

WORKING PAPER

A Stochastic Maximum Principle for Volterra SDEs

Adjoint Equations and Pontryagin Necessary Conditions under Rough Drivers

Supervised by Dr. T. Zamrik

19-SEP-2024



Phase portrait of the Brusselator: $\dot{x} = a + x^2y - (b+1)x$, $\dot{y} = bx - x^2y$, $a = 1.05$, $b = 2.60$



WP-2026-48040577
zamrik.com



Zamrik Quant Research

The research arm of Zamrik Quant Lab

Copyright

© 2024 T. Zamrik. All rights reserved. Published by Zamrik Quant Research. No part of this document may be reproduced, stored, distributed or adapted, in whole or in part, by any means, without the prior written permission of the publisher, except for brief quotation with full citation. Requests for permission: info@zamrik.com.

Suggested Citation

Optimal Control & Decision Desk, Zamrik Quant Research (2024). 'A Stochastic Maximum Principle for Volterra SDEs: Adjoint Equations and Pontryagin Necessary Conditions under Rough Drivers'. *Zamrik Quant Research Working Paper WP-2026-48040577*. Available at <https://zamrik.com/research-items/wp-2026-48040577/>.

About Zamrik Quant Research

Zamrik Quant Research is the research arm of Zamrik Quant Lab. It works through five research desks, from market microstructure and derivatives to optimal control, stochastic analysis and applied mathematics, and publishes everything it finds in the open: research papers, notebooks that run on Kaggle, free tools that answer one practical question, and essays.

The five desks are Market Microstructure & Trading; Derivatives & Volatility; Optimal Control & Decision; Stochastic Analysis; and Applied Mathematics. Each paper, notebook, tool and essay belongs to exactly one desk. Zamrik Quant Research is run by a small, independent team, headed by Dr. T. Zamrik. We are self-funded: no sponsors, no advertisers, no outside capital.

Every publication passes through the same five functions before anyone outside sees it: research, data, software, verification and production. We never retrofit the mathematics to a conclusion chosen in advance. Every number is checked two independent ways. Every figure is re-run from its code. When we find an error, or replace a piece with a better one, we record it in the public corrections log at <https://zamrik.com/corrections/>.

Legal Notice

This paper is research and education. Nothing in it is investment, financial, legal or engineering advice; nothing is a recommendation to buy or sell shares of stock, derivative contracts, bonds, or any other instrument or property; and nothing is a forecast. You are responsible for your own decisions. Figures and code are provided as they are. Care and independent checks reduce errors but cannot rule them out, so no warranty is given. The views expressed are those of the Optimal Control & Decision Desk.

WORKING PAPER



A Stochastic Maximum Principle for Volterra SDEs

*Adjoint Equations and Pontryagin Necessary Conditions under
Rough Drivers*

Supervised by Dr. T. Zamrik¹

19-SEP-2024

WP-2026-48040577

Abstract. We establish a stochastic maximum principle for finite-horizon optimal control of stochastic Volterra integral equations with the fractional kernel $k(t) = t^{H-1/2}/\Gamma(H+1/2)$, Hurst index $H \in (0, 1/2]$, Lipschitz coefficients and a convex compact control set. The adjoint process solves a backward stochastic Volterra integral equation (BSVIE) in which the forward kernel is read over later times and the gradient of the terminal cost enters with the singular weight $k(T-t)$. An optimal control satisfies the first-order form of the maximum condition, which is pointwise maximisation of the Hamiltonian when the Hamiltonian is concave in the control; when the terminal cost is convex and the Hamiltonian is jointly concave in state and control, the condition is also sufficient. A linear-quadratic example has a closed-form adjoint in Mittag-Leffler functions, which we check against direct numerical solutions.

WP-2026-48040577 · Optimal Control & Decision Desk · Working Paper · Published
19-SEP-2024 · Reading time: about 21 min

1. Introduction

The classical stochastic maximum principle (Yong–Zhou [1], Peng [2]) gives necessary conditions for optimality in controlled stochastic differential equations driven by Brownian motion: the adjoint process solves a backward stochastic differential equation (BSDE), an optimal control maximises the Hamiltonian, and under concavity the condition is also sufficient. The derivation rests on the semimartingale structure of the state, through Itô’s product rule for the adjoint and the variation of the state.

Rough-volatility models remove that structure. Estimates of the regularity of realised log-volatility give a Hurst index of order 0.1 (Gatheral–Jaisson–Rosenbaum [3]), and pricing under such models (Bayer–Friz–Gatheral [4]) works with Volterra dynamics,

$$X_t = x_0 + \int_0^t k(t-s) b(s, X_s, u_s) ds + \int_0^t k(t-s) \sigma(s, X_s, u_s) dW_s, \quad (1.1)$$

in which the kernel k carries the memory of the driver. For a fractional kernel with $H < 1/2$ the state is neither a Markov process nor a semimartingale, and the adjoint BSDE does not apply.

¹Correspondence: Optimal Control & Decision Desk, Zamrik Quant Research, info@zamrik.com.



Yong [5] introduced backward stochastic Volterra integral equations and used them as adjoint equations in a Pontryagin-type maximum principle for stochastic Volterra integral equations; the linear–quadratic problem for such equations was studied by Wang [6]. This paper gives a self-contained treatment for the singular fractional kernel. Two features distinguish the result from the Markovian one: the adjoint reads the forward kernel over later times, and the gradient of the terminal cost enters with the weight $k(T - t)$, which is singular at maturity — a control applied just before T moves the terminal state with an unbounded weight.

1.1 Contributions

1. **Forward equation (Theorem 1, Lemma 1).** The controlled Volterra SDE has a unique strong solution, with Hölder paths of every order below H . The solution depends Hölder-continuously, not Lipschitz-continuously, on the control in the L^2 norm when $H < 1/2$, and Lemma 1 shows that the Lipschitz bound fails.
2. **Adjoint BSVIE (Definition 6, Theorem 2).** For each admissible control the adjoint equation has a unique solution, by a contraction in a weighted norm.
3. **Duality (Lemma 3).** The derivative of the cost in a direction of control is an integral of the control gradient of the Hamiltonian along the adjoint.
4. **Maximum principle (Theorems 3 and 4).** The first-order necessary condition, which is pointwise maximisation under concavity in the control, and its sufficiency under joint concavity.
5. **Example (Proposition 2).** A linear–quadratic problem with a closed-form adjoint and optimal control, checked numerically in §6.

§2 fixes the setting, §3 treats the forward equation, §4 the adjoint and the duality identity, §5 the maximum principle, §6 the example and its numerical check, and §7 concludes. Appendix A collects the kernel estimates.

2. Preliminaries

Throughout, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$ is a complete filtered probability space supporting a d -dimensional Brownian motion W , $(\mathcal{F}_t)$ is its augmented natural filtration, and $T \in (0, \infty)$ is a fixed horizon. All processes are progressively measurable. We write $\mathbb{S}^p$ for the processes ξ with $\sup_t \mathbb{E} \|\xi_t\|^p < \infty$, $\mathbb{L}$ for those with $\int_0^T (\mathbb{E} \|\xi_t\|^2)^{1/2} dt < \infty$, and $L^2_{\mathcal{F}}$ for those with $\mathbb{E} \int_0^T \|\xi_t\|^2 dt < \infty$. For $\xi \in L^2(\mathcal{F}_T)$ the martingale representation theorem gives $\xi = \mathbb{E}\xi + \int_0^T \zeta_r^\xi dW_r$ with a unique $\zeta^\xi \in L^2_{\mathcal{F}}$, the *representation integrand* of ξ . Matrices are paired by $\langle A, B \rangle = \text{tr}(A^\top B)$.

2.1 Volterra SDE and the fractional kernel

The state process is the solution of a controlled stochastic Volterra integral equation.



Definition 1 (Controlled Volterra SDE). For a kernel $k : (0, T] \rightarrow \mathbb{R}_+$, coefficients $b : [0, T] \times \mathbb{R}^n \times U \rightarrow \mathbb{R}^n$, $\sigma : [0, T] \times \mathbb{R}^n \times U \rightarrow \mathbb{R}^{n \times d}$ and an admissible control u (Definition 3), the *controlled Volterra SDE* is

$$X_t = x_0 + \int_0^t k(t-s) b(s, X_s, u_s) ds + \int_0^t k(t-s) \sigma(s, X_s, u_s) dW_s, \quad t \in [0, T],$$

with initial value $x_0 \in \mathbb{R}^n$.

Definition 2 (Fractional kernel). For a Hurst index $H \in (0, 1/2]$ and $\alpha := H + 1/2$, the *fractional kernel* is

$$k(r) = \frac{r^{H-1/2}}{\Gamma(H+1/2)} = \frac{r^{\alpha-1}}{\Gamma(\alpha)}, \quad r \in (0, T].$$

For $H < 1/2$ it is singular at $r = 0$ but square-integrable on $[0, T]$. For $H = 1/2$, $k \equiv 1$ and Definition 1 is an ordinary Itô SDE. For $H < 1/2$ the increments of the stochastic convolution $\int_0^t k(t-s) dW_s$ have variance of order $|t-s|^{2H}$, so its paths are rougher than Brownian ones.

The kernel parametrises a family of Volterra equations from Brownian dynamics ($H = 1/2$) to arbitrarily rough ones ($H \rightarrow 0^+$); Decreusefond–Üstünel [7] develop the stochastic analysis of the corresponding fractional processes. Appendix A collects the kernel estimates used below.

2.2 Admissible controls

The control values lie in a fixed convex compact set $U \subset \mathbb{R}^m$.

Definition 3 (Admissible controls). The set of *admissible controls* is

$$\mathcal{U}_{\text{ad}} = \left\{ u : [0, T] \times \Omega \rightarrow U \mid u \text{ is progressively measurable} \right\}.$$

No integrability condition is needed: every $u \in \mathcal{U}_{\text{ad}}$ is bounded by the diameter of U , which we denote D .

2.3 Cost functional

The decision-maker chooses $u \in \mathcal{U}_{\text{ad}}$ to minimise a cost.

Definition 4 (Cost functional). For a running cost $f : [0, T] \times \mathbb{R}^n \times U \rightarrow \mathbb{R}$ and a terminal cost $g : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$J(u) = \mathbb{E} \left[\int_0^T f(t, X_t^u, u_t) dt + g(X_T^u) \right], \quad J^* = \inf_{u \in \mathcal{U}_{\text{ad}}} J(u),$$

where X^u solves Definition 1 under the control u .



2.4 Standing assumptions

Assumption 1 (Lipschitz coefficients). The functions b, σ, f, g are continuous and Lipschitz in (x, u) uniformly in t :

$$|b(t, x_1, u_1) - b(t, x_2, u_2)| + |\sigma(t, x_1, u_1) - \sigma(t, x_2, u_2)| \leq L(\|x_1 - x_2\| + \|u_1 - u_2\|),$$

and likewise for f and g . The control set U is convex and compact.

Assumption 2 (Differentiability). The functions b, σ, f are continuously differentiable in (x, u) on $[0, T] \times \mathbb{R}^n \times U$ with derivatives bounded by L , and g is continuously differentiable with $\|g_x\|_\infty \leq L$.

Assumption 1 gives the forward well-posedness of §3; Assumption 2 is used from Proposition 1 on.

2.5 The Hamiltonian

The maximum principle is stated with the following Hamiltonian.

Definition 5 (Hamiltonian). For $(t, x, u, y, z) \in [0, T] \times \mathbb{R}^n \times U \times \mathbb{R}^n \times \mathbb{R}^{n \times d}$,

$$\mathcal{H}(t, x, u, y, z) = \langle y, b(t, x, u) \rangle + \langle z, \sigma(t, x, u) \rangle - f(t, x, u).$$

This is the sign convention of Yong–Zhou [1] for a minimisation problem: the running cost enters with a minus sign, the adjoint of §4 is minus the gradient of the cost-to-go, and an optimal control maximises $\mathcal{H}$. Under Assumption 2, $\mathcal{H}$ is C^1 in (x, u) and affine in (y, z) .

3. The Controlled Volterra SDE

This section shows that Definition 1 is well-posed, with Hölder paths, and how its solution depends on the control. The argument is a fixed point in a weighted norm that absorbs the singularity of the kernel.

3.1 Forward well-posedness

Theorem 1 (Existence and uniqueness). Under Assumption 1, with the fractional kernel of Definition 2, for every $u \in \mathcal{U}_{\text{ad}}$ the controlled Volterra SDE has a unique solution $X^u \in \mathbb{S}^2$, and $X^u \in \mathbb{S}^p$ for every $p \geq 2$. It has a continuous modification which is Hölder continuous of every order $\gamma < H$: for each such γ there is a finite random variable C_γ with

$$\|X_t^u - X_s^u\| \leq C_\gamma |t - s|^\gamma, \quad 0 \leq s \leq t \leq T.$$

Proof. Let $\mathcal{T}[X]_t$ be the right-hand side of Definition 1 with X in place of the solution, and $\|X\|_\beta^2 := \sup_t e^{-\beta t} \mathbb{E}\|X_t\|^2$ on $\mathbb{S}^2$. With $K_T := \int_0^T k$, the Cauchy–Schwarz inequality



on the drift and the Itô isometry on the diffusion give, by Assumption 1,

$$\mathbb{E}\|\mathcal{T}[X]_t - \mathcal{T}[\tilde{X}]_t\|^2 \leq 2L^2 \int_0^t \kappa(t-s) \mathbb{E}\|X_s - \tilde{X}_s\|^2 ds, \quad \kappa(r) := K_T k(r) + k(r)^2,$$

hence $\|\mathcal{T}[X] - \mathcal{T}[\tilde{X}]\|_\beta^2 \leq 2L^2 (\int_0^T \kappa(r) e^{-\beta r} dr) \|X - \tilde{X}\|_\beta^2$. The integral is finite because $k \in L^1 \cap L^2$ (Lemma 4) and tends to 0 as $\beta \rightarrow \infty$ by dominated convergence; its k^2 part is at most $\Gamma(2H) \beta^{-2H} / \Gamma(\alpha)^2$. So $\mathcal{T}$, which maps $\mathbb{S}^2$ into itself because b and σ grow at most linearly in x , is a contraction for large β , and the Banach fixed-point theorem gives existence and uniqueness. For $p > 2$ the same estimate with the Burkholder–Davis–Gundy inequality, and Jensen’s inequality for the probability measure proportional to $k(t-s)^2 ds$, gives $X \in \mathbb{S}^p$. For the increments, the drift part of $X_t - X_s$ has L^p norm at most $C |t-s|^\alpha$ by Minkowski’s inequality and the computation of Lemma 5, and the stochastic part

$$\int_s^t k(t-r) \sigma_r dW_r + \int_0^s [k(t-r) - k(s-r)] \sigma_r dW_r$$

has p -th moment at most $C_p \sup_r \mathbb{E}\|\sigma_r\|^p (\int_0^{t-s} k^2 + \int_0^\infty [k(r+t-s) - k(r)]^2 dr)^{p/2} = C |t-s|^{pH}$ by Lemma 4 (i) and (iii). Kolmogorov’s criterion gives a modification that is Hölder of every order $\gamma < H - 1/p$, and p is arbitrary. $\square$

Remark 1. When σ is bounded away from zero the order H cannot be improved: the increment variance is then of exact order $|t-s|^{2H}$, and in the example of §6 its observed order is computed in Table 2. At $H = 1/2$, Theorem 1 is the classical existence theorem for Lipschitz SDEs.

3.2 Stability with respect to the control

The dependence of the state on the control is where the singular kernel first changes the Markovian picture.

Lemma 1 (Control stability). Under Assumption 1 there is a constant $C = C(T, L, H)$ such that, for $u_1, u_2 \in \mathcal{U}_{\text{ad}}$ and κ as in the proof of Theorem 1,

$$\sup_{t \in [0, T]} \mathbb{E}\|X_t^{u_1} - X_t^{u_2}\|^2 \leq C \sup_{t \in [0, T]} \mathbb{E} \int_0^t \kappa(t-s) \|u_1(s) - u_2(s)\|^2 ds.$$

Consequently, for every $\theta \in (0, 2H)$ (and $\theta = 1$ when $H = 1/2$) there is C_θ with

$$\sup_{t \in [0, T]} \mathbb{E}\|X_t^{u_1} - X_t^{u_2}\|^2 \leq C_\theta \left(\mathbb{E} \int_0^T \|u_1(s) - u_2(s)\|^2 ds \right)^\theta.$$

For $H < 1/2$ the exponent $\theta = 1$ is not available: with $b = 0$, $\sigma(t, x, u) = u$, $n = d = m = 1$, $U = [0, c]$, $u_1 = c \mathbf{1}_{[T-\delta, T]}$ and $u_2 = 0$, the left-hand side is at least $c^2 \delta^{2H} / (2H \Gamma(\alpha)^2)$ while $\mathbb{E} \int_0^T \|u_1 - u_2\|^2 = c^2 \delta$.

Proof. Split each coefficient difference into an X -difference and a u -difference. As in Theorem 1, $\varphi(t) := \mathbb{E}\|X_t^{u_1} - X_t^{u_2}\|^2$ satisfies $\varphi(t) \leq 4L^2 \int_0^t \kappa(t-s) [\varphi(s) + \mathbb{E}\|u_1(s) -$



$u_2(s)\|^2] ds$. In the weighted norm of Theorem 1, choose β with $4L^2 \int_0^T \kappa(r)e^{-\beta r} dr \leq 1/2$; then $\sup_t e^{-\beta t} \varphi(t) \leq 8L^2 \sup_t \int_0^t \kappa(t-s) \mathbb{E}\|u_1(s) - u_2(s)\|^2 ds$, which is the first bound with $C = 8L^2 e^{\beta T}$. For the second, take $1 < q < 1/(1-2H)$, so that $\kappa \in L^q(0, T)$, and $q' = q/(q-1)$. Hölder's inequality and $\|u_1 - u_2\| \leq D$ give

$$\int_0^t \kappa(t-s) \mathbb{E}\|u_1(s) - u_2(s)\|^2 ds \leq \|\kappa\|_{L^q} D^{2(1-1/q')} \left(\mathbb{E} \int_0^T \|u_1(s) - u_2(s)\|^2 ds \right)^{1/q'},$$

and $\theta = 1/q' = 1 - 1/q$ ranges over $(0, 2H)$; for $H = 1/2$, κ is bounded and $q = \infty$. The counterexample is the Itô isometry at $t = T$ with $\int_0^\delta k^2 = \delta^{2H}/(2H \Gamma(\alpha)^2)$. $\square$

The constant grows exponentially in T through $e^{\beta T}$. The loss of the Lipschitz bound is a feature of the kernel: a control change concentrated on a short interval before t is amplified by the singularity of k .

Lemma 2 (Cost functional well-posedness). Under Assumption 1 the cost J is finite on $\mathcal{U}_{\text{ad}}$ and continuous for the norm $(\mathbb{E} \int_0^T \|u_t\|^2 dt)^{1/2}$.

Proof. By Assumption 1 and compactness of U , $|f(t, x, u)| + |g(x)| \leq C(1 + \|x\|)$, and Theorem 1 gives $X^u \in \mathbb{S}^2$, so $J(u)$ is finite. By the Lipschitz property, $|J(u_1) - J(u_2)| \leq L(T+1) \sup_t (\mathbb{E}\|X_t^{u_1} - X_t^{u_2}\|^2)^{1/2} + L\sqrt{T} (\mathbb{E} \int_0^T \|u_1 - u_2\|^2)^{1/2}$, and Lemma 1 bounds the supremum by a positive power of the same norm. $\square$

3.3 Forward variational equation

For the maximum principle of §5 we need the first-order variation of the state. Fix $u^*, v \in \mathcal{U}_{\text{ad}}$, put $\delta u := v - u^*$ and $u^\varepsilon := u^* + \varepsilon \delta u$ for $\varepsilon \in (0, 1]$, which is admissible because U is convex, and write $X^* := X^{u^*}$, $X^\varepsilon := X^{u^\varepsilon}$, $b_x(s) := b_x(s, X_s^*, u_s^*)$ and similarly for b_u, σ_x, σ_u .

Proposition 1 (Forward variation). Under Assumptions 1–2, $(X^\varepsilon - X^*)/\varepsilon \rightarrow \eta$ in $\mathbb{S}^2$ as $\varepsilon \downarrow 0$, where η is the unique solution in $\mathbb{S}^2$ of the linear Volterra equation

$$\eta_t = \int_0^t k(t-s) [b_x(s)\eta_s + b_u(s)\delta u_s] ds + \int_0^t k(t-s) [\sigma_x(s)\eta_s + \sigma_u(s)\delta u_s] dW_s.$$

Proof. The linear equation has coefficients bounded by L , so Theorem 1 gives its unique solution, and $\eta \in \mathbb{S}^p$ for every p . The quotient $\eta^\varepsilon := (X^\varepsilon - X^*)/\varepsilon$ solves the same equation with $b_x(s)$ replaced by the average $\int_0^1 b_x(s, X_s^* + \lambda(X_s^\varepsilon - X_s^*), u_s^* + \lambda\varepsilon \delta u_s) d\lambda$, and likewise for b_u, σ_x, σ_u . The difference $\eta^\varepsilon - \eta$ therefore solves a linear equation with coefficients bounded by L and a forcing ρ^ε made of the differences between averaged and exact derivatives, multiplied by η or δu . The estimate of Theorem 1 gives $\sup_t \mathbb{E}\|\eta_t^\varepsilon - \eta_t\|^2 \leq C \sup_t \int_0^t \kappa(t-s) \mathbb{E}\|\rho_s^\varepsilon\|^2 ds$. By Lemma 1, $X^\varepsilon \rightarrow X^*$ in $\mathbb{S}^2$, so the averaged derivatives converge in probability for each s ; they are bounded by L , so $\mathbb{E}\|\rho_s^\varepsilon\|^2 \rightarrow 0$ for each s by dominated convergence, with a bound uniform in s and ε . Hölder's inequality with $\kappa \in L^q$, as in Lemma 1, turns this into uniform convergence of the right-hand side to 0. $\square$



Proposition 1 is the bridge to the adjoint: the duality identity of Lemma 3 pairs η with the adjoint and removes it from the derivative of the cost.

4. The Adjoint Backward Volterra Equation

This section defines the adjoint process, proves that it exists and is unique, and proves the duality identity that links it to the forward variation.

4.1 The adjoint BSVIE

Fix $u \in \mathcal{U}_{\text{ad}}$ with state $X = X^u$, and write $b_x(t) := b_x(t, X_t, u_t)$ and similarly for the other derivatives. For a process $P \in \mathbb{L}$ let ζ^g be the representation integrand of $g_x(X_T)$ and ζ^{P_s} that of P_s , and define

$$Y_t[P] := \mathbb{E} \left[-k(T-t) g_x(X_T) + \int_t^T k(s-t) P_s \, ds \mid \mathcal{F}_t \right], \quad (4.1)$$

$$Z_t[P] := -k(T-t) \zeta_t^g + \int_t^T k(s-t) \zeta_t^{P_s} \, ds. \quad (4.2)$$

Definition 6 (Adjoint BSVIE). The *adjoint equation* associated with u is the fixed-point problem

$$P_t = \mathcal{H}_x(t, X_t, u_t, Y_t[P], Z_t[P]), \quad dt \otimes d\mathbb{P}\text{-a.e.},$$

and the *adjoint pair* is $(Y, Z) := (Y[P], Z[P])$. Equivalently, with $\Theta_t := -k(T-t) g_x(X_T) + \int_t^T k(s-t) \mathcal{H}_x(s, X_s, u_s, Y_s, Z_s) \, ds$ and $\mathcal{Z}(t, \cdot)$ the representation integrand of Θ_t , the pair solves the backward stochastic Volterra integral equation

$$Y_t = \Theta_t - \int_t^T \mathcal{Z}(t, r) \, dW_r, \quad t \in [0, T],$$

and Z_t is the value of $\mathcal{Z}(t, \cdot)$ on the diagonal $r = t$, which the formula for $Z_t[P]$ defines for every t .

Here $\mathcal{H}_x = b_x^\top Y + \sum_j (\sigma_x^j)^\top Z^j - f_x$, with σ^j and Z^j the j -th columns, is affine in (Y, Z) with coefficients bounded by L .

Remark 2 (What the kernel does). At $H = 1/2$, $k \equiv 1$, $\Theta_t = \Theta_0 - \int_0^t \mathcal{H}_x \, ds$, and $\mathcal{Z}(t, r)$ does not depend on t for $r \geq t$; Definition 6 becomes the adjoint BSDE of Yong–Zhou [1], $Y_t = -g_x(X_T) + \int_t^T \mathcal{H}_x \, ds - \int_t^T Z_r \, dW_r$. For $H < 1/2$ two things change. The adjoint at t weighs the Hamiltonian gradient at later times s by $k(s-t)$, the forward kernel read over later times. And the terminal gradient enters with the weight $k(T-t)$, singular at maturity: a control applied just before T moves X_T with an unbounded weight, so Y_t can be unbounded as $t \rightarrow T$, as it is in §6. The same weight is why Z lies in general in $\mathbb{L}$ rather than in $L_{\mathcal{F}}^2$.



4.2 Existence and uniqueness

Theorem 2 (BSVIE existence and uniqueness). Under Assumptions 1–2, with the fractional kernel of Definition 2, for every $u \in \mathcal{U}_{\text{ad}}$ the adjoint equation of Definition 6 has a unique solution $P \in \mathbb{L}$, and $Y, Z \in \mathbb{L}$. If g is affine, $P, Y, Z \in L^2_{\mathcal{F}}$.

Proof. Let $\|\xi\|_{\beta} := \int_0^T e^{\beta t} (\mathbb{E}\|\xi_t\|^2)^{1/2} dt$, a norm equivalent to that of $\mathbb{L}$. The map $\Phi(P) := \mathcal{H}_x(\cdot, X, u, Y[P], Z[P])$ is affine; its linear part is $P \mapsto b_x^{\top} Y^0[P] + \sum_j (\sigma_x^j)^{\top} Z^0[P]^j$, where Y^0, Z^0 omit the terms in g . By the conditional Jensen and Minkowski inequalities, $(\mathbb{E}\|Y_t^0[P]\|^2)^{1/2} \leq \int_t^T k(s-t) (\mathbb{E}\|P_s\|^2)^{1/2} ds$, so by Fubini

$$\|Y^0[P]\|_{\beta} \leq \left(\int_0^T k(r) e^{-\beta r} dr \right) \|P\|_{\beta}.$$

For Z^0 , Minkowski's inequality, Fubini and the Cauchy–Schwarz inequality in t give

$$\|Z^0[P]\|_{\beta} \leq \int_0^T e^{\beta s} \left(\int_0^s k(s-t)^2 e^{-2\beta(s-t)} dt \right)^{1/2} \left(\mathbb{E} \int_0^s \|\zeta_t^{P_s}\|^2 dt \right)^{1/2} ds,$$

and the Itô isometry $\mathbb{E} \int_0^s \|\zeta_t^{P_s}\|^2 dt \leq \mathbb{E}\|P_s\|^2$ bounds the right-hand side by $(\int_0^T k(r)^2 e^{-2\beta r} dr)^{1/2} \|P\|_{\beta}$. Both prefactors tend to 0 as $\beta \rightarrow \infty$ because $k \in L^1 \cap L^2$ (Lemma 4), and the derivatives are bounded by L , so Φ is a contraction on $\mathbb{L}$ for large β . Its constant term $\Phi(0)$ lies in $\mathbb{L}$: f_x is bounded, and the terms in g , $Y_t^g := -\mathbb{E}[k(T-t)g_x(X_T) | \mathcal{F}_t]$ and $Z_t^g := -k(T-t)\zeta_t^g$, satisfy $(\mathbb{E}\|Y_t^g\|^2)^{1/2} \leq Lk(T-t)$ and $\int_0^T (\mathbb{E}\|Z_t^g\|^2)^{1/2} dt \leq \|k\|_{L^2} (\mathbb{E}\|g_x(X_T)\|^2)^{1/2} \leq L\|k\|_{L^2}$. The Banach fixed-point theorem gives a unique $P \in \mathbb{L}$, and the two bounds put Y and Z in $\mathbb{L}$. If g is affine, $Z^g = 0$ and Y^g is deterministic and square-integrable, and the same estimates with the Cauchy–Schwarz inequality in place of Minkowski's run the contraction in $L^2_{\mathcal{F}}$ with the norm $(\mathbb{E} \int_0^T e^{\beta t} \|\xi_t\|^2 dt)^{1/2}$. $\square$

Theorem 2 is the Volterra analogue of the existence theorem of Pardoux–Peng [8] for BSDEs. Its Picard iteration is the scheme used in §6.

4.3 The duality identity

The identity that links Proposition 1 to Definition 6 is a change of the order of integration together with the Itô isometry.

Lemma 3 (Duality). Under Assumptions 1–2 let $u \in \mathcal{U}_{\text{ad}}$ have state X , adjoint (Y, Z) and $P_t = \mathcal{H}_x(t) := \mathcal{H}_x(t, X_t, u_t, Y_t, Z_t)$, and let $\Delta_t = \int_0^t k(t-s)a_s ds + \int_0^t k(t-s)m_s dW_s$ with $a \in \mathbb{S}^2$ and $m \in \mathbb{S}^2$. Then

$$\mathbb{E} \left[-\langle g_x(X_T), \Delta_T \rangle + \int_0^T \langle \mathcal{H}_x(t), \Delta_t \rangle dt \right] = \mathbb{E} \int_0^T [\langle Y_t, a_t \rangle + \langle Z_t, m_t \rangle] dt.$$



In particular, for $u = u^*$ and the forward variation η of Proposition 1,

$$\mathbb{E}\left[\langle g_x(X_T^*), \eta_T \rangle + \int_0^T \langle f_x(t), \eta_t \rangle dt\right] = -\mathbb{E}\int_0^T \langle \mathcal{H}_u(t) + f_u(t), \delta u_t \rangle dt.$$

Proof. For almost every t , $P_t \in L^2(\Omega)$, and conditioning on $\mathcal{F}_s$ and the Itô isometry give

$$\mathbb{E}\langle P_t, \Delta_t \rangle = \int_0^t k(t-s) \mathbb{E}[\langle \mathbb{E}[P_t | \mathcal{F}_s], a_s \rangle + \langle \zeta_s^{P_t}, m_s \rangle] ds,$$

and the same holds at $t = T$ with P_T replaced by $\xi := -g_x(X_T)$. Integrate over $t \in [0, T]$ and change the order of integration by Lemma 6; the double integrals converge absolutely by the bounds in the proof of Theorem 2 and $a, m \in \mathbb{S}^2$. The left-hand side becomes $\mathbb{E}\int_0^T [\langle Y_s, a_s \rangle + \langle Z_s, m_s \rangle] ds$, because the factors multiplying a_s and m_s are $\mathbb{E}[k(T-s)\xi + \int_s^T k(t-s)P_t dt | \mathcal{F}_s] = Y_s$ and $k(T-s)\zeta_s^\xi + \int_s^T k(t-s)\zeta_s^{P_t} dt = Z_s$ of Definition 6. For the second statement take $\Delta = \eta$, $a = b_x \eta + b_u \delta u$ and $m = \sigma_x \eta + \sigma_u \delta u$: the terms $\langle Y, b_x \eta \rangle + \langle Z, \sigma_x \eta \rangle = \langle \mathcal{H}_x + f_x, \eta \rangle$ cancel the $\mathcal{H}_x$ term on the left, and $\langle Y, b_u \delta u \rangle + \langle Z, \sigma_u \delta u \rangle = \langle \mathcal{H}_u + f_u, \delta u \rangle$. $\square$

Lemma 3 holds for any Δ of Volterra form, not only for the linearisation η ; Theorem 4 uses it with $\Delta = X^u - X^*$, which is how the sufficiency proof avoids a second-order expansion.

5. The Stochastic Maximum Principle

This section combines the forward variation (Proposition 1), the adjoint (Theorem 2) and the duality identity (Lemma 3) into the necessary and sufficient conditions.

5.1 Necessary condition

Theorem 3 (Necessary condition). Under Assumptions 1–2 let $u^* \in \mathcal{U}_{\text{ad}}$ be optimal, $J(u^*) = J^*$, with state X^* and adjoint (Y, Z) (Theorem 2), and write $\mathcal{H}_u(t) := \mathcal{H}_u(t, X_t^*, u_t^*, Y_t, Z_t)$. Then, $dt \otimes d\mathbb{P}$ -almost everywhere,

$$\langle \mathcal{H}_u(t), v - u_t^* \rangle \leq 0 \quad \text{for every } v \in U.$$

If moreover $u \mapsto \mathcal{H}(t, X_t^*, u, Y_t, Z_t)$ is concave on U , this is equivalent to

$$\mathcal{H}(t, X_t^*, v, Y_t, Z_t) \leq \mathcal{H}(t, X_t^*, u_t^*, Y_t, Z_t) \quad \text{for every } v \in U.$$

Proof. Let $v \in \mathcal{U}_{\text{ad}}$, $\delta u = v - u^*$ and $u^\varepsilon = u^* + \varepsilon \delta u$. Writing the differences of f and g with averaged derivatives, as in the proof of Proposition 1, and using the convergence of Proposition 1,

$$\lim_{\varepsilon \downarrow 0} \frac{J(u^\varepsilon) - J(u^*)}{\varepsilon} = \mathbb{E}\left[\int_0^T (\langle f_x(t), \eta_t \rangle + \langle f_u(t), \delta u_t \rangle) dt + \langle g_x(X_T^*), \eta_T \rangle\right],$$

and by Lemma 3 the right-hand side equals $-\mathbb{E}\int_0^T \langle \mathcal{H}_u(t), \delta u_t \rangle dt$. Optimality makes



the limit non-negative, so $\mathbb{E} \int_0^T \langle \mathcal{H}_u(t), v_t - u_t^* \rangle dt \leq 0$ for every $v \in \mathcal{U}_{\text{ad}}$. For $w \in U$ and a progressively measurable set $A \subset [0, T] \times \Omega$, the control $v := w \mathbf{1}_A + u^* \mathbf{1}_{A^c}$ is admissible, so $\mathbb{E} \int_0^T \mathbf{1}_A \langle \mathcal{H}_u(t), w - u_t^* \rangle dt \leq 0$ for every such A ($\mathcal{H}_u$ is integrable because $Y, Z \in \mathbb{L}$), hence $\langle \mathcal{H}_u(t), w - u_t^* \rangle \leq 0$ almost everywhere for each w . A countable dense subset of U and continuity in w give the first statement. If $\mathcal{H}$ is concave in u , then $\mathcal{H}(t, X_t^*, v, Y_t, Z_t) \leq \mathcal{H}(t, X_t^*, u_t^*, Y_t, Z_t) + \langle \mathcal{H}_u(t), v - u_t^* \rangle$, which gives the maximum; conversely, a maximiser over the convex set U satisfies the first-order condition. $\square$

Remark 3 (Comparison with the Markovian case). At $H = 1/2$ the adjoint is the BSDE of Remark 2, and Theorem 3 is the maximum principle of Yong–Zhou [1] for a convex control domain. Without concavity in the control, the classical global maximum condition is obtained by spike variations and, when the diffusion depends on the control, a second-order adjoint equation (Peng [2]); spike variations of a Volterra equation with a singular kernel are not treated here.

5.2 Sufficient condition

Under convexity of the terminal cost and concavity of the Hamiltonian, the first-order condition is also sufficient.

Theorem 4 (Sufficient condition). Under Assumptions 1–2 let $u^* \in \mathcal{U}_{\text{ad}}$ have state X^* and adjoint (Y, Z) . Suppose that g is convex, that $(x, u) \mapsto \mathcal{H}(t, x, u, Y_t, Z_t)$ is concave on $\mathbb{R}^n \times U$ for $dt \otimes d\mathbb{P}$ -almost every (t, ω) , and that u^* satisfies the condition of Theorem 3,

$$\langle \mathcal{H}_u(t, X_t^*, u_t^*, Y_t, Z_t), v - u_t^* \rangle \leq 0 \quad \forall v \in U, \quad dt \otimes d\mathbb{P}\text{-a.e.}$$

Then u^* is optimal: $J(u^*) = J^*$.

Proof. Let $u \in \mathcal{U}_{\text{ad}}$, $X = X^u$, $\Delta := X - X^*$, $\Delta b_t := b(t, X_t, u_t) - b(t, X_t^*, u_t^*)$ and $\Delta \sigma_t$ likewise, and write $\mathcal{H}(t)$ and $\mathcal{H}^*(t)$ for $\mathcal{H}$ at (t, X_t, u_t, Y_t, Z_t) and $(t, X_t^*, u_t^*, Y_t, Z_t)$, with the adjoint of u^* in both. Δ has the form of Lemma 3 with $a = \Delta b$ and $m = \Delta \sigma$, both in $\mathbb{S}^2$ by Theorem 1. By convexity, $g(X_T) - g(X_T^*) \geq \langle g_x(X_T^*), \Delta_T \rangle$, and by Definition 5, $f(t, X_t, u_t) - f(t, X_t^*, u_t^*) = \langle Y_t, \Delta b_t \rangle + \langle Z_t, \Delta \sigma_t \rangle - (\mathcal{H}(t) - \mathcal{H}^*(t))$. Lemma 3 gives $\mathbb{E} \langle g_x(X_T^*), \Delta_T \rangle = \mathbb{E} \int_0^T [\langle \mathcal{H}_x^*(t), \Delta_t \rangle - \langle Y_t, \Delta b_t \rangle - \langle Z_t, \Delta \sigma_t \rangle] dt$. Adding,

$$J(u) - J(u^*) \geq \mathbb{E} \int_0^T [\langle \mathcal{H}_x^*(t), \Delta_t \rangle - (\mathcal{H}(t) - \mathcal{H}^*(t))] dt \geq -\mathbb{E} \int_0^T \langle \mathcal{H}_u^*(t), u_t - u_t^* \rangle dt \geq 0,$$

by concavity of $\mathcal{H}$ in (x, u) , $\mathcal{H}(t) - \mathcal{H}^*(t) \leq \langle \mathcal{H}_x^*(t), \Delta_t \rangle + \langle \mathcal{H}_u^*(t), u_t - u_t^* \rangle$, and then by the first-order condition. $\square$

The hypotheses are those of a minimisation problem: g convex and $\mathcal{H}$ concave, which holds for instance when b and σ are affine in (x, u) and f is convex in (x, u) .



5.3 Discussion

Theorems 3 and 4 have the form of the Markovian maximum principle (Yong–Zhou [1]) — Hamiltonian, adjoint, first-order condition, concavity for sufficiency — with the BSDE replaced by the BSVIE of Definition 6 and Itô’s product rule by the duality of Lemma 3. The kernel enters only through Definition 6, so one proof covers $H = 1/2$ and $H < 1/2$.

6. A Linear–Quadratic Illustration

This section works one example in closed form and checks it numerically.

6.1 The example

Take $n = d = m = 1$, $b(t, x, u) = -\lambda x + u$, $\sigma \equiv \sigma_0$, $f(t, x, u) = cx + ru^2/2$, $g(x) = ax$ and $U = [-M, M]$, with $\lambda, r, M > 0$. Assumptions 1–2 hold (f is Lipschitz on $\mathbb{R} \times U$ because U is bounded), g is convex, and $\mathcal{H}(t, x, u, y, z) = y(-\lambda x + u) + \sigma_0 z - cx - ru^2/2$ is jointly concave in (x, u) . Write $E_{\alpha, \beta}(z) := \sum_{j \geq 0} z^j / \Gamma(\alpha j + \beta)$ for the Mittag-Leffler function, $E_\alpha := E_{\alpha, 1}$, $R(t) := t^{\alpha-1} E_{\alpha, \alpha}(-\lambda t^\alpha)$ and $\Pi_U(y) := \max(-M, \min(M, y))$.

Proposition 2 (Closed forms). In this example the adjoint of every admissible control is deterministic, $Z \equiv 0$, and with $\tau := T - t$,

$$Y_t = -a \tau^{\alpha-1} E_{\alpha, \alpha}(-\lambda \tau^\alpha) - \frac{c}{\lambda} (1 - E_\alpha(-\lambda \tau^\alpha)).$$

The control $u_t^* := \Pi_U(Y_t/r)$ is optimal, and every optimal control equals it $dt \otimes d\mathbb{P}$ -a.e. The state is

$$X_t^u = x_0 E_\alpha(-\lambda t^\alpha) + \int_0^t R(t-s) u_s ds + \sigma_0 \int_0^t R(t-s) dW_s.$$

At $H = 1/2$, $R(t) = e^{-\lambda t}$ and $Y_t = -a e^{-\lambda \tau} - (c/\lambda)(1 - e^{-\lambda \tau})$.

Proof. Here $\mathcal{H}_x = -\lambda Y - c$ involves neither Z , X nor u , and $g_x \equiv a$, so $\zeta^g = 0$ and the map Φ of the proof of Theorem 2 sends deterministic processes to deterministic ones with $Z[P] = 0$. Its unique fixed point is therefore deterministic and the same for every control, and $y(\tau) := Y_{T-\tau}$ solves $y = -ak - cK - \lambda k * y$ on $(0, T]$, with $K(\tau) := \int_0^\tau k = \tau^\alpha / \Gamma(\alpha+1)$. The Laplace transforms of k , K and $\tau^{\beta-1} E_{\alpha, \beta}(-\lambda \tau^\alpha)$ are $s^{-\alpha}$, $s^{-\alpha-1}$ and $s^{\alpha-\beta} / (s^\alpha + \lambda)$, so the equation reads $\hat{y}(1 + \lambda s^{-\alpha}) = -as^{-\alpha} - cs^{-\alpha-1}$, that is $\hat{y} = -a/(s^\alpha + \lambda) - cs^{-1}/(s^\alpha + \lambda)$, whose inverse transform is the stated Y by the series identity $\tau^\alpha E_{\alpha, \alpha+1}(-\lambda \tau^\alpha) = (1 - E_\alpha(-\lambda \tau^\alpha))/\lambda$. The Hamiltonian is strictly concave in u with maximiser $\Pi_U(Y_t/r)$ over U , so u^* satisfies the condition of Theorem 3 and is optimal by Theorem 4; every optimal control satisfies the condition of Theorem 3 with the same adjoint, and for a strictly concave Hamiltonian that condition has the single solution u_t^* . The state formula is the variation-of-constants formula for $X + \lambda k * X = x_0 + k * u + \sigma_0 k * dW$: the resolvent $R = k - \lambda k * R$ has Laplace transform $1/(s^\alpha + \lambda)$, the stochastic Fubini theorem applies pathwise, and $1 - \lambda \int_0^t R = E_\alpha(-\lambda t^\alpha)$. At $\alpha = 1$ the Mittag-Leffler functions are exponentials. $\square$



As $\tau \rightarrow 0$, $Y_t \approx -a\tau^{H-1/2}/\Gamma(\alpha)$: for $a > 0$ the weight $k(T-t)$ of Remark 2 drives the adjoint to $-\infty$ at maturity and the control bound binds there. At $H = 1/2$ the adjoint is a convex combination of $-a$ and $-c/\lambda$ and stays bounded.

6.2 Numerical check

We take $T = 1$, $H = 0.1$ ($\alpha = 0.6$), $\lambda = 1$, $c = 1$, $a = 0.5$, $r = 1$, $M = 1.5$, $x_0 = 1$ and $\sigma_0 = 0.3$. The script `python/volterra_lq_check.py` computes every number below; the Mittag-Leffler functions are summed from their power series, which converge fast here because $\lambda\tau^\alpha \leq 1$. Proposition 2 gives $Y_0 = -0.6722$, and Y_t crosses $-rM$ at $T-t = 0.0193$, so $u^* = -M$ on the last 1.93 % of the horizon; at $H = 1/2$, $|Y_t| \leq 1$ and the bound never binds.

The adjoint, by the Picard iteration of Theorem 2. We discretise the backward equation of Definition 6 on N cells of length $h = T/N$, with Y constant on each cell and every kernel integral exact (the double integral of k over two cells is a second difference of $\tau^{\alpha+1}/\Gamma(\alpha+2)$), and solve it by the Picard iteration of the proof of Theorem 2 (Algorithm 1, Appendix B), which never uses the closed form. The iteration reaches rounding level in 27 steps for every N in Table 1 and agrees with a direct triangular solve to 2×10^{-14} . Against the exact cell averages of the closed form of Proposition 2, the L^1 error falls from 1.6×10^{-4} at $N = 250$ to 5.9×10^{-6} at $N = 4000$ with observed order 1.20.

Table 1: The adjoint of Proposition 2 against the discretised backward equation solved by the Picard iteration of Theorem 2. Errors are in the cell averages of Y ; the observed order is that of the L^1 error between successive N .

N	Picard steps	L^1 error of Y	max error, $t \leq T - 0.1$	observed order
250	27	1.60×10^{-4}	1.02×10^{-4}	—
500	27	7.04×10^{-5}	3.34×10^{-5}	1.184
1000	27	3.08×10^{-5}	1.10×10^{-5}	1.190
2000	27	1.35×10^{-5}	3.61×10^{-6}	1.193
4000	27	5.89×10^{-6}	1.19×10^{-6}	1.195

The state, by the Picard iteration of Theorem 1. The mean $m_t = \mathbb{E}X_t$ under u^* solves the deterministic Volterra equation $m + \lambda k * m = x_0 + k * u^*$, since the noise term has mean zero. Discretised in the same way and solved by the Picard iteration of the proof of Theorem 1 (28 steps), it agrees with the closed form $x_0 E_\alpha(-\lambda t^\alpha) + R * u^*$ of Proposition 2 to 1.6×10^{-4} in the cell averages at $N = 250$ and to 6.0×10^{-6} at $N = 4000$, with $\mathbb{E}X_T = -0.1127$.

The roughness of the state. By Proposition 2 the noise part $\tilde{X} := \sigma_0 R * dW$ is Gaussian, with increment variance $D(\delta) := \mathbb{E}|\tilde{X}_T - \tilde{X}_{T-\delta}|^2 = \sigma_0^2 [\int_0^\delta R^2 + \int_0^{T-\delta} (R(r+\delta) - R(r))^2 dr]$, which we compute by adaptive quadrature with algebraic end-point weights; $\text{Var } X_T =$



0.1241. Theorem 1 and Lemma 4 (iii) predict $D(\delta) \approx \sigma_0^2 c_H \delta^{2H}$ for small δ , with $c_H := \Gamma(\alpha)^{-2} [1/(2H) + \int_0^\infty ((1+x)^{\alpha-1} - x^{\alpha-1})^2 dx] = 3.5245$ computed from the kernel alone. Table 2 shows the observed order tending to $2H = 0.2$ and the ratio to this law tending to 1.

Table 2: Increment variance of the noise part of the state at maturity, by quadrature of the resolvent R , against the small-lag law from the kernel. The observed order is that of D between successive δ ; Theorem 1 predicts $2H$.

δ	$D(\delta)$	observed order	$D(\delta)/(\sigma_0^2 c_H \delta^{2H})$
10^{-1}	0.18287	—	0.91370
10^{-2}	0.12334	0.1710	0.97673
10^{-3}	0.07920	0.1924	0.99406
10^{-4}	0.05020	0.1981	0.99850
10^{-5}	0.03171	0.1995	0.99962

The gradient, by forward sensitivities. For the discretised cost J_N of piecewise-constant controls we compute $\partial J_N / \partial u_j$ by solving the forward equation once for each unit perturbation of the control, as in Proposition 1, with no adjoint. By Lemma 3 the quantity $-(h^{-1} \partial J_N / \partial u_j - ru_j)$ is the cell average of Y : it agrees with the discrete backward solution above to 3×10^{-12} for $N \leq 2000$, and with the closed form to 1.3×10^{-5} in L^1 at $N = 2000$.

The optimal control, by direct minimisation. On $N = 200$ cells we minimise J_N over $[-M, M]^{200}$ with L-BFGS-B and finite-difference gradients, from three starting points: zero, the upper bound and a uniform random control. The three runs end within 1×10^{-12} of each other at $J_N = 0.48421779$. Their controls differ from u^* of Proposition 2 by at most 1.1×10^{-3} , the discretisation error of the adjoint at this N , and from the control of the discrete adjoint by at most 6.4×10^{-6} ; on this grid u^* costs 9.3×10^{-9} more than the direct minimiser. The maximum condition of Theorem 3 thus picks out the minimiser, which Theorem 4 says is global, and Figure 1 shows both controls.

The terminal weight. Replacing the weight $k(T-t)$ on g_x by 1 — the adjoint BSDE carried over unchanged, $Y_t = -aE_\alpha(-\lambda\tau^\alpha) - (c/\lambda)(1 - E_\alpha(-\lambda\tau^\alpha))$ — gives a control that differs from u^* by up to 1.00 and costs $J_N = 0.5270$ against $J_N(u^*) = 0.4842$ at $N = 2000$: it gives up 14.7 % of the improvement of u^* over the null control ($J_N = 0.7755$). At $H = 1/2$ the two adjoints coincide, and the series reproduces the exponential form of Proposition 2 to 3×10^{-16} .

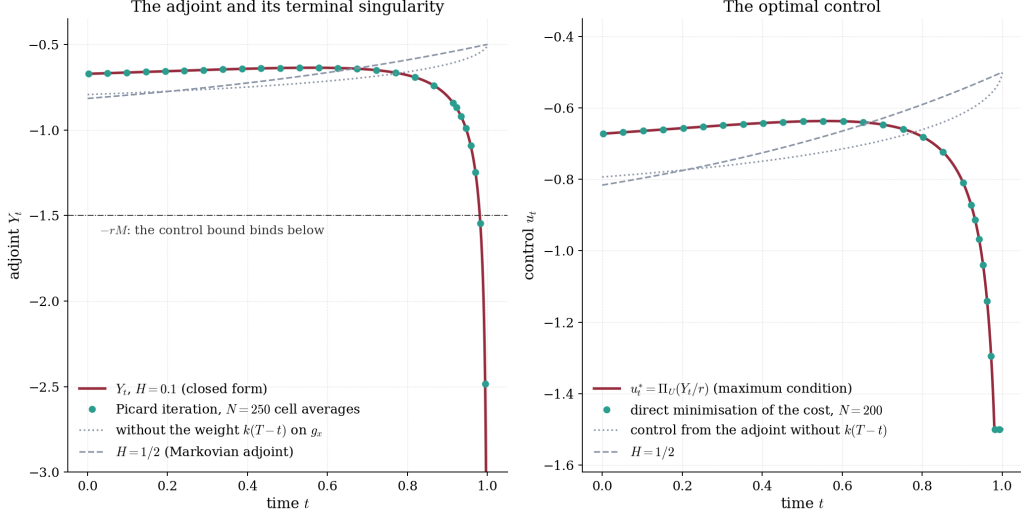


Figure 1: The linear–quadratic example of §6, $H = 0.1$. Left: the adjoint Y_t of Proposition 2 (curve) and the cell averages of the backward equation solved by the Picard iteration of Theorem 2 with $N = 250$ (points); dotted, the adjoint without the terminal weight $k(T - t)$; dashed, the Markovian adjoint ($H = 1/2$). Y_t tends to $-\infty$ at maturity and crosses $-rM$ at $T - t = 0.0193$. Right: the control $u^* = \Pi_U(Y/r)$ of the maximum condition (curve) and the minimiser of the discretised cost found by L-BFGS-B on 200 cells without the adjoint (points), with the same two comparisons.

7. Conclusion

We proved a stochastic maximum principle for controlled Volterra equations with the fractional kernel: forward well-posedness with Hölder paths of every order below H (Theorem 1), a control-stability estimate that is Hölder rather than Lipschitz for $H < 1/2$ (Lemma 1), existence and uniqueness of the adjoint BSVIE (Theorem 2), the duality identity (Lemma 3), and the necessary and sufficient conditions (Theorems 3 and 4). The adjoint differs from the Markovian one in two places — the kernel read over later times and the singular weight $k(T - t)$ on the terminal gradient — and the example of §6 shows the second one changing the optimal control near maturity.

7.1 Scope

The results assume Lipschitz coefficients and a convex compact control set. Square-root models such as the rough Heston model of El Euch–Rosenbaum [9] fall outside Assumption 1 and need a different existence theory for the forward and the adjoint equations. Without concavity of the Hamiltonian in the control, the global maximum condition needs spike variations, not treated here. Beyond linear–quadratic examples the adjoint has no closed form, and a numerical BSVIE solver, with regression estimates of the conditional expectations in Definition 6, is the natural next step.



A. Fractional Kernel Estimates

This appendix collects the kernel estimates used in §§3–4, for the fractional kernel $k(r) = r^{\alpha-1}/\Gamma(\alpha)$, $\alpha = H + 1/2$, $H \in (0, 1/2]$.

A.1 Integrability and increments

Lemma 4 (Kernel estimates). For $T > 0$ and $\delta > 0$: (i) $\int_0^T k(r) dr = T^\alpha/\Gamma(\alpha + 1)$ and $\int_0^T k(r)^2 dr = T^{2H}/(2H\Gamma(\alpha)^2)$; (ii) $(k * k)(r) = \int_0^r k(r-s)k(s) ds = r^{2H}/\Gamma(2H + 1)$; (iii) $\int_0^\infty (k(r+\delta) - k(r))^2 dr = \delta^{2H} \Gamma(\alpha)^{-2} \int_0^\infty ((1+x)^{\alpha-1} - x^{\alpha-1})^2 dx$, and the last integral is finite.

Proof. is elementary. For (ii), the substitution $s = rv$ gives $r^{2\alpha-1}B(\alpha, \alpha)/\Gamma(\alpha)^2$ with the Beta function $B(\alpha, \alpha) = \Gamma(\alpha)^2/\Gamma(2\alpha)$, and $2\alpha - 1 = 2H$. For (iii), substitute $r = \delta x$; near $x = 0$ the integrand is $O(x^{2H-1})$, which is integrable, and at infinity it is $O(x^{2H-3})$ by the mean-value theorem. These estimates are standard for fractional processes (Decreusefond–Üstünel [7]). $\square$

The square-integrability of k , from $2H - 1 > -1$, gives the stochastic convolution in Definition 1 a finite second moment; (iii) gives its increments the order $|t - s|^{2H}$ used in Theorem 1.

A.2 Hölder regularity of convolutions

Lemma 5 (Hölder regularity). Let $\varphi \in L^\infty([0, T])$ and $h(t) = \int_0^t k(t-s)\varphi(s) ds$. Then $|h(t) - h(s)| \leq 2\|\varphi\|_\infty |t - s|^\alpha/\Gamma(\alpha + 1)$ for $0 \leq s \leq t \leq T$.

Proof. Write

$$h(t) - h(s) = \int_s^t k(t-r)\varphi(r) dr + \int_0^s [k(t-r) - k(s-r)]\varphi(r) dr.$$

The first integral is at most $\|\varphi\|_\infty K(t-s)$ with $K(\tau) = \tau^\alpha/\Gamma(\alpha + 1)$. Since k is non-increasing, the second is at most $\|\varphi\|_\infty \int_0^s [k(s-r) - k(t-r)] dr = \|\varphi\|_\infty [K(s) - K(t) + K(t-s)] \leq \|\varphi\|_\infty K(t-s)$. $\square$

Theorem 1 uses Lemma 5 in L^p , through Minkowski's inequality, for the drift part of the increments, which are of order $\alpha = H + 1/2$; the stochastic part is of the lower order H by Lemma 4 (iii).



A.3 Change of the order of integration

Lemma 6 (Volterra integration-by-parts identity). For $k \in L^1(0, T)$ and bounded measurable u, v on $[0, T]$, and more generally whenever the double integral converges absolutely,

$$\int_0^T u(t) \int_0^t k(t-s) v(s) ds dt = \int_0^T v(s) \int_s^T k(t-s) u(t) dt ds.$$

Proof. Fubini's theorem on $\{(s, t) : 0 \leq s \leq t \leq T\}$; the double integral of $|k(t-s) u(t) v(s)|$ is at most $\|u\|_\infty \|v\|_\infty T \|k\|_{L^1}$. $\square$

In Lemma 3 the identity is applied to expectations, under the absolute-convergence bounds of Theorem 2. It is why the adjoint of Definition 6 reads the forward kernel $k(t-s)$ over the later variable: the function is not reversed, only the direction of integration.

B. The Discretisation of Section 6

This appendix states the scheme behind the numerical check of §6, as implemented in `python/volterra_lq_check.py`.

B.1 Cell-average Picard scheme

Algorithm 1 — Cell-average Picard scheme

Solves a linear Volterra equation $v = \varphi - \lambda k * v$ on $(0, T]$ for the cell averages of v , by the Picard iteration of the proofs of Theorems 1 and 2. In §6 it solves the backward equation of Definition 6 in the variable $\tau = T - t$, with $\varphi = -ak - cK$, and the forward equation for the mean of the state, with $\varphi = x_0 + k * u$.

```

1 Input:  kernel order al = H + 1/2, horizon T, number of cells N, rate lam,
2         cell averages phi[0..N-1] of the free term, tolerance eps.
3 Output: cell averages v[0..N-1] of the solution of v = phi - lam (k * v).
4
5 1. h = T / N;  G(x) = max(x, 0)^(al+1) / Gamma(al+2)
6 2. For m = 0, 1, ..., N-1:
7     Om[m] = G((m+1) h) - 2 G(m h) + G((m-1) h)
8     # the exact integral of k(t - s) 1{s < t} over two cells m apart
9 3. v = phi
10 4. Repeat:
11     a. For i = 0, 1, ..., N-1:
12         w[i] = phi[i] - (lam / h) * sum_{j=0..i} Om[i-j] * v[j]
13     b. d = max_i |w[i] - v[i]|;  v = w
14     until d <= eps * max(1, max_i |v[i]|)
15 5. return v

```

The free terms are integrated exactly as well: the cell averages of k and K are differences of K and of $\tau^{\alpha+1}/\Gamma(\alpha+2)$, and that of $k*u$ for a piecewise-constant control is $h^{-1} \sum_j \Omega_{i-j} u_j$. Each sweep of step 4 is one discrete convolution, of cost $O(N^2)$ as implemented. The



number of sweeps does not grow with N (Table 1, with $\varepsilon = 10^{-14}$). In continuous time the n -th power of the operator $v \mapsto \lambda k * v$ has norm at most $(\lambda T^\alpha)^n / \Gamma(n\alpha + 1)$ on bounded functions, which is why the weighted-norm contraction of Theorem 2 exists. The closed forms of Proposition 2 are evaluated from the Mittag-Leffler series and their antiderivatives, $\int_0^\tau s^{\beta-1} E_{\alpha,\beta}(-\lambda s^\alpha) ds = \tau^\beta E_{\alpha,\beta+1}(-\lambda \tau^\alpha)$, so the comparison in Table 1 is between exact cell averages.



References

- [1] Yong, J., & Zhou, X. Y. (1999). *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Springer.
- [2] Peng, S. (1990). A general stochastic maximum principle for optimal control problems. *SIAM Journal on Control and Optimization*, 28(4), 966–979.
<https://doi.org/10.1137/0328054>
- [3] Gatheral, J., Jaisson, T., & Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, 18(6), 933–949. <https://doi.org/10.1080/14697688.2017.1393551>
- [4] Bayer, C., Friz, P., & Gatheral, J. (2016). Pricing under rough volatility. *Quantitative Finance*, 16(6), 887–904. <https://doi.org/10.1080/14697688.2015.1099717>
- [5] Yong, J. (2006). Backward stochastic Volterra integral equations and some related problems. *Stochastic Processes and their Applications*, 116(5), 779–795.
<https://doi.org/10.1016/j.spa.2006.01.005>
- [6] Wang, T. (2018). Linear quadratic control problems of stochastic Volterra integral equations. *ESAIM: Control, Optimisation and Calculus of Variations*, 24, 1849–1879.
<https://doi.org/10.1051/cocv/2017002>
- [7] Decreusefond, L., & Üstünel, A. S. (1999). Stochastic analysis of the fractional Brownian motion. *Potential Analysis*, 10, 177–214.
<https://doi.org/10.1023/a:1008634027843>
- [8] Pardoux, E., & Peng, S. (1990). Adapted solution of a backward stochastic differential equation. *Systems & Control Letters*, 14, 55–61.
[https://doi.org/10.1016/0167-6911\(90\)90082-6](https://doi.org/10.1016/0167-6911(90)90082-6)
- [9] El Euch, O., & Rosenbaum, M. (2019). The characteristic function of rough Heston models. *Mathematical Finance*, 29(1), 3–38. <https://doi.org/10.1111/mafi.12173>



Zamrik Quant Research

The research arm of Zamrik Quant Lab

About the Desk

Optimal Control & Decision Desk

Zamrik Quant Research

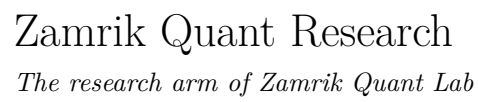
What to do under uncertainty: optimal control and HJB, mean-field games, agent-based models, irreversible decisions, betting, lotteries and insurance.

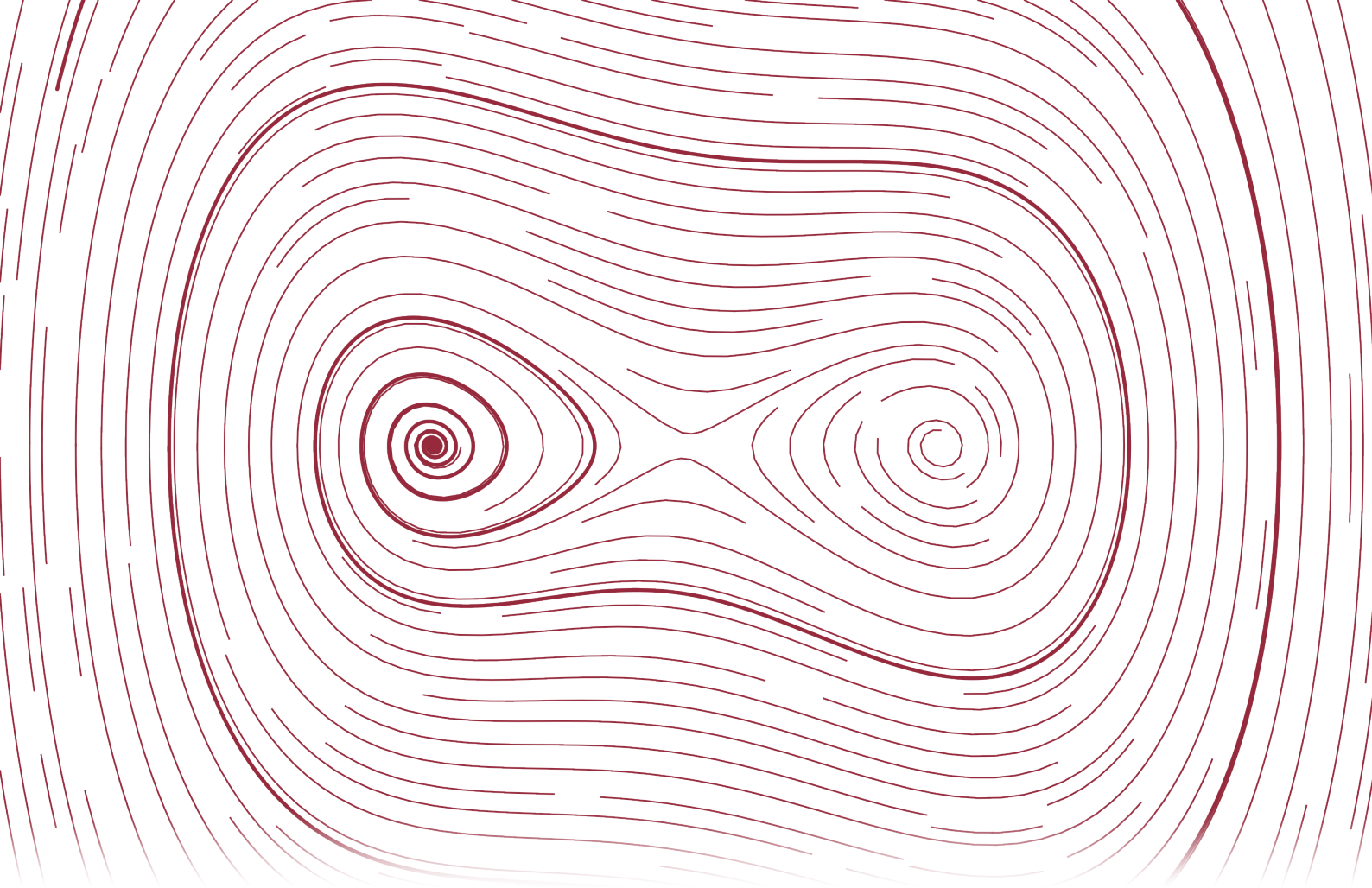
Most of this desk's questions have the same shape: someone must act now, without knowing what comes next, and live with the result. The Optimal Control & Decision desk turns such questions into Hamilton–Jacobi–Bellman equations, optimal stopping and singular control problems, and solves them.

When several parties optimise at once, the problem becomes a game: Dynkin games and wars of attrition between two or three players, and mean field games, in which a large population and its representative member shape each other through coupled Hamilton–Jacobi–Bellman and Fokker–Planck equations. Where the equations give out, agent-based models take over, and the desk says plainly which tool fits which case.

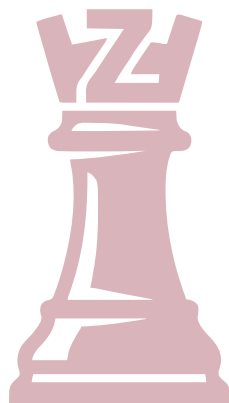
The applications are concrete: when to build an irreversible plant, how a company should pay dividends, how to run a home battery against a time-of-use tariff, when a wind turbine should turn into the wind, how a central bank should answer stagflation, and how much to bet when an edge is small. Its notebooks and essays carry the same reasoning to retirement, market timing, lotteries, insurance and prize competitions, each worked through to a decision a reader could actually make.

The desk is supervised by **Dr. T. Zamrik**, Head of Zamrik Quant Research, PhD, Imperial College London (2011).





Phase portrait of the damped Duffing oscillator: $\dot{x} = y$, $\dot{y} = -\delta y - \alpha x - \beta x^3$, $\delta = 0.18$, $\alpha = -0.60$, $\beta = 0.76$



ZAMRIK QUANT RESEARCH

The research arm of Zamrik Quant Lab



WP-2026-48040577
zamrik.com