  $N - 1$ . Moreover, for every  $m \in I$  we define

$$g_m : [0, T] \rightarrow \mathbb{C}^N, \quad g_m(t) := \begin{pmatrix} \varphi_m^1(t) f^1(t) \\ \vdots \\ \varphi_m^N(t) f^N(t) \end{pmatrix},$$

which is the Hadamard product of  $\varphi_m$  and  $f$ . Then (6.11) and Fubini's theorem imply

$$\begin{aligned}
\langle f, \tilde{\mathbf{G}} f \rangle_{L^2} &= \int_I \int_0^T \int_0^T \sigma_m \sum_{i,j=1}^N f^i(t) \overline{\varphi_m^i(t)} \varphi_m^j(s) f^j(s) ds dt d\mu(m) \\
&= \int_I \int_0^T \int_0^T \sigma_m \overline{g_m(t)}^\top \mathbf{1}_N g_m(s) ds dt d\mu(m) \\
&= \sum_{m \in I} \sigma_m \left( \int_0^T g_m(t) dt \right)^\top \mathbf{1}_N \left( \int_0^T g_m(s) ds \right) \geq 0.
\end{aligned}$$

This finishes the proof.  $\square$

**Remark 6.5.** In the proof of Proposition 6.3, we use Fubini's theorem and the counting measure  $\mu$  to justify interchanging integration and summation, since the uniform convergence of

$$\tilde{G}_{ij}(t, s) = \sum_{m \in I} \sigma_m \overline{\varphi_m^i(t)} \varphi_m^j(s),$$

on  $[0, T]^2$  does not imply the uniform convergence of

$$\sum_{m \in I} \sigma_m f^i(t) \overline{\varphi_m^i(t)} \varphi_m^j(s) f^j(s)$$

on  $[0, T]^2$  for any  $f \in L^2([0, T], \mathbb{R}^N)$  as  $f$  may be unbounded.

## 6.2 Convolution kernels

For the remainder of this section we restrict our discussion to convolution kernels of the form  $G(t, s) := \mathbb{1}_{\{t \geq s\}} H(t - s)$  as in Theorem 2.14.

*Proof of Theorem 2.14.* Without loss of generality, we can assume that

$$H(T) = \lim_{t \uparrow T} H(t) \in \mathbb{R}^{N \times N},$$

where the limit always exists, since  $H$  is continuous, nonincreasing, nonnegative and symmetric on  $(0, T)$  by assumption.

**Case 1:** We assume that  $H$  be bounded. Then we can also assume without loss of generality that

$$H(0) = \lim_{t \downarrow 0} H(t) \in \mathbb{R}^{N \times N},$$

where the limit exists by a similar argument. Now we continuously extend  $H$  to  $[0, \infty)$  by defining

$$\bar{H}(t) := \begin{cases} H(t) & \text{for } t < T, \\ H(T) & \text{for } t \geq T. \end{cases} \quad (6.12)$$

Then  $\bar{H}$  is continuous, symmetric, nonnegative, nonincreasing and convex on  $[0, \infty)$ . Thus, it follows from Theorem 2 in [3] that the mirrored kernel associated with the Volterra kernel  $\bar{G}(t, s) := \mathbb{1}_{\{t \geq s\}} \bar{H}(t - s)$  on  $[0, \infty)^2$  satisfies (6.3). Therefore, the continuous mirrored kernel  $\bar{G}$  associated with  $G$  satisfies (6.3) as well. This allows us to apply Proposition 6.3 to conclude that for every  $f \in L^2([0, T], \mathbb{R}^N)$ ,

$$\int_0^T \int_0^t f(t)^\top H(t - s) f(s) ds dt = \langle f, \mathbf{G}f \rangle_{L^2} = \frac{1}{2} \langle f, \tilde{\mathbf{G}}f \rangle_{L^2} \geq 0.$$

**Case 2:** We consider the case where  $H$  is unbounded. Then it follows from the assumptions of the theorem that  $H$  must have a singularity at  $t = 0$ , i.e. the limit of  $H(t)$  as  $t$  decreases to 0 does not exist in  $\mathbb{R}^{N \times N}$ . Let  $\bar{H}$  denote the extension of  $H$  to  $[0, \infty)$  as in (6.12). Then it follows that  $\bar{H}$  is nonincreasing, continuous, convex, nonnegative and symmetric on  $(0, \infty)$ . Define the sequence of convolution kernels  $(H_m)_{m \geq 1}$  by

$$H_m : [0, T] \rightarrow \mathbb{R}^{N \times N}, \quad H_m(t) := \bar{H}(t + m^{-1}), \quad \text{for } m \geq 1, \quad (6.13)$$

and notice that it converges pointwise to  $H$  on  $(0, T]$ , as  $\bar{H}$  is continuous on  $(0, \infty)$ . Furthermore, for every  $m \geq 1$ ,  $H_m$  is nonincreasing, convex, nonnegative, symmetric, continuous and thus bounded on  $[0, T]$ , so that the proof of Case 1 applies to them.

In order to proceed we notice that if  $C = (C_{ij})_{i,j \in \{1, \dots, N\}}$  is symmetric nonnegative definite matrix, then all of its diagonal entries  $C_{ii}$  are nonnegative and we have

$$|C_{ij}| \leq \frac{1}{2} (C_{ii} + C_{jj}) \leq \max_{1 \leq i \leq N} C_{ii},$$



where we have used the fact that  $x^\top Cx \geq 0$  for  $x = e_i \pm e_j$  and the symmetry of  $C$ . This implies the second inequality of the following,

$$\left(\max_{1 \leq i \leq N} C_{ii}\right)^2 \leq \|C\|^2 \leq N^2 \left(\max_{1 \leq i \leq N} C_{ii}\right)^2, \quad (6.14)$$

where the first inequality follows from the definition of the Frobenius norm.

Let  $f \in L^2([0, T], \mathbb{R}^N)$ . Using, in order of appearance below, the submultiplicativity of the Frobenius norm, Definition (6.13) and Young's inequality, the right-hand side of (6.14), the facts that the diagonal set  $\Delta := \{(t, s) \in [0, T]^2 : t = s\}$  has  $\mathbb{R}^2$ -Lebesgue measure zero and that the diagonal entries of  $\bar{H}(t)$  are nonincreasing on  $(0, \infty)$ , it holds for all  $m \geq 1$ ,

$$\begin{aligned} & \int_0^T \int_0^T |f(t)^\top \mathbb{1}_{\{t \geq s\}} H_m(t-s) f(s)| ds dt \\ & \leq \int_0^T \int_0^T \mathbb{1}_{\{t \geq s\}} \|f(t)\| \|H_m(t-s)\| \|f(s)\| ds dt \\ & \leq \int_0^T \int_0^T \mathbb{1}_{\{t \geq s\}} \left( \|\bar{H}(t-s+m^{-1})\|^2 + \|f(t)\|^2 \|f(s)\|^2 \right) ds dt \\ & \leq N^2 \int_0^T \int_0^T \mathbb{1}_{\{t \geq s\}} \left( \left( \max_{1 \leq i \leq N} \bar{H}_{ii}(t-s+m^{-1}) \right)^2 + \|f(t)\|^2 \|f(s)\|^2 \right) ds dt \quad (6.15) \\ & \leq N^2 \int_0^T \int_0^T \mathbb{1}_{\{t > s\}} \left( \left( \max_{1 \leq i \leq N} \bar{H}_{ii}(t-s) \right)^2 + \|f(t)\|^2 \|f(s)\|^2 \right) ds dt \\ & \leq N^2 \int_0^T \int_0^T \mathbb{1}_{\{t > s\}} \left( \|\bar{H}(t-s)\|^2 + \|f(t)\|^2 \|f(s)\|^2 \right) ds dt \\ & \leq N^2 \int_0^T \int_0^T \|G(t, s)\|^2 ds dt + N^2 \|f\|_{L^2}^4 \\ & < \infty, \end{aligned}$$

where we have used the left-hand side of (6.14), the fact that  $G(t, s) = \mathbb{1}_{\{t \geq s\}} H(t-s) = \mathbb{1}_{\{t \geq s\}} \bar{H}(t-s)$  for  $(s, t) \in [0, T]^2$ , and finally that  $G$  is in  $L^2([0, T]^2, \mathbb{R}^{N \times N})$  in the last three inequalities.

Now the sequence of functions

$$\left( f(t)^\top \mathbb{1}_{\{t \geq s\}} H_m(t-s) f(s) \right)_{m \geq 1}$$

defined on  $[0, T]^2$  converges pointwise on  $[0, T]^2 \setminus \Delta$  and thus almost everywhere to

$$f(t)^\top \mathbb{1}_{\{t \geq s\}} H(t-s) f(s).$$

Therefore, it follows from (6.15) that dominated convergence can be applied to get,

$$\int_0^T \int_0^t f(t)^\top H(t-s) f(s) ds dt = \lim_{m \rightarrow \infty} \int_0^T \int_0^t f(t)^\top H_m(t-s) f(s) ds dt \geq 0,$$

which completes the proof.  $\square$

*Proof of Corollary 2.17.* First, note that the function given by

$$[0, T] \rightarrow \mathbb{R}, \quad t \mapsto \int_0^t \phi(t-s)^2 ds$$

is the convolution of the integrable function  $\phi^2$  and the constant function  $\mathbb{1}_{[0, T]}$ . It follows that the convolution is integrable again (see Theorem 2.2.2 (i) in [17]), i.e.

$$\int_0^T \int_0^t \phi(t-s)^2 ds dt < \infty.$$

Therefore the Volterra kernel  $\overline{G}(t, s) = \mathbb{1}_{\{t \geq s\}} H(t-s)$  is in  $L^2([0, T]^2, \mathbb{R}^{N \times N})$  for  $H$  as in (2.17). Next, for any  $x \in \mathbb{R}^N$  we have that  $x^\top C x \geq 0$ . This implies that the function

$$t \mapsto x^\top H(t)x = x^\top C x \phi(t)$$

is nonincreasing and convex on  $(0, T)$  because  $\phi$  is nonincreasing and convex on  $(0, T)$  by assumption. Finally, the matrix

$$H(t) = C \phi(t)$$

is symmetric nonnegative definite for every  $t \in (0, T)$ , since  $\phi$  is nonnegative and  $C$  is symmetric nonnegative definite. Thus, the conditions of Theorem 2.14 are satisfied.  $\square$

## 7 Proof of Lemma 2.3

*Proof of Lemma 2.3.* Note that all propagator matrices from Example 2.2 except for the singular power-law kernel in Example 2.2(ii) and the non-convolution kernel in Example 2.2(v) are of the general form

$$G(t, s) = Q^\top \mathbb{1}_{\{t \geq s\}} \text{diag} (g_1(t-s), \dots, g_N(t-s)) Q$$

for an invertible matrix  $Q \in \mathbb{R}^{N \times N}$  and nonnegative, nonincreasing, convex functions  $g_1, \dots, g_N : [0, T] \rightarrow \mathbb{R}$ . For the matrix exponential kernel, this follows from Example 2 in [3]. For the other kernels, an eigendecomposition of the corresponding symmetric nonnegative definite matrix  $C$  yields the desired representation. Next, in order to see that  $G \in \mathcal{G}$ , we proceed by verifying the conditions of Theorem 2.14 for the kernel  $H(t) := Q^\top \text{diag} (g_1(t), \dots, g_N(t)) Q$ . Let  $x \in \mathbb{R}^N$  and  $y := Qx$ . Then it follows that

$$x^\top H(t)x = y^\top \text{diag} (g_1(t), \dots, g_N(t)) y = \sum_{i=1}^N y_i^2 g_i(t)$$

is nonnegative, nonincreasing and convex on  $[0, T]$ . Moreover  $H(t)$  is symmetric for any  $t \in [0, T]$ . Therefore, Theorem 2.14 applies, since  $G$  is bounded and thus in  $L^2([0, T]^2, \mathbb{R}^{N \times N})$ .

The singular power-law kernel in Example 2.2(ii) given by

$$G(t, s) = \mathbf{1}_{\{t>s\}}(t - s)^{-\beta} C$$

for  $\beta \in (0, \frac{1}{2})$  and nonnegative definite  $C \in \mathbb{R}^{N \times N}$  is in  $\mathcal{G}$  due to Corollary 2.17.

For the non-convolution kernel in Example 2.2(v) we follow similar lines as in the proof of Lemma 5.4, using the fact that  $H : [0, T] \rightarrow \mathbb{R}$  is nonnegative definite (see Example 2.7 in [16]) to get

$$\begin{aligned} & \int_0^T \int_0^T f(t) \alpha(T - t) \mathbf{1}_{\{t>s\}} H(t - s) f(s) ds dt \\ &= \alpha \int_0^T \int_0^{T-r} \int_0^{T-r} f(t) \mathbf{1}_{\{t>s\}} H(t - s) f(s) ds dt dr \\ &\geq 0, \quad \text{for all } f \in L^2([0, T], \mathbb{R}). \end{aligned}$$

Hence  $G$  is a product of a nonnegative, nonincreasing, convex function taking values in  $\mathbb{R}$  and a nonnegative definite matrix  $C \in \mathbb{R}^{N \times N}$ , so the result follows from the argument used for examples (i)–(iv) above.  $\square$

## References

- [1] E. Abi Jaber and E. Neuman. Optimal liquidation with signals: the general propagator case. *arXiv preprint arXiv:2211.00447*, 2022.
- [2] E. Abi Jaber, E. Neuman, and M. Voß. Equilibrium in functional stochastic games with mean-field interaction. *arXiv preprint arXiv:2306.05433*, 2023.
- [3] A. Alfonsi, F. Klöck, and A. Schied. Multivariate transient price impact and matrix-valued positive definite functions. *Mathematics of Operations Research*, 41(3):914–934, 2016. ISSN 0364765X, 15265471. URL <http://www.jstor.org/stable/24736389>.
- [4] R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40, 2001.
- [5] M. Benzaquen, I. Mastromatteo, Z. Eisler, and J.-P. Bouchaud. Dissecting cross-impact on stock markets: An empirical analysis. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(2):023406, 2017.

- [6] J.-P. Bouchaud, J. Bonart, J. Donier, and M. Gould. *Trades, Quotes and Prices: Financial Markets Under the Microscope*. Cambridge University Press, 2018. doi: 10.1017/9781316659335.
- [7] D. Brigo, F. Graceffa, and E. Neuman. Price impact on term structure. *Quantitative Finance*, 22(1):171–195, 2022.
- [8] F. Capponi and R. Cont. Multi-asset market impact and order flow commonality. *Available at SSRN 3706390*, 2020.
- [9] R. Cont, M. Cucuringu, and C. Zhang. Cross-impact of order flow imbalance in equity markets. *Quantitative Finance*, 23(10):1373–1393, 2023.
- [10] E. De Vito, V. Umanità, and S. Villa. An extension of mercer theorem to matrix-valued measurable kernels. *Applied and Computational Harmonic Analysis*, 34(3):339–351, 2013.
- [11] I. Ekeland and R. Temam. *Convex analysis and variational problems*. SIAM, 1999.
- [12] I. Ekren and J. Muhle-Karbe. Portfolio choice with small temporary and transient price impact. *Mathematical Finance*, 29(4):1066–1115, 2019.
- [13] N. Gârleanu and L. H. Pedersen. Dynamic trading with predictable returns and transaction costs. *The Journal of Finance*, 68(6):2309–2340, 2013. doi: <https://doi.org/10.1111/jofi.12080>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12080>.
- [14] N. Gârleanu and L. H. Pedersen. Dynamic portfolio choice with frictions. *Journal of Economic Theory*, 165:487–516, 2016.
- [15] J. Gatheral. No-dynamic-arbitrage and market impact. *Quant. Finance*, 10:749–759, 2010.
- [16] J. Gatheral, A. Schied, and A. Slynko. Transient linear price impact and Fredholm integral equations. *Math. Finance*, 22:445–474, 2012.
- [17] G. Gripenberg, S.-O. Londen, and O. Staffans. *Volterra integral and functional equations*. Encyclopedia of mathematics and its applications 34. Cambridge University Press, 1990.
- [18] U. Horst and X. Xia. Multi-dimensional optimal trade execution under stochastic resilience. *Finance and Stochastics*, 23:889–923, 2019.
- [19] G. Huberman and W. Stanzl. Price manipulation and quasi-arbitrage. *Econometrica*, 72(4):1247–1275, 2004.

- [20] V. Le Coz, I. Mastromatteo, D. Challet, and M. Benzaquen. When is cross impact relevant? *arXiv preprint arXiv:2305.16915*, 2023.
- [21] C.-A. Lehalle and E. Neuman. Incorporating signals into optimal trading. *Finance and Stochastics*, 23:275–311, 2019.
- [22] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 2024/02/28/ 1952. doi: 10.2307/2975974. URL <http://www.jstor.org/stable/2975974>.
- [23] I. Mastromatteo, M. Benzaquen, Z. Eisler, and J.-P. Bouchaud. Trading lightly: Cross-impact and optimal portfolio execution. *arXiv preprint arXiv:1702.03838*, 2017.
- [24] A. A. Obizhaeva and J. Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial markets*, 16(1):1–32, 2013.
- [25] M. Schneider and F. Lillo. Cross-impact and no-dynamic-arbitrage. *Quantitative Finance*, 19(1):137–154, 2019. doi: 10.1080/14697688.2018.1467033.
- [26] M. Tomas, I. Mastromatteo, and M. Benzaquen. How to build a cross-impact model from first principles: Theoretical requirements and empirical results. *Quantitative Finance*, 22(6):1017–1036, 2022.
- [27] A. Torchinsky. *Problems in real and functional analysis*, volume 166. American Mathematical Soc., 2015.
- [28] G. Tsoukalas, J. Wang, and K. Giesecke. Dynamic portfolio execution. *Management Science*, 65(5):2015–2040, 2019.