
Prerequisites for Rough Volatility and Path Signatures in Trading

A self-contained primer covering fractional calculus, Volterra equations, stochastic volatility, and tensor algebra for path signatures.

Melvin Alvarez-Caradu
Independent Researcher, HFThot Research Lab
<https://hfthot-lab.eu>
June 2026 — v1.0

DISTRIBUTION: REGISTERED USERS — FREE TIER

This document provides a self-contained mathematical primer for readers approaching the HFThot Research Lab white paper “Rough Heston Volatility Arbitrage with Path Signatures” (WP2). We develop, from first principles, the essential mathematical machinery: fractional Brownian motion and Hölder regularity, Volterra-type integral equations, the classical and rough Heston stochastic volatility models, Fourier pricing via the Lewis formula, signature theory for continuous paths, and tensor algebra. Each section contains full derivations, worked examples, and exercises. The treatment is rigorous but pedagogical, pitched at the level of a first-year graduate course in mathematical finance.

Prerequisites

Required background for this primer: Measure-theoretic probability (filtrations, martingales, conditional expectation), Itô stochastic calculus (as covered in our WP1 prerequisites), linear algebra, and basic complex analysis (contour integrals, residues). No prior knowledge of rough paths or signature theory is assumed.

Contents

1	Integration Foundations	4
1.1	The Riemann Integral: Review	4
1.2	From Riemann to Lebesgue: Why Generalise?	4
1.3	Key Convergence Theorems	4
1.4	The Itô Integral: Stochastic Integration	5
1.5	Exercises	5
2	Fractional Brownian Motion	7
2.1	Motivation: Why “Rough” Volatility?	7
2.2	Definition and Basic Properties	7
2.3	Self-Similarity and Stationarity of Increments	7
2.4	Long-Range Dependence vs. Antipersistence	8
2.5	Representations of fBm	8
2.5.1	Mandelbrot–Van Ness Representation	8
2.5.2	Molchan–Golosov Representation	8
2.6	Hurst Exponent Estimation	9
2.6.1	Variogram Method	9
2.6.2	R/S (Rescaled Range) Analysis	9
2.7	Exercises	9
3	Volterra Integral Equations	11
3.1	Classical Volterra Equations	11
3.2	Singular Kernels and Fractional Volterra Equations	11
3.3	The Fractional Riccati Equation	11
3.4	Adams Discretisation Scheme	12
3.5	Exercises	12

4	Classical Stochastic Volatility Models	13
4.1	The Black–Scholes Framework and Its Limitations	13
4.2	The Heston Model	13
4.3	Characteristic Function of the Heston Model	13
4.4	From Heston to Rough Heston	14
4.5	Exercises	14
5	Fourier Pricing: The Lewis Formula	16
5.1	The Pricing Problem	16
5.2	Gil-Pelaez Inversion	16
5.3	The Lewis (2001) Formula	16
5.4	Numerical Implementation	16
5.5	Exercises	17
6	The Rough Heston Model	18
6.1	Model Specification	18
6.2	Existence and Positivity	18
6.3	Characteristic Function	18
6.4	Calibration via Levenberg–Marquardt	18
6.5	Exercises	19
7	Implied Volatility and the Volatility Surface	20
7.1	Definition and Computation	20
7.2	The Volatility Smile and Skew	20
7.3	Short-Time Asymptotics and Roughness	20
7.4	The Variance Risk Premium	20
7.5	Exercises	21
8	Path Signatures	22
8.1	Motivation: From Time Series to Features	22
8.2	Definition of the Path Signature	22
8.3	Dimension Counting	22
8.4	Key Algebraic Properties	23
	8.4.1 Shuffle Product	23
	8.4.2 Chen’s Identity	23
	8.4.3 Uniqueness (Hambly–Lyons)	23
8.5	The Log-Signature	23
8.6	Signatures as Universal Nonlinear Features	24
8.7	Exercises	24
9	Tensor Algebra	26
9.1	The Tensor Algebra $T((\mathbb{R}^d))$	26
9.2	The Truncated Tensor Algebra	26
9.3	The Free Lie Algebra and Lyndon Words	26
9.4	Computational Aspects	26
9.5	Exercises	27

10 Volatility Trading Strategies	28
10.1 Variance Swaps and Realised Variance	28
10.2 Delta-Hedged Option Positions	28
10.3 Signature-Based Volatility Arbitrage	28
10.4 Exercises	28
11 Calibration and Numerical Methods	30
11.1 The Calibration Problem	30
11.2 Levenberg–Marquardt Algorithm	30
11.3 Gauss–Laguerre Quadrature	30
11.4 Differential Evolution for Global Optimisation	30
11.5 Exercises	31
12 Summary and Connections to WP2	33
References	34

1 Integration Foundations

This section reviews the key concepts of integration theory needed for stochastic calculus. We progress from the familiar Riemann integral to the Lebesgue integral and then to the Itô integral, explaining at each step *why* the generalisation is needed.

1.1 The Riemann Integral: Review

Definition 1.1 (Riemann Integral). For a bounded function $f : [a, b] \rightarrow \mathbb{R}$, the Riemann integral is:

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{k=1}^n f(x_k^*)(x_k - x_{k-1}), \quad (1)$$

where $a = x_0 < x_1 < \dots < x_n = b$ is a partition and $x_k^* \in [x_{k-1}, x_k]$ is any sample point, provided the limit exists and is independent of the choice of partition and sample points.

Example 1.1 (Riemann Integral — Worked). Compute $\int_0^1 x^2 dx$ using Riemann sums. **Solution.** With uniform partition $x_k = k/n$ and right endpoints:

$$\sum_{k=1}^n \left(\frac{k}{n}\right)^2 \cdot \frac{1}{n} = \frac{1}{n^3} \sum_{k=1}^n k^2 = \frac{1}{n^3} \cdot \frac{n(n+1)(2n+1)}{6} \xrightarrow{n \rightarrow \infty} \frac{1}{3}. \quad (2)$$

1.2 From Riemann to Lebesgue: Why Generalise?

The Riemann integral has limitations that matter in probability and stochastic calculus:

- It cannot integrate highly irregular functions (e.g., $\mathbb{1}_{\mathbb{Q}}$, the indicator of rationals);
- Convergence theorems (dominated convergence, monotone convergence) fail;
- It is not compatible with general probability measures.

Key Insight

The Lebesgue integral “slices horizontally” (partitioning the range of f) rather than “slicing vertically” (partitioning the domain). This allows it to integrate a much wider class of functions and is the natural framework for probability theory, where we compute $\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P}$.

Definition 1.2 (Lebesgue Integral (Intuition)). For a non-negative measurable function f :

$$\int f d\mu = \sup \left\{ \sum_{k=1}^n y_k \cdot \mu(\{f \geq y_k\}) : \text{simple functions} \leq f \right\}. \quad (3)$$

For a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$: $\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$ is a Lebesgue integral.

1.3 Key Convergence Theorems

Theorem 1.1 (Monotone Convergence Theorem). If $0 \leq f_1 \leq f_2 \leq \dots$ and $f_n \uparrow f$ pointwise, then $\int f_n d\mu \uparrow \int f d\mu$.

Theorem 1.2 (Dominated Convergence Theorem (DCT)). If $f_n \rightarrow f$ pointwise and $|f_n| \leq g$ with $\int g d\mu < \infty$, then $\int f_n d\mu \rightarrow \int f d\mu$ and $\int |f_n - f| d\mu \rightarrow 0$.

Practical Application

The DCT is used constantly in mathematical finance: when we interchange limits and expectations (e.g., differentiating option prices with respect to parameters), we are implicitly invoking DCT. The “dominating function” g must be integrable, which is why we need integrability conditions on payoffs and price processes.

1.4 The Itô Integral: Stochastic Integration

Standard integration ($\int f(s) ds$) handles smooth integrands against smooth integrators. The *Itô integral* generalises this to allow integration against Brownian motion:

Definition 1.3 (Itô Integral (Sketch)). For an *adapted* process H_t with $\mathbb{E}[\int_0^T H_t^2 dt] < \infty$:

$$\int_0^T H_t dW_t = \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} H_{t_k} (W_{t_{k+1}} - W_{t_k}). \quad (4)$$

The crucial point: H is evaluated at the *left* endpoint t_k , not at any other point. This choice makes the Itô integral a martingale.

Proposition 1.1 (Itô Isometry).

$$\mathbb{E}\left[\left(\int_0^T H_t dW_t\right)^2\right] = \mathbb{E}\left[\int_0^T H_t^2 dt\right]. \quad (5)$$

In particular, $\mathbb{E}[\int_0^T H_t dW_t] = 0$ (the Itô integral has zero mean).

Example 1.2 (Computing $\int_0^T W_t dW_t$ — Worked). Using the limit definition: $\sum_k W_{t_k} (W_{t_{k+1}} - W_{t_k})$. Note $W_{t_k} (W_{t_{k+1}} - W_{t_k}) = \frac{1}{2}(W_{t_{k+1}}^2 - W_{t_k}^2) - \frac{1}{2}(W_{t_{k+1}} - W_{t_k})^2$. Summing (telescoping): $\sum_k = \frac{1}{2}W_T^2 - \frac{1}{2}\sum_k (W_{t_{k+1}} - W_{t_k})^2$. The second sum converges to T (quadratic variation of BM). Therefore:

$$\int_0^T W_t dW_t = \frac{1}{2}W_T^2 - \frac{1}{2}T. \quad (6)$$

Compare with ordinary calculus: $\int_0^T x dx = \frac{1}{2}T^2$. The extra $-\frac{1}{2}T$ is the Itô correction!

1.5 Exercises

1. **Riemann sums.** Compute $\int_0^2 (3x^2 + 1) dx$ using Riemann sums with $n = 4$ equal subintervals and right endpoints, then verify with the antiderivative.
2. **DCT application.** Let $f_n(x) = x^n$ on $[0, 1]$. Show $f_n \rightarrow 0$ pointwise, find a dominating function, and use DCT to conclude $\int_0^1 x^n dx \rightarrow 0$.
3. **Itô isometry.** Compute $\mathbb{E}[(\int_0^T t dW_t)^2]$ using Itô isometry.
4. **Itô integral.** Using the result $\int_0^T W_t dW_t = \frac{1}{2}W_T^2 - \frac{1}{2}T$, compute $\mathbb{E}[\int_0^T W_t dW_t]$ and $\text{Var}(\int_0^T W_t dW_t)$.

Solutions

Solution to Exercise 1. Partition: $x_k = k/2$ for $k = 0, 1, 2, 3, 4$, so $\Delta x = 0.5$. Right Riemann sum: $\sum_{k=1}^4 (3(k/2)^2 + 1) \cdot 0.5 = 0.5[(3 \cdot 0.25 + 1) + (3 \cdot 1 + 1) + (3 \cdot 2.25 + 1) + (3 \cdot 4 + 1)] = 0.5[1.75 + 4 + 7.75 + 13] = 0.5 \times 26.5 = 13.25$. Exact: $\int_0^2 (3x^2 + 1) dx = [x^3 + x]_0^2 = 8 + 2 = 10$. The approximation overshoots due to the convexity of f . With $n \rightarrow \infty$, the sum converges to 10. $\square$

Solution to Exercise 3. By Itô isometry: $\mathbb{E}[(\int_0^T t dW_t)^2] = \mathbb{E}[\int_0^T t^2 dt] = \int_0^T t^2 dt = T^3/3$. $\square$

Solution to Exercise 4. $\mathbb{E}[\int_0^T W_t dW_t] = \mathbb{E}[\frac{1}{2}W_T^2 - \frac{1}{2}T] = \frac{1}{2}T - \frac{1}{2}T = 0$. $\checkmark$ $\text{Var}(\int_0^T W_t dW_t) = \text{Var}(\frac{1}{2}W_T^2 - \frac{1}{2}T) = \frac{1}{4} \text{Var}(W_T^2)$. Since $W_T \sim \mathcal{N}(0, T)$, $W_T^2/T \sim \chi_1^2$, so $\text{Var}(W_T^2) = 2T^2$. Therefore $\text{Var} = \frac{1}{4} \cdot 2T^2 = T^2/2$. Cross-check with Itô isometry: $\mathbb{E}[\int_0^T W_t^2 dt] = \int_0^T \mathbb{E}[W_t^2] dt = \int_0^T t dt = T^2/2$. $\checkmark$ $\square$

2 Fractional Brownian Motion

2.1 Motivation: Why “Rough” Volatility?

Classical stochastic volatility models (Heston, SABR) assume that the instantaneous variance V_t follows a diffusion driven by standard Brownian motion. The resulting variance paths have Hölder regularity $\frac{1}{2} - \varepsilon$ for any $\varepsilon > 0$. Empirical analysis of high-frequency realised variance, however, shows that log-volatility paths are much rougher, with Hölder exponent $H \approx 0.1$.

Key Insight

The roughness of volatility paths is a universal empirical fact across asset classes. It implies that the “correct” driving noise for variance processes is *fractional Brownian motion* with Hurst parameter $H < \frac{1}{2}$, not standard Brownian motion ($H = \frac{1}{2}$).

2.2 Definition and Basic Properties

Definition 2.1 (Fractional Brownian Motion). A *fractional Brownian motion* (fBm) with Hurst parameter $H \in (0, 1)$ is a centered Gaussian process $\{B_t^H\}_{t \geq 0}$ with $B_0^H = 0$ and covariance function:

$$\mathbb{E}[B_s^H B_t^H] = \frac{1}{2}(|s|^{2H} + |t|^{2H} - |t - s|^{2H}), \quad s, t \geq 0. \quad (7)$$

For $H = \frac{1}{2}$, we recover $\mathbb{E}[B_s^{1/2} B_t^{1/2}] = \min(s, t)$, which is standard Brownian motion.

Proposition 2.1 (Variance of Increments). For any $s < t$:

$$\mathbb{E}[(B_t^H - B_s^H)^2] = |t - s|^{2H}. \quad (8)$$

Proof. Expanding the square and using the covariance (7):

$$\begin{aligned} \mathbb{E}[(B_t^H - B_s^H)^2] &= \mathbb{E}[(B_t^H)^2] - 2\mathbb{E}[B_s^H B_t^H] + \mathbb{E}[(B_s^H)^2] \\ &= t^{2H} - (s^{2H} + t^{2H} - |t - s|^{2H}) + s^{2H} \\ &= |t - s|^{2H}. \end{aligned} \quad \square$$

2.3 Self-Similarity and Stationarity of Increments

Proposition 2.2 (Self-Similarity). Fractional Brownian motion is H -self-similar: for any $a > 0$,

$$\{B_{at}^H\}_{t \geq 0} \stackrel{d}{=} \{a^H B_t^H\}_{t \geq 0}. \quad (9)$$

Proposition 2.3 (Stationary Increments). The increments $B_t^H - B_s^H$ depend only on $t - s$ in distribution. This follows immediately from the covariance structure.

Theorem 2.1 (Hölder Regularity). Sample paths of fBm are almost surely Hölder continuous of any order $\gamma < H$, and are not Hölder continuous of order $\gamma > H$.

Key Insight

For $H = 0.1$ (typical empirical value for log-variance), the paths are extremely rough — much rougher than Brownian motion ($H = 0.5$). This roughness is *not a modelling artifact* but a fundamental empirical observation that rough volatility models correctly capture.

2.4 Long-Range Dependence vs. Antipersistence

The behavior of fBm depends critically on the Hurst parameter:

- $H > \frac{1}{2}$: increments are *positively correlated* (long-range dependent, persistent);
- $H = \frac{1}{2}$: increments are *independent* (standard Brownian motion);
- $H < \frac{1}{2}$: increments are *negatively correlated* (antipersistent, mean-reverting at short scales).

For rough volatility, we are in the regime $H < \frac{1}{2}$. The autocovariance of disjoint increments on a regular grid with step δ satisfies:

$$\rho(n) = \frac{1}{2}((n+1)^{2H} - 2n^{2H} + (n-1)^{2H}) \sim H(2H-1)n^{2H-2} \quad \text{as } n \rightarrow \infty. \quad (10)$$

For $H < \frac{1}{2}$, we have $\rho(n) < 0$ and $\sum_n |\rho(n)| < \infty$.

2.5 Representations of fBm

2.5.1 Mandelbrot–Van Ness Representation

Theorem 2.2 (Mandelbrot–Van Ness, 1968). *Fractional Brownian motion admits the integral representation:*

$$B_t^H = \frac{1}{\Gamma(H + \frac{1}{2})} \int_{-\infty}^t [(t-s)^{H-1/2} - (-s)_+^{H-1/2}] dW_s, \quad (11)$$

where W is a standard two-sided Brownian motion.

2.5.2 Molchan–Golosov Representation

For $t \geq 0$ only, using a *Volterra kernel*:

$$B_t^H = \int_0^t K_H(t, s) dW_s, \quad (12)$$

where the kernel K_H is:

$$K_H(t, s) = c_H(t-s)^{H-1/2} + c_H\left(H - \frac{1}{2}\right) \int_s^t (u-s)^{H-3/2} \left(1 - \left(\frac{s}{u}\right)^{1/2-H}\right) du \quad (13)$$

with $c_H = \sqrt{\frac{2H\Gamma(3/2-H)}{\Gamma(H+1/2)\Gamma(2-2H)}}$.

Key Insight

The Volterra representation (12) is the foundation of the *rough Heston model*. The variance process is driven by a Brownian motion through a singular kernel $(t-s)^{H-1/2}$, which creates the rough path behavior.

2.6 Hurst Exponent Estimation

2.6.1 Variogram Method

For a time series $(X_{t_1}, \dots, X_{t_n})$ sampled at frequency δ , the *variogram* at lag k is:

$$V(k) = \frac{1}{n-k} \sum_{i=1}^{n-k} (X_{t_{i+k}} - X_{t_i})^2. \quad (14)$$

If X is fBm, then $\mathbb{E}[V(k)] = (k\delta)^{2H}$, so:

$$\log V(k) \approx 2H \log(k\delta) + \text{const}. \quad (15)$$

The slope of $\log V(k)$ vs $\log k$ gives $2H$.

2.6.2 R/S (Rescaled Range) Analysis

For a window $[t, t+n]$, define:

$$R(n) = \max_{1 \leq k \leq n} S_k - \min_{1 \leq k \leq n} S_k, \quad S_k = \sum_{i=1}^k (X_{t+i} - \bar{X}_n), \quad (16)$$

where $\bar{X}_n$ is the sample mean. The R/S statistic is $R(n)/\hat{\sigma}(n)$, and:

$$\mathbb{E}[R(n)/\hat{\sigma}(n)] \sim c \cdot n^H \quad \text{as } n \rightarrow \infty. \quad (17)$$

Example 2.1 (Hurst Estimation for SPX Realised Variance). Using 5-minute returns on SPX (2015–2025), the variogram method gives $\hat{H}_{\text{var}} = 0.098$ and R/S analysis gives $\hat{H}_{\text{R/S}} = 0.112$. Both estimates confirm $H \approx 0.1 \ll 0.5$, validating the rough volatility hypothesis for SPX.

2.7 Exercises

1. Verify that the covariance function (7) reduces to $\min(s, t)$ when $H = \frac{1}{2}$.
2. Show that fBm is *not* a semimartingale for $H \neq \frac{1}{2}$, and explain why this makes classical Itô calculus inapplicable.
3. Compute the quadratic variation of fBm on $[0, T]$ and show it is 0 for $H > \frac{1}{2}$ and ∞ for $H < \frac{1}{2}$.
4. Implement both Hurst estimation methods on simulated fBm paths and verify they recover the true H .

Solutions

Solution to Exercise 1. For $H = 1/2$, the covariance function becomes: $R(s, t) = \frac{1}{2}(|s|^1 + |t|^1 - |t-s|^1) = \frac{1}{2}(s+t-|t-s|)$. For $0 \leq s \leq t$: $|t-s| = t-s$, so $R(s, t) = \frac{1}{2}(s+t-t+s) = s = \min(s, t)$. This is exactly the covariance of standard Brownian motion. ✓ $\square$

Solution to Exercise 3. The p -variation of fBm on $[0, T]$ is $\lim \sum_k |B_{t_{k+1}}^H - B_{t_k}^H|^p$. Since $\mathbb{E}[|B_{t_{k+1}}^H - B_{t_k}^H|^p] = c_p |t_{k+1} - t_k|^{pH}$ and there are n intervals of length T/n : $\sum_k \mathbb{E}[|\Delta B^H|^p] \sim n \cdot c_p (T/n)^{pH} = c_p T^{pH} n^{1-pH}$. This converges to 0 if $pH > 1$ and diverges if $pH < 1$. For $p = 2$ (quadratic variation): converges if $H > 1/2$, diverges if $H < 1/2$. For $H = 1/2$: QV = T (standard BM). For $H < 1/2$: infinite QV means fBm is “rougher” than BM, and classical Itô calculus (which requires finite QV) cannot be directly applied. $\square$

Solution to Exercise 2. A semimartingale must have finite quadratic variation. From Exercise 3, fBm with $H \neq 1/2$ has QV that is either 0 or ∞ , neither of which equals T (as required for a semimartingale with non-trivial diffusion). More precisely, for $H < 1/2$ the QV is infinite (fBm is too rough), and for $H > 1/2$ the QV is zero (fBm is too smooth). Since Itô calculus applies only to semimartingales, alternative theories (rough path theory, fractional calculus) are needed. $\square$

3 Volterra Integral Equations

3.1 Classical Volterra Equations

Definition 3.1 (Volterra Equation of the Second Kind). A *Volterra integral equation of the second kind* takes the form:

$$f(t) = g(t) + \int_0^t K(t, s)f(s) ds, \quad t \in [0, T], \quad (18)$$

where g is a known function, K is the *kernel*, and f is the unknown function.

Theorem 3.1 (Existence and Uniqueness). If $g \in L^2[0, T]$ and $K \in L^2([0, T]^2)$, then equation (18) has a unique solution $f \in L^2[0, T]$.

3.2 Singular Kernels and Fractional Volterra Equations

In the rough Heston model, we encounter *singular* Volterra kernels of the form:

$$K(t, s) = \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)}, \quad \alpha = H + \frac{1}{2} \in (0, 1). \quad (19)$$

This kernel is *not* square-integrable when $\alpha < \frac{1}{2}$ (i.e., $H < 0$), but for the rough volatility regime $H \in (0, \frac{1}{2})$ we have $\alpha \in (\frac{1}{2}, 1)$ and $K \in L^2$.

Definition 3.2 (Riemann–Liouville Fractional Integral). The *left Riemann–Liouville fractional integral* of order $\alpha > 0$ is:

$$I_{0+}^\alpha f(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} f(s) ds. \quad (20)$$

The rough Heston variance satisfies a fractional Volterra equation:

$$V_t = V_0 + \frac{1}{\Gamma(H + \frac{1}{2})} \int_0^t (t-s)^{H-1/2} \lambda(\theta - V_s) ds + \frac{1}{\Gamma(H + \frac{1}{2})} \int_0^t (t-s)^{H-1/2} \nu \sqrt{V_s} dW_s. \quad (21)$$

Key Insight

The presence of the singular kernel $(t-s)^{H-1/2}$ means the variance at time t depends on the *entire past history* with a power-law weighting — recent values get much more weight ($H - \frac{1}{2} < 0$, so the kernel diverges near $s = t$). This is the mathematical reason for rough paths.

3.3 The Fractional Riccati Equation

For option pricing under rough Heston, the characteristic function involves solving a *fractional Riccati equation*:

$$D_{0+}^\alpha \psi(t) = \frac{1}{2}(u^2 - u) + (\rho \nu u i - \lambda)\psi(t) + \frac{\nu^2}{2}\psi(t)^2, \quad \psi(0) = 0, \quad (22)$$

where D_{0+}^α is the Caputo fractional derivative of order $\alpha = H + \frac{1}{2}$.

Definition 3.3 (Caputo Fractional Derivative). For $0 < \alpha < 1$:

$$D_{0+}^\alpha f(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{f'(s)}{(t-s)^\alpha} ds. \quad (23)$$

Proposition 3.1. The fractional derivative and integral are inverse operations: $D_{0+}^\alpha I_{0+}^\alpha f = f$ for $f \in L^1$.

3.4 Adams Discretisation Scheme

The fractional Riccati equation (22) does not have a closed-form solution. We use the Adams predictor-corrector scheme:

Algorithm Adams Scheme for Fractional Riccati

Given time grid $0 = t_0 < t_1 < \dots < t_N = T$ with step Δt :

Predictor:

$$\psi_{n+1}^P = \psi_0 + \frac{(\Delta t)^\alpha}{\Gamma(\alpha + 1)} \sum_{j=0}^n b_{n+1-j} F(\psi_j), \quad (24)$$

Corrector:

$$\psi_{n+1} = \psi_0 + \frac{(\Delta t)^\alpha}{\Gamma(\alpha + 2)} \left[a_{n+1,0} F(\psi_{n+1}^P) + \sum_{j=0}^n a_{n+1,j+1} F(\psi_j) \right], \quad (25)$$

where $F(\psi) = \frac{1}{2}(u^2 - u) + (\rho\nu ui - \lambda)\psi + \frac{\nu^2}{2}\psi^2$, and $a_{j,k}$, b_k are explicit weights depending on α .

Proposition 3.2 (Convergence). *The Adams scheme converges with order $\min(1 + \alpha, 2)$, i.e., for $\alpha = H + \frac{1}{2} \approx 0.6$, the convergence order is ≈ 1.6 .*

3.5 Exercises

1. Verify that for $\alpha = 1$ (i.e., $H = \frac{1}{2}$), the fractional Riccati equation reduces to the classical Heston Riccati ODE.
2. Implement the Adams scheme in Python and verify convergence by comparing with finer grids.
3. Show that the Caputo derivative of a constant is zero, consistent with the initial condition $\psi(0) = 0$.

Solutions

Solution to Exercise 1. For $\alpha = 1$, $D_{0+}^1 \psi(t) = \psi'(t)$ (ordinary derivative). The fractional Riccati becomes: $\psi'(t) = \frac{1}{2}(u^2 - u) + (\rho\nu ui - \lambda)\psi + \frac{\nu^2}{2}\psi^2$, which is the classical Heston Riccati ODE for $B(u, T - t)$. The solution is $\psi(t) = B(u, t) = \frac{b-d}{\nu^2} \cdot \frac{1-e^{-dt}}{1-ge^{-dt}}$ where b, d, g are as in the Heston characteristic function. $\checkmark$ $\square$

Solution to Exercise 3. If $f(t) = c$ (constant), then $f'(s) = 0$ for all s . Therefore: $D_{0+}^\alpha c = \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{0}{(t-s)^\alpha} ds = 0$. This is consistent with $\psi(0) = 0$: the constant function $\psi = 0$ satisfies the initial condition and has vanishing Caputo derivative, so it would be a (trivial) solution if the right-hand side were zero. $\square$

4 Classical Stochastic Volatility Models

4.1 The Black–Scholes Framework and Its Limitations

The Black–Scholes model assumes constant volatility:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t. \quad (26)$$

This produces a *flat* implied volatility surface, contradicting the well-documented smile and skew observed in option markets.

4.2 The Heston Model

Definition 4.1 (Heston, 1993). The Heston stochastic volatility model specifies:

$$\frac{dS_t}{S_t} = r dt + \sqrt{V_t} dW_t^S, \quad (27)$$

$$dV_t = \lambda(\theta - V_t) dt + \nu \sqrt{V_t} dW_t^V, \quad (28)$$

where $\langle W^S, W^V \rangle_t = \rho t$, and the parameters are:

- $\lambda > 0$: mean-reversion speed;
- $\theta > 0$: long-run variance;
- $\nu > 0$: volatility of variance (“vol of vol”);
- $\rho \in [-1, 1]$: correlation (typically $\rho < 0$ for equities);
- $V_0 > 0$: initial variance.

Theorem 4.1 (Feller Condition). *The variance process V_t remains strictly positive if and only if:*

$$2\lambda\theta \geq \nu^2. \quad (29)$$

Proof. The CIR process V_t has the boundary classification: the point 0 is *entrance* (inaccessible from the interior) if and only if $2\lambda\theta/\nu^2 \geq 1$. This follows from the Feller test for explosions applied to the scale function of the CIR diffusion. $\square$

4.3 Characteristic Function of the Heston Model

Theorem 4.2 (Heston, 1993). *Under the risk-neutral measure, the characteristic function of $\log S_T$ is:*

$$\phi(u) = \mathbb{E}[e^{iu \log S_T}] = \exp(iu(\log S_0 + rT) + A(u, T) + B(u, T)V_0), \quad (30)$$

where:

$$A(u, T) = \frac{\lambda\theta}{\nu^2} \left[(b-d)T - 2 \log \frac{1 - ge^{-dT}}{1 - g} \right], \quad (31)$$

$$B(u, T) = \frac{b-d}{\nu^2} \cdot \frac{1 - e^{-dT}}{1 - ge^{-dT}}, \quad (32)$$

with:

$$b = \lambda - i\rho\nu u, \quad d = \sqrt{b^2 - \nu^2(iu - u^2)}, \quad g = \frac{b-d}{b+d}. \quad (33)$$

4.4 From Heston to Rough Heston

The classical Heston model (28) has V as a Markov process. The *rough Heston model* replaces the CIR dynamics with the fractional Volterra equation (21):

$$V_t = V_0 + \frac{1}{\Gamma(H + \frac{1}{2})} \int_0^t (t-s)^{H-1/2} [\lambda(\theta - V_s) ds + \nu \sqrt{V_s} dW_s^V]. \quad (34)$$

The key difference: V_t is now a *non-Markov*, non-semimartingale process. The characteristic function still has the exponentially affine form (30), but A and B satisfy a *fractional* Riccati equation (22) instead of an ODE.

Practical Application

Practical impact: The rough Heston model calibrates to option surfaces with RMSE of ~ 12 bps versus ~ 48 bps for classical Heston, and it does so with essentially the same number of parameters (+1 for H).

4.5 Exercises

1. Derive the classical Heston Riccati ODE for $B(u, T)$ starting from the Feynman–Kac PDE. Verify that (32) is the solution.
2. Implement the Heston characteristic function in Python and compute ATM implied volatility for typical parameters ($\lambda = 2, \theta = 0.04, \nu = 0.5, \rho = -0.7, V_0 = 0.04$).
3. Explain qualitatively why $\rho < 0$ produces a downward-sloping volatility skew.

Solutions

Solution to Exercise 1. Assume the log-price $x_t = \log S_t$ and variance V_t follow the Heston dynamics. The characteristic function $\phi(u) = \mathbb{E}[e^{iux_T}]$ has the affine form $\phi(u) = \exp(A(u, T) + B(u, T)V_0 + iux_0)$. Applying the Feynman–Kac theorem, ϕ satisfies the PDE:

$$\partial_\tau \phi = \left(-\frac{1}{2}u^2 + \frac{1}{2}iu\right)v\phi + \lambda(\theta - v)\partial_v \phi + \frac{1}{2}\nu^2 v \partial_{vv} \phi + \rho\nu v iu \partial_v \phi,$$

with $\tau = T - t$ and $\phi|_{\tau=0} = e^{iux}$. Substituting the exponentially affine ansatz $\phi = \exp(A + Bv + iux)$ gives $\dot{A} = \lambda\theta B$ and $\dot{B} = \frac{1}{2}(iu - u^2) + (\rho\nu iu - \lambda)B + \frac{1}{2}\nu^2 B^2$, which is the Heston Riccati ODE for $B(u, \tau)$ with $B(u, 0) = 0$. Define $d = \sqrt{(\rho\nu iu - \lambda)^2 - \nu^2(iu - u^2)}$ and $g = (\lambda - \rho\nu iu - d)/(\lambda - \rho\nu iu + d)$. Then:

$$B(u, \tau) = \frac{\lambda - \rho\nu iu - d}{\nu^2} \cdot \frac{1 - e^{-d\tau}}{1 - ge^{-d\tau}}.$$

□

Solution to Exercise 3. When $\rho < 0$, the Brownian motions driving price and variance are negatively correlated: $\langle dW^S, dW^V \rangle = \rho dt < 0$. This means:

- A drop in S_t ($dW^S < 0$) tends to increase V_t ($dW^V > 0$), so after a sell-off, volatility rises—the “leverage effect.”
- Higher volatility after a drop increases the value of downside protection (OTM puts), which have $K < S$ and thus $k = \log(K/S) < 0$.
- OTM calls ($K > S, k > 0$) are less valuable because rallies decrease variance.

Hence σ_{imp} is a decreasing function of k , producing a negative (downward) skew: $\partial\sigma_{\text{imp}}/\partial k < 0$. $\square$

4. **(New)** Show that in the Heston model the at-the-money (ATM) implied variance σ_{ATM}^2 satisfies $\sigma_{\text{ATM}}^2 \approx \mathbb{E}^{\mathbb{Q}}[\bar{V}]$ where $\bar{V} = T^{-1} \int_0^T V_s ds$ for short maturities. *Hint:* expand the Black–Scholes formula around $k = 0$.

Solution to Exercise 4. At the money $k = 0$, for short T , the Black–Scholes price satisfies $C_{\text{BS}} \approx S_0(\sigma\sqrt{T}/(2\sqrt{\pi}))$ up to $O(T)$ terms. In the Heston model, the price of the ATM call is $C = \mathbb{E}[(S_T - S_0)^+] \approx S_0 \mathbb{E}[\sqrt{\bar{V}}\sqrt{T}/(2\sqrt{\pi})]$ by conditioning on the volatility path. Matching the two: $\sigma_{\text{imp}} \approx \mathbb{E}[\sqrt{\bar{V}}]$ and thus $\sigma_{\text{imp}}^2 \approx \mathbb{E}[\bar{V}]$ to leading order (by Jensen’s inequality the correction is second-order in the vol-of-vol ν). $\checkmark$ $\square$

5 Fourier Pricing: The Lewis Formula

5.1 The Pricing Problem

A European call option with strike K and maturity T has price:

$$C(S_0, K, T) = e^{-rT} \mathbb{E}[(S_T - K)^+]. \quad (35)$$

If we know the characteristic function $\phi(u) = \mathbb{E}[e^{iu \log S_T}]$, we can recover the price via Fourier inversion.

5.2 Gil-Pelaez Inversion

The CDF of $\log S_T$ can be recovered from ϕ via:

$$\mathbb{P}(\log S_T \leq x) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \text{Im} \left(\frac{e^{-iux} \phi(u)}{u} \right) du. \quad (36)$$

5.3 The Lewis (2001) Formula

Theorem 5.1 (Lewis, 2001). *The European call price is:*

$$C = S_0 - \frac{\sqrt{KS_0}}{\pi} e^{-rT/2} \int_0^\infty \text{Re} \left(\frac{e^{-iuk} \phi(u - i/2)}{u^2 + 1/4} \right) du, \quad (37)$$

where $k = \log(K/S_0)$ is the log-moneyness.

Sketch of proof. Starting from the Fourier representation of the payoff $(e^x - K)^+$:

$$(e^x - K)^+ = \frac{1}{2\pi} \int_{\mathcal{C}} \frac{e^{(1+iz)x} K^{-iz}}{iz(1+iz)} dz, \quad (38)$$

where the contour $\mathcal{C}$ runs along $\text{Im}(z) = \frac{1}{2}$. Taking the risk-neutral expectation, using $\mathbb{E}[e^{(1+iz) \log S_T}] = e^{(1+iz)(rT + \log S_0)} \phi(z + i/2)$ (after careful adjustment for the drift), and simplifying gives (37). $\square$

Key Insight

The Lewis formula has a key advantage: the integrand decays as $O(u^{-2})$ for large u , making numerical quadrature efficient. The formula is valid for *any* model with a known characteristic function, including rough Heston.

5.4 Numerical Implementation

For the rough Heston model, the implementation proceeds as:

1. For each quadrature point u , solve the fractional Riccati ODE (22) with the Adams scheme to get $\psi(T; u - i/2)$.
2. Compute $\phi(u - i/2) = \exp(A(u - i/2, T) + B(u - i/2, T)V_0)$ where $B = \psi$ and $A = \lambda \theta I_{0+}^\alpha \psi(T)$.
3. Evaluate the integral (37) via Gauss-Laguerre or adaptive Simpson quadrature.

Practical Application

Computational cost: Each option price requires ~ 50 – 100 evaluations of the fractional Riccati, each involving $N = 200$ Adams steps. Total: $\sim 10^4$ elementary operations per price. This is fast enough for real-time calibration on modern hardware ($\sim 200 \mu\text{s}$ per price in our Rust implementation).

5.5 Exercises

1. Implement the Lewis formula for the classical Heston model (where ϕ is available in closed form) and verify against the analytic Heston formula.
2. Show that the Lewis integrand for ATM options ($k = 0$) reduces to $\text{Re}[\phi(u - i/2)]/(u^2 + 1/4)$.
3. Investigate the effect of truncating the integration at $u_{\max}$ on pricing accuracy. What is the minimum $u_{\max}$ needed for < 0.01 bps error?

Solutions

Solution to Exercise 1. For the Heston model, $\phi(u) = \exp(A(u, T) + B(u, T)V_0 + iu \log S_0)$ is available in closed form. Substituting into (37):

$$C_{\text{Lewis}} = S_0 - \frac{\sqrt{KS_0}}{\pi} e^{-rT/2} \int_0^\infty \frac{\text{Re}[e^{-iuk}\phi(u - i/2)]}{u^2 + 1/4} du.$$

To verify, we compare against the analytic Heston formula (which uses two separate probability integrals P_1, P_2): $C_{\text{Heston}} = S_0 P_1 - K e^{-rT} P_2$. Numerically, using Gauss–Laguerre quadrature with 64 nodes and typical parameters, one finds $|C_{\text{Lewis}} - C_{\text{Heston}}| < 10^{-12}$ (machine precision), confirming correctness. $\square$

Solution to Exercise 2. At the money, $k = \log(K/S_0) = 0$, so $e^{-iuk} = 1$. The integrand becomes:

$$\frac{\text{Re}[\phi(u - i/2)]}{u^2 + 1/4}.$$

This simplification means we only need the *real part* of ϕ evaluated along the line $\text{Im}(u) = -1/2$ in the complex plane. Since $\phi(-i/2) = \mathbb{E}[S_T/\sqrt{S_0 K}] \cdot e^{rT/2}$ is real and positive, the integrand is smooth and the quadrature converges rapidly. $\checkmark$ $\square$

Solution to Exercise 3. The integrand decays as $|\phi(u - i/2)|/(u^2 + 1/4) \sim u^{-2}$ for large u , so convergence is algebraic, not exponential. Experimentally, for the Heston model with typical parameters:

$u_{\max}$	Error (bps)
10	~ 0.5
20	~ 0.05
50	~ 0.003
100	~ 0.0003

Thus $u_{\max} \approx 30$ – 50 typically suffices for < 0.01 bps error. For rough Heston with small H , the characteristic function oscillates more at high frequencies, and $u_{\max} \approx 80$ – 100 may be needed. $\square$

6 The Rough Heston Model

6.1 Model Specification

Definition 6.1 (El Euch & Rosenbaum, 2019). The *rough Heston model* specifies:

$$\frac{dS_t}{S_t} = r dt + \sqrt{V_t} dW_t^S, \quad (39)$$

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} [\lambda(\theta - V_s) ds + \nu \sqrt{V_s} dW_s^V], \quad (40)$$

where $\alpha = H + \frac{1}{2} \in (\frac{1}{2}, 1)$, $\langle W^S, W^V \rangle_t = \rho t$.

The parameters are the same as classical Heston, plus the Hurst parameter H :

Parameter	Symbol	Range	Typical
Hurst exponent	H	$(0, 0.5)$	0.10
Mean-reversion speed	λ	$(0, \infty)$	0.30
Long-run variance	θ	$(0, \infty)$	0.04
Vol of vol	ν	$(0, \infty)$	0.45
Correlation	ρ	$[-1, 1]$	-0.78
Initial variance	V_0	$(0, \infty)$	0.035

6.2 Existence and Positivity

Theorem 6.1 (El Euch & Rosenbaum, 2019). *The rough Heston variance V_t exists, is unique in law, and remains non-negative almost surely. No Feller condition is needed (unlike classical Heston).*

Remark 6.1. The non-negativity of V_t in the rough case is more subtle than for CIR. It relies on comparison theorems for Volterra equations and the weak singularity of the kernel.

6.3 Characteristic Function

Theorem 6.2 (El Euch & Rosenbaum, 2019). *The characteristic function of $\log S_T$ under rough Heston has the exponentially affine form:*

$$\phi(u) = \exp(iu(\log S_0 + rT) + \theta \lambda I^\alpha \psi(T) + V_0 I^1 \psi(T)), \quad (41)$$

where ψ solves the fractional Riccati equation (22) and I^α , I^1 denote fractional and ordinary integrals respectively.

6.4 Calibration via Levenberg–Marquardt

To calibrate the model to observed option prices:

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_{i=1}^M w_i (\sigma_i^{\text{market}} - \sigma_i^{\text{model}}(\theta))^2, \quad (42)$$

where σ_i denotes implied volatility and w_i are vega weights. The Levenberg–Marquardt algorithm is used because the Jacobian $\partial \sigma^{\text{model}} / \partial \theta_j$ can be computed via finite differences of the characteristic function.

HFThot Lab Reference

The HFThot Lab `optimizr` Rust library provides a parallel Levenberg–Marquardt implementation that calibrates rough Heston in $< 200 \mu\text{s}$ per strike, enabling real-time recalibration.

6.5 Exercises

1. Show that setting $H = \frac{1}{2}$ in the rough Heston model recovers the classical Heston model exactly.
2. Implement the rough Heston characteristic function using your Adams solver from Section 3 and compare implied volatilities with the classical Heston model for different values of H .
3. Investigate the sensitivity of the ATM skew $\partial\sigma/\partial k|_{k=0}$ to H . Show that it scales approximately as $T^{H-1/2}$ for short maturities.

Solutions

Solution to Exercise 1. The rough Heston variance is $V_t = V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} [\lambda(\theta - V_s) ds + \nu\sqrt{V_s} dW_s^V]$ with $\alpha = H + 1/2$. For $H = 1/2$, $\alpha = 1$, the kernel $(t-s)^{\alpha-1} = (t-s)^0 = 1$ and $1/\Gamma(1) = 1$, so: $V_t = V_0 + \int_0^t [\lambda(\theta - V_s) ds + \nu\sqrt{V_s} dW_s^V]$. Differentiating: $dV_t = \lambda(\theta - V_t) dt + \nu\sqrt{V_t} dW_t^V$, which is the classical Heston SDE. The fractional Riccati also reduces to ODE form as shown in the Volterra exercise. ✓ □

Solution to Exercise 3. From the short-maturity expansion (Bayer–Friz–Gatheral, 2016), the ATM implied volatility skew is:

$$\left. \frac{\partial\sigma_{\text{imp}}}{\partial k} \right|_{k=0} \sim C(H, \rho, \nu) \cdot T^{H-1/2} \quad \text{as } T \rightarrow 0,$$

where C depends on model parameters but not on T . Numerically, one can verify this by computing $\sigma_{\text{imp}}(k, T)$ for $k \in \{-0.01, 0, 0.01\}$ and $T \in \{1/252, 5/252, 21/252, 63/252\}$, estimating the skew via finite differences, and plotting $\log|\text{skew}|$ vs. $\log T$. The slope should equal $H - 1/2$:

- $H = 0.1$: slope ≈ -0.4 (steep short-maturity skew)
- $H = 0.3$: slope ≈ -0.2 (moderate)
- $H = 0.5$: slope ≈ 0 (flat, classical Heston)

This confirms the power-law scaling $T^{H-1/2}$. ✓ □

4. **(New)** Show that the rough Heston variance process V_t is *not* a semimartingale when $H < 1/2$, using the fact that its paths have infinite quadratic variation on compact intervals.

Solution to Exercise 4. The variance V_t is driven by the Volterra-type stochastic integral: $V_t - V_0 = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \nu\sqrt{V_s} dW_s + \dots$. For the stochastic part, the kernel $K(t, s) = (t-s)^{\alpha-1}/\Gamma(\alpha)$ with $\alpha = H + 1/2 < 1$ makes V_t behave locally like fBm with index H . As shown in the fBm section, for $H < 1/2$ the quadratic variation of fBm is infinite on any interval: $\langle V \rangle_T = +\infty$. Since every semimartingale has finite quadratic variation, V_t cannot be a semimartingale. This has practical consequences: standard Itô calculus does not apply directly to functions of V_t , which is why the Volterra/fractional calculus framework is essential. ✓ □

7 Implied Volatility and the Volatility Surface

7.1 Definition and Computation

Definition 7.1 (Implied Volatility). The *implied volatility* $\sigma_{\text{imp}}(K, T)$ is the unique value σ such that:

$$C_{\text{BS}}(S_0, K, T; \sigma) = C_{\text{market}}(K, T), \quad (43)$$

where C_{BS} is the Black–Scholes formula.

This is computed numerically via Newton’s method, using the vega $\partial C_{\text{BS}}/\partial \sigma$ for the derivative.

7.2 The Volatility Smile and Skew

- **Smile:** σ_{imp} is a U-shaped function of K (or log-moneyness k). Both deep ITM and OTM options have higher IV.
- **Skew:** For equity indices, $\partial \sigma_{\text{imp}}/\partial k < 0$ — OTM puts (low K) have higher IV than OTM calls (high K).
- **Term structure:** Short-maturity smiles are steeper and more pronounced than long-maturity smiles.

7.3 Short-Time Asymptotics and Roughness

Theorem 7.1 (Bayer, Friz, Gatheral, 2016). *Under the rough Heston model, the ATM implied variance skew satisfies:*

$$\left. \frac{\partial \sigma_{\text{imp}}^2}{\partial k} \right|_{k=0} \propto T^{H-1/2} \quad \text{as } T \rightarrow 0. \quad (44)$$

For classical Heston ($H = 0.5$), the skew is $O(1)$ as $T \rightarrow 0$, which is *too flat* compared to market data. For rough Heston ($H \approx 0.1$), the skew blows up as $T^{-0.4}$ for short maturities, matching empirical observations.

Key Insight

This asymptotic result is perhaps the strongest theoretical argument for rough volatility: the short-maturity skew behavior is a *signature* of roughness, and classical models cannot reproduce it regardless of parameter choice.

7.4 The Variance Risk Premium

The *variance risk premium* (VRP) is:

$$\text{VRP}(\tau) = \mathbb{E}^{\mathbb{P}}[\text{RV}_\tau] - \mathbb{E}^{\mathbb{Q}}[\text{RV}_\tau] = \text{RV}_\tau^{\text{realised}} - \text{IV}_\tau^2, \quad (45)$$

where RV is realised variance and IV is implied variance over horizon τ .

Theorem 7.2 (Rough VRP Term Structure). *Under rough volatility, the VRP follows a power-law term structure:*

$$|\text{VRP}(\tau)| \propto \tau^{2H-1}. \quad (46)$$

For $H = 0.1$: $|\text{VRP}(\tau)| \propto \tau^{-0.8}$, implying that the VRP is largest at short horizons and decays as a power law.

7.5 Exercises

1. For the Black–Scholes model, verify that $\sigma_{\text{imp}}(K, T) = \sigma$ (constant) for all K, T .
2. Implement Newton’s method for implied volatility computation. What convergence rate do you observe?
3. Plot the implied volatility surface for rough Heston with $H \in \{0.05, 0.1, 0.2, 0.3, 0.5\}$ and observe how roughness affects the short-maturity skew.

Solutions

Solution to Exercise 1. In Black–Scholes, $S_t = S_0 \exp((r - \sigma^2/2)t + \sigma W_t)$ with *constant* volatility σ . The model’s option price is $C_{\text{BS}}(S_0, K, T; \sigma)$ by definition. Since the market price is *also* generated by the same model, $C_{\text{market}} = C_{\text{BS}}(\sigma)$. Therefore the implied volatility satisfies $C_{\text{BS}}(\sigma_{\text{imp}}) = C_{\text{BS}}(\sigma)$, which gives $\sigma_{\text{imp}} = \sigma$ for all K, T .

This shows that any deviation from a flat IV surface is evidence against constant volatility, which motivates stochastic and rough volatility models. $\square$

Solution to Exercise 2. Newton’s method for σ_{imp} solves $f(\sigma) = C_{\text{BS}}(\sigma) - C_{\text{market}} = 0$:

$$\sigma_{n+1} = \sigma_n - \frac{C_{\text{BS}}(\sigma_n) - C_{\text{market}}}{\partial C_{\text{BS}} / \partial \sigma} = \sigma_n - \frac{C_{\text{BS}}(\sigma_n) - C_{\text{market}}}{\text{vega}(\sigma_n)},$$

where $\text{vega} = S_0 \sqrt{T} \varphi(d_1) > 0$ (with φ the standard normal PDF). Since C_{BS} is a *convex* function of σ (positive $\partial^2 C / \partial \sigma^2 = \text{vomma} \geq 0$, with equality only at $d_1 = 0$), Newton’s method converges *quadratically*: $|\sigma_{n+1} - \sigma^*| \leq M |\sigma_n - \sigma^*|^2$ for some $M > 0$. Typically 3–5 iterations suffice for 12-digit accuracy with a reasonable initial guess $\sigma_0 = 0.2$. $\square$

Solution to Exercise 3. Using the Lewis formula with rough Heston CF for different H values and plotting $\sigma_{\text{imp}}(k, T)$, one observes:

- $H = 0.5$ (classical): Mild smile that flattens at short maturity.
- $H = 0.3$: Steeper short-maturity skew, smile slightly more pronounced.
- $H = 0.1$: Very steep short-maturity skew, diverging as $T \rightarrow 0$ with power law $T^{H-1/2} = T^{-0.4}$.
- $H = 0.05$: Extreme short-maturity skew, approaching $T^{-0.45}$.

The key observation: as H decreases, the short-maturity end of the surface becomes progressively *steeper*, while the long-maturity end remains similar across models. This matches empirical observations in equity and FX markets, where 1-week skew is much steeper than 1-year skew. $\square$

8 Path Signatures

8.1 Motivation: From Time Series to Features

Traditional feature engineering for time series (returns, moving averages, RSI) is ad hoc and lossy. *Path signatures* provide a principled, mathematically grounded feature map that is:

- **Universal:** captures all information in the path;
- **Graded:** naturally truncated at any *level* for bias-variance control;
- **Invariant:** to time reparametrisation (for the signature itself).

8.2 Definition of the Path Signature

Definition 8.1 (Path Signature). Let $X : [0, T] \rightarrow \mathbb{R}^d$ be a continuous path of bounded variation. The *signature* of X is the collection of iterated integrals:

$$S(X)_{0,T}^{i_1 \dots i_k} = \int_{0 < t_1 < \dots < t_k < T} dX_{t_1}^{i_1} \dots dX_{t_k}^{i_k}, \quad (47)$$

for all multi-indices $(i_1, \dots, i_k)$ with $i_j \in \{1, \dots, d\}$ and all levels $k \geq 1$.

Example 8.1 (Level 1). At level 1, the signature is simply the increment:

$$S(X)^i = \int_0^T dX_t^i = X_T^i - X_0^i. \quad (48)$$

Example 8.2 (Level 2). At level 2, we get the “area” terms:

$$S(X)^{ij} = \int_0^T \int_0^t dX_s^i dX_t^j. \quad (49)$$

For $d = 2$, the antisymmetric part $S(X)^{12} - S(X)^{21}$ equals twice the signed area enclosed by the path projected onto the (x_1, x_2) -plane.

8.3 Dimension Counting

Proposition 8.1. *The signature truncated at level M for a d -dimensional path has total dimension:*

$$\dim(S^{\leq M}) = \sum_{k=1}^M d^k = \frac{d(d^M - 1)}{d - 1}. \quad (50)$$

For $d = 5$ (typical for volatility trading features: price, volume, IV, RV, VRP):

Level M	Dimension	Comment
1	5	Increments only
2	30	+ pairwise interactions
3	155	+ triple interactions
4	780	Risk of overfitting
5	3,905	Severe overfitting

Key Insight

The signature truncation level M is the key hyperparameter. In WP2 we find that $M = 3$ gives the best out-of-sample Sharpe ratio: $M < 3$ underfits, $M > 3$ overfits.

8.4 Key Algebraic Properties**8.4.1 Shuffle Product**

Theorem 8.1 (Ree, 1958). *The signature satisfies the shuffle product identity: for any words w_1, w_2 :*

$$S(X)^{w_1} \cdot S(X)^{w_2} = \sum_{w \in w_1 \sqcup w_2} S(X)^w, \quad (51)$$

where $w_1 \sqcup w_2$ is the set of all shuffles of w_1 and w_2 (interleavings that preserve the internal ordering of each word).

Example 8.3. For $w_1 = (1)$ and $w_2 = (2)$:

$$S(X)^1 \cdot S(X)^2 = S(X)^{12} + S(X)^{21}. \quad (52)$$

This says: “the product of increments equals the sum of ordered pairwise integrals.”

8.4.2 Chen’s Identity

Theorem 8.2 (Chen, 1957). *For concatenated paths $X * Y$:*

$$S(X * Y) = S(X) \otimes S(Y), \quad (53)$$

where $\otimes$ is the tensor product in the tensor algebra. This means signatures are multiplicative under path concatenation.

Practical Application

Practical use of Chen’s identity: When processing streaming data, we can compute the signature incrementally. Given the signature $S(X_{[0,t]})$ and a new segment $X_{[t,t+\delta]}$, we update via: $S(X_{[0,t+\delta]}) = S(X_{[0,t]}) \otimes S(X_{[t,t+\delta]})$, without recomputing from scratch.

8.4.3 Uniqueness (Hambly–Lyons)

Theorem 8.3 (Hambly & Lyons, 2010). *The signature uniquely determines the path up to tree-like equivalence (sequences of excursions that cancel pairwise).*

8.5 The Log-Signature

Definition 8.2 (Log-Signature). The *log-signature* is the logarithm of the signature in the tensor algebra:

$$\text{LogSig}(X) = \log S(X) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} (S(X) - \mathbf{1})^n. \quad (54)$$

The log-signature is more compact than the signature, with dimension growing as d^M/M rather than d^M , and it uses the Lyndon basis for further compression.

8.6 Signatures as Universal Nonlinear Features

Theorem 8.4 (Kiraly & Oberhauser, 2019). *Let f be any continuous functional on (un-parametrised) paths. Then f can be approximated arbitrarily well by a linear functional of the signature. That is, signatures form a universal feature set for path-valued data.*

Key Insight

This universality theorem is the theoretical justification for using *linear regression on signature features* as a powerful nonlinear model. The nonlinearity is entirely captured by the signature transform; the downstream model can be simple and interpretable.

8.7 Exercises

1. Compute the level-2 signature of the path $X_t = (t, t^2)$ on $[0, 1]$ by hand.
2. Verify the shuffle product for $S(X)^1 \cdot S(X)^2 = S(X)^{12} + S(X)^{21}$ on a numerical example.
3. Implement a pure-Python signature computation up to level 3 for discrete paths and verify on a known example.
4. Explain why the signature is invariant to time reparametrisation (hint: change of variables in the iterated integral).

Solutions

Solution to Exercise 1. Let $X_t = (t, t^2)$ on $[0, 1]$. The level-1 terms are:

$$S(X)^1 = \int_0^1 dX_t^1 = \int_0^1 dt = 1, \quad S(X)^2 = \int_0^1 dX_t^2 = \int_0^1 2t dt = 1.$$

The level-2 terms are:

$$\begin{aligned} S(X)^{11} &= \int_0^1 \int_0^{t_2} dt_1 dt_2 = \frac{1}{2}, \\ S(X)^{12} &= \int_0^1 \int_0^{t_2} dt_1 \cdot 2t_2 dt_2 = \int_0^1 t_2 \cdot 2t_2 dt_2 = \frac{2}{3}, \\ S(X)^{21} &= \int_0^1 \int_0^{t_2} 2t_1 dt_1 dt_2 = \int_0^1 t_2^2 dt_2 = \frac{1}{3}, \\ S(X)^{22} &= \int_0^1 \int_0^{t_2} 2t_1 \cdot 2t_2 dt_1 dt_2 = \int_0^1 t_2^2 \cdot 2t_2 dt_2 = \frac{1}{2}. \end{aligned}$$

Note: $S(X)^{12} + S(X)^{21} = \frac{2}{3} + \frac{1}{3} = 1 = S(X)^1 \cdot S(X)^2$, confirming the shuffle identity. Also, the signed area is $\frac{1}{2}(S^{12} - S^{21}) = \frac{1}{2}(\frac{2}{3} - \frac{1}{3}) = \frac{1}{6}$. $\square$

Solution to Exercise 2. Let $X : [0, 1] \rightarrow \mathbb{R}^2$ be any path. By definition:

$$S(X)^{12} + S(X)^{21} = \int_0^1 \int_0^t dX_s^1 dX_t^2 + \int_0^1 \int_0^t dX_s^2 dX_t^1.$$

In the first integral, exchange variable names ($s \leftrightarrow t$) and reverse the order of integration: $\int_0^1 \int_0^t dX_s^1 dX_t^2 = \int_0^1 X_{[0,t]}^1 dX_t^2$ where $X_{[0,t]}^1 = \int_0^t dX_s^1$. Summing both integrals and using integration by parts (the “area formula”): $S^{12} + S^{21} = (\int_0^1 dX^1)(\int_0^1 dX^2) = S^1 \cdot S^2$. $\checkmark$

Numerical verification: for the path $X_n = (\cos(2\pi n/N), \sin(2\pi n/N))$ with $N = 100$ (a circle), compute signatures numerically and verify $S^{12} + S^{21} = S^1 \cdot S^2 \approx 0$ (since $S^1 = S^2 \approx 0$). $\square$

Solution to Exercise 4. Let $\varphi : [0, T] \rightarrow [0, T]$ be a monotone increasing reparametrisation with $\varphi(0) = 0$, $\varphi(T) = T$. The reparametrised path is $\tilde{X}_t = X_{\varphi(t)}$. For the level-1 signature:

$$S(\tilde{X})^i = \int_0^T d\tilde{X}_t^i = \int_0^T dX_{\varphi(t)}^i = \int_0^T X_{\varphi(t)}^i \varphi'(t) dt \stackrel{s=\varphi(t)}{=} \int_0^T dX_s^i = S(X)^i.$$

The same change of variables $s_k = \varphi(t_k)$ applies to each iterated integral at every level, since the Jacobian factors $\varphi'(t_k)$ cancel with the $dt_k \rightarrow ds_k$ change of measure. Hence $S(\tilde{X})^{i_1 \dots i_k} = S(X)^{i_1 \dots i_k}$ for all multi-indices.

Intuitively, the signature measures “how much the path moves in each direction and in what order,” regardless of the speed at which it is traversed. $\square$

9 Tensor Algebra

9.1 The Tensor Algebra $T((\mathbb{R}^d))$

Definition 9.1. The *tensor algebra* over $\mathbb{R}^d$ is:

$$T((\mathbb{R}^d)) = \prod_{k=0}^{\infty} (\mathbb{R}^d)^{\otimes k} = \mathbb{R} \oplus \mathbb{R}^d \oplus (\mathbb{R}^d \otimes \mathbb{R}^d) \oplus \dots \quad (55)$$

An element $\mathbf{a} \in T((\mathbb{R}^d))$ is a sequence $\mathbf{a} = (a_0, a_1, a_2, \dots)$ where $a_k \in (\mathbb{R}^d)^{\otimes k}$.

The tensor algebra is equipped with:

- **Addition:** componentwise;
- **Multiplication:** $(\mathbf{a} \otimes \mathbf{b})_k = \sum_{j=0}^k a_j \otimes b_{k-j}$;
- **Unit:** $\mathbf{1} = (1, 0, 0, \dots)$.

9.2 The Truncated Tensor Algebra

For computation, we work with the *truncated* tensor algebra:

$$T^M(\mathbb{R}^d) = \bigoplus_{k=0}^M (\mathbb{R}^d)^{\otimes k}. \quad (56)$$

This is a finite-dimensional algebra of dimension $1 + d + d^2 + \dots + d^M$.

9.3 The Free Lie Algebra and Lyndon Words

The *free Lie algebra* $\mathfrak{g}^M(\mathbb{R}^d)$ is the Lie subalgebra of $T^M(\mathbb{R}^d)$ generated by the bracket $[a, b] = a \otimes b - b \otimes a$. The log-signature lives in this Lie algebra, and its dimension is:

$$\dim(\mathfrak{g}^M(\mathbb{R}^d)) = \sum_{k=1}^M \frac{1}{k} \sum_{j|k} \mu(k/j) d^j, \quad (57)$$

where μ is the Möbius function. A basis is provided by *Lyndon words*.

Definition 9.2 (Lyndon Word). A *Lyndon word* over alphabet $\{1, \dots, d\}$ is a word that is strictly lexicographically smaller than all its proper suffixes.

Example 9.1. For $d = 2$: Lyndon words up to length 3 are: 1, 2, 12, 112, 122. Dimension of log-signature: $2 + 1 + 2 = 5$ (vs. $2 + 4 + 8 = 14$ for the signature).

9.4 Computational Aspects

Computing the signature of a discrete d -dimensional path with n time steps at truncation level M costs:

$$O\left(n \cdot \frac{d(d^M - 1)}{d - 1}\right) \quad (58)$$

using Chen's identity applied step by step. For the Rust implementation in `optimizr`, this is highly parallelisable via SIMD.

9.5 Exercises

1. Verify that $(\mathbb{R}^2)^{\otimes 2}$ has dimension 4 with basis $\{e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2\}$.
2. Show that $[e_1, e_2] = e_1 \otimes e_2 - e_2 \otimes e_1$ spans a 1-dimensional subspace of $(\mathbb{R}^2)^{\otimes 2}$, consistent with the Lyndon word count.
3. List all Lyndon words of length ≤ 4 for $d = 2$ and verify the dimension formula.

Solutions

Solution to Exercise 1. A basis for $(\mathbb{R}^2)^{\otimes 2}$ consists of all tensor products of pairs from $\{e_1, e_2\}$: $e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2$. This has $2^2 = 4$ elements. Any element $A \in (\mathbb{R}^2)^{\otimes 2}$ can be written as $A = \sum_{i,j \in \{1,2\}} a_{ij} e_i \otimes e_j$, so these 4 tensors form a basis. $\checkmark$ $\square$

Solution to Exercise 2. $[e_1, e_2] = e_1 \otimes e_2 - e_2 \otimes e_1$. This is a single non-zero vector in the 4-dimensional space $(\mathbb{R}^2)^{\otimes 2}$, spanning a 1-dimensional subspace. From the Lyndon word point of view, the only Lyndon word of length 2 over $\{1, 2\}$ is “12” (since $1 < 2$). The word “21” is not Lyndon because $21 > 1$ (its suffix). Hence $\dim(\mathfrak{g}^2) = 1$ at level 2, matching the bracket $[e_1, e_2]$.

The symmetric part $e_1 \otimes e_2 + e_2 \otimes e_1$ does *not* lie in the Lie algebra—it is captured by the shuffle relation $S^1 \cdot S^2 = S^{12} + S^{21}$, showing that only the antisymmetric part carries independent information. $\square$

Solution to Exercise 3. Lyndon words over $\{1, 2\}$ by length:

- Length 1: 1, 2 (2 words)
- Length 2: 12 (1 word)
- Length 3: 112, 122 (2 words)
- Length 4: 1112, 1122, 1222 (3 words)

Total: $2 + 1 + 2 + 3 = 8$ Lyndon words.

Using the Möbius dimension formula:

$$\dim(\mathfrak{g}^k) = \frac{1}{k} \sum_{j|k} \mu(k/j) 2^j.$$

- $k = 1$: $\frac{1}{1} \mu(1) \cdot 2^1 = 2$
- $k = 2$: $\frac{1}{2} [\mu(1) \cdot 4 + \mu(2) \cdot 2] = \frac{1}{2} [4 - 2] = 1$
- $k = 3$: $\frac{1}{3} [\mu(1) \cdot 8 + \mu(3) \cdot 2] = \frac{1}{3} [8 - 2] = 2$
- $k = 4$: $\frac{1}{4} [\mu(1) \cdot 16 + \mu(2) \cdot 4 + \mu(4) \cdot 2] = \frac{1}{4} [16 - 4 + 0] = 3$

Cumulative: $2 + 1 + 2 + 3 = 8$. $\checkmark$ $\square$

10 Volatility Trading Strategies

10.1 Variance Swaps and Realised Variance

Definition 10.1 (Realised Variance). The *annualised realised variance* over $[t, t + \tau]$ from n log-returns $r_i = \log(S_{t+i\Delta}/S_{t+(i-1)\Delta})$, $\Delta = \tau/n$, is:

$$\text{RV}_{t,\tau} = \frac{252}{n} \sum_{i=1}^n r_i^2. \quad (59)$$

Definition 10.2 (Variance Swap). A variance swap with notional N_{var} pays:

$$\text{Payoff} = N_{\text{var}} \cdot (\text{RV}_{t,\tau} - K_{\text{var}}), \quad (60)$$

where K_{var} is the strike (fixed at inception).

10.2 Delta-Hedged Option Positions

A delta-hedged long call position earns the *gamma P&L*:

$$dP\&L_t = \frac{1}{2} S_t^2 \Gamma_t (\sigma_t^2 - \sigma_{\text{imp}}^2) dt, \quad (61)$$

where Γ_t is the Black–Scholes gamma and σ_t is the realised volatility. This position profits when realised variance exceeds implied variance.

10.3 Signature-Based Volatility Arbitrage

The signature-based strategy uses the path signature of a multivariate volatility feature vector as input to a linear trading rule:

$$w_t = \beta^\top S(X)_{[t-\tau, t]}^{\leq M}, \quad (62)$$

where $X_t = (\text{price}_t, \text{IV}_t, \text{RV}_t, \text{VRP}_t, \text{volume}_t) \in \mathbb{R}^5$ and β is estimated by ridge regression on training data.

Practical Application

Key result from WP2: The optimal truncation level is $M = 3$, giving 155 features. Ridge regularisation with cross-validated λ achieves out-of-sample Sharpe ratio 1.55, versus 0.92 for classical volatility arbitrage.

10.4 Exercises

1. Derive the gamma P&L formula from the Black–Scholes PDE.
2. Show that a variance swap can be replicated by a portfolio of options with weights $1/K^2$.
3. Implement a signature-based linear model for volatility prediction and compare with a simple moving-average model.

Solutions

Solution to Exercise 1. Under the Black–Scholes model, a delta-hedged long call has value $\Pi_t = C(S_t, t) - \Delta_t S_t$ where $\Delta_t = \partial C / \partial S$. By Itô's lemma:

$$d\Pi_t = \left(\frac{\partial C}{\partial t} + \frac{1}{2} \sigma_t^2 S_t^2 \frac{\partial^2 C}{\partial S^2} \right) dt + \frac{\partial C}{\partial S} dS_t - \Delta_t dS_t.$$

Since $\Delta_t = \partial C / \partial S$, the dS_t terms cancel. But if the hedge uses *implied* vol σ_{imp} , the BS PDE gives $\partial C / \partial t = rC - rS\Delta - \frac{1}{2} \sigma_{\text{imp}}^2 S^2 \Gamma$. Substituting:

$$d\Pi_t = \frac{1}{2} S_t^2 \Gamma_t (\sigma_t^2 - \sigma_{\text{imp}}^2) dt,$$

where $\Gamma_t = \partial^2 C / \partial S^2$ is the gamma, after subtracting the risk-free financing cost. The position profits when realised variance exceeds implied variance. $\square$

Solution to Exercise 2. The fair strike of a variance swap equals the risk-neutral expected realised variance. By the Carr–Madan (1998) replication argument:

$$\mathbb{E}^{\mathbb{Q}} \left[\int_0^T \sigma_t^2 dt \right] = \frac{2}{T} \left[\int_0^F \frac{P(K)}{K^2} dK + \int_F^\infty \frac{C(K)}{K^2} dK \right],$$

where $F = S_0 e^{rT}$ is the forward price. This follows from $-\log(S_T/F) = \int_0^T \frac{(K-S_T)^+}{K^2} dK + \int_F^\infty \frac{(S_T-K)^+}{K^2} dK - \frac{S_T-F}{F}$, and $\mathbb{E}^{\mathbb{Q}}[-2\log(S_T/F)]/T = \mathbb{E}^{\mathbb{Q}}[\int_0^T \sigma_t^2 dt]/T$ by Itô's lemma applied to $\log S_T$. The weights $1/K^2$ arise naturally from the payoff $-\log(S_T/F)$, not from an approximation. $\square$

Solution to Exercise 3. The signature-based model predicts next-day realised variance as: $\hat{\sigma}_{t+1}^2 = \beta^\top S(X)_{[t-\tau, t]}^{\leq M}$, where X is the feature path. A moving-average model uses: $\hat{\sigma}_{t+1}^2 = \frac{1}{\tau} \sum_{s=t-\tau+1}^t r_s^2$. The signature model captures *path-dependent* patterns (trends in volatility, mean-reversion speed changes, leverage effects) via the higher-order iterated integrals, while the MA model only uses the marginal distribution. Log-signature features with ridge regression typically improve MSE by 15–25% over the MA model, with the most significant gains during regime transitions. $\square$

11 Calibration and Numerical Methods

11.1 The Calibration Problem

Model calibration seeks parameters $\boldsymbol{\theta} = (H, \lambda, \theta, \nu, \rho, V_0)$ that minimize the distance between model and market implied volatilities:

$$\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}} \sum_{i=1}^M w_i (\sigma_i^{\text{market}} - \sigma_i^{\text{model}}(\boldsymbol{\theta}))^2. \quad (63)$$

11.2 Levenberg–Marquardt Algorithm

Algorithm Levenberg–Marquardt

Given residuals $\mathbf{r}(\boldsymbol{\theta}) = (\sigma_i^{\text{market}} - \sigma_i^{\text{model}})_{i=1}^M$:

1. Compute Jacobian $J_{ij} = \partial r_i / \partial \theta_j$ via finite differences.
2. Solve the regularised normal equations: $(J^\top J + \mu I) \Delta \boldsymbol{\theta} = -J^\top \mathbf{r}$.
3. Update $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \Delta \boldsymbol{\theta}$.
4. Adjust μ : decrease if objective improved, increase otherwise.
5. Iterate until convergence.

The damping parameter μ interpolates between Gauss–Newton ($\mu \rightarrow 0$) and gradient descent ($\mu \rightarrow \infty$).

11.3 Gauss–Laguerre Quadrature

For the Lewis integral (37), we use Gauss–Laguerre quadrature:

$$\int_0^\infty f(u) du \approx \sum_{k=1}^{N_q} w_k f(u_k), \quad (64)$$

where (u_k, w_k) are the Gauss–Laguerre nodes and weights. With $N_q = 64$ points, this achieves accuracy better than 0.01 bps for typical parameter ranges.

11.4 Differential Evolution for Global Optimisation

For the initial parameter guess, we use *differential evolution* (DE), a global optimiser that avoids local minima:

Algorithm Differential Evolution

1. Initialise a population of N_p parameter vectors uniformly in bounds.
2. For each vector $\boldsymbol{\theta}_i$:
 - (a) Select three distinct vectors $\boldsymbol{\theta}_{r_1}, \boldsymbol{\theta}_{r_2}, \boldsymbol{\theta}_{r_3}$.
 - (b) Mutant: $\mathbf{v} = \boldsymbol{\theta}_{r_1} + F(\boldsymbol{\theta}_{r_2} - \boldsymbol{\theta}_{r_3})$.
 - (c) Trial: crossover $\mathbf{u}$ from $\mathbf{v}$ and $\boldsymbol{\theta}_i$ with probability CR .

(d) Replace θ_i with $\mathbf{u}$ if $f(\mathbf{u}) < f(\theta_i)$.

3. Iterate until convergence.

HFThot Lab Reference

The `optimizr` library implements a parallel adaptive DE with population size $4d + 10$, $F = 0.8$, $CR = 0.9$, and automatic convergence detection based on population variance.

11.5 Exercises

1. Implement the Levenberg–Marquardt algorithm for calibrating the classical Heston model and verify convergence from different starting points.
2. Compare the accuracy of Gauss–Laguerre with $N_q \in \{16, 32, 64, 128\}$ quadrature points for the Lewis integral.
3. Implement differential evolution in Python and use it to find the global minimum of the Rastrigin function.

Solutions

Solution to Exercise 1. Levenberg–Marquardt for Heston calibration:

1. **Initialise:** $\theta_0 = (\lambda = 2, \theta = 0.04, \nu = 0.5, \rho = -0.7, V_0 = 0.04), \mu = 10^{-3}$.
2. **Residuals:** For each of M market strikes/maturities, compute $r_i = \sigma_i^{\text{market}} - \sigma_i^{\text{model}}(\theta)$ using the Lewis formula with analytic Heston CF.
3. **Jacobian:** $J_{ij} = -\partial \sigma_i^{\text{model}} / \partial \theta_j$ via central finite differences: $J_{ij} \approx (\sigma_i(\theta_j + h) - \sigma_i(\theta_j - h)) / (2h)$ with $h = 10^{-5}|\theta_j| + 10^{-8}$.
4. **Update:** Solve $(J^\top J + \mu I) \Delta \theta = J^\top r$ and set $\theta \leftarrow \theta + \Delta \theta$.
5. **Damping:** If MSE decreases, accept and set $\mu \leftarrow \mu/10$. Otherwise reject and set $\mu \leftarrow 10\mu$.

Typically converges in 15–30 iterations. Multiple starting points (3–5) guard against local minima. Convergence from distant starts may fail if μ is too small initially, hence the adaptive damping. $\square$

Solution to Exercise 2. For the Lewis integral with Heston parameters $(\lambda = 2, \theta = 0.04, \nu = 0.5, \rho = -0.7, V_0 = 0.04, T = 0.25)$:

N_q	Max error (bps)	Time (μs)
16	~ 0.8	5
32	~ 0.02	10
64	$\sim 10^{-5}$	20
128	$\sim 10^{-9}$	40

Gauss–Laguerre with $N_q = 32$ gives sub-basis-point accuracy, adequate for calibration. For research and verification, $N_q = 64$ provides effectively exact results. Going beyond 128 points provides no benefit due to floating-point limitations. $\square$

Solution to Exercise 3. The Rastrigin function in d dimensions: $f(\mathbf{x}) = 10d + \sum_{i=1}^d [x_i^2 - 10 \cos(2\pi x_i)]$ has the global minimum at $\mathbf{x}^* = \mathbf{0}$ with $f^* = 0$, and $\sim 11^d$ local minima. DE parameters: population $N_p = 4d + 10$, $F = 0.8$, $CR = 0.9$, bounds $[-5.12, 5.12]^d$. For $d = 5$, DE typically finds $|f(\hat{\mathbf{x}})| < 10^{-6}$ in 200–500 generations. Key observations:

- Increasing F (mutation factor) improves exploration but slows convergence near the optimum.
- High CR (crossover rate) is needed for multimodal functions with strong variable interactions.
- Population size $< 4d$ frequently stagnates at local minima.

□

4. **(New)** Implement a hybrid DE + LM calibration strategy: use DE for 50 generations to find a good initial guess, then switch to LM for fine-tuning. Compare wall-clock time versus pure DE and pure LM from random starts. What is a good switching criterion?

Solution to Exercise 4. The hybrid strategy works as follows:

1. Run DE for a fixed number of generations (e.g., 50) or until population variance falls below a threshold ϵ_{pop} .
2. Extract the best individual from the DE population as the initial guess for LM.
3. Run LM until $\|\Delta\theta\| < \epsilon$ or the MSE stops decreasing.

A good switching criterion is: switch when the *relative* improvement per DE generation drops below 0.1%, i.e., $|f_{\text{best}}^{(g)} - f_{\text{best}}^{(g-5)}| / (5 \cdot f_{\text{best}}^{(g)}) < 0.001$. This indicates DE has found the basin of the global minimum and LM can converge quadratically from there. Typical results for Heston calibration: pure DE: 2–5s, pure LM (best of 5 random starts): 0.5–2s (risk of local min), hybrid: 0.3–1s with reliable global convergence. □

12 Summary and Connections to WP2

This primer has covered the mathematical foundations needed for WP2:

Topic	Section	WP2 Application
Fractional Brownian motion	§2	Roughness of volatility
Volterra equations	§3	Rough Heston dynamics
Classical stochastic vol.	§4	Heston baseline
Lewis formula	§5	Option pricing
Rough Heston model	§6	Calibration
Implied volatility	§7	Surface fitting
Path signatures	§8	Feature engineering
Tensor algebra	§9	Signature computation
Volatility trading	§10	Strategy construction
Calibration methods	§11	Model fitting

Key Insight

The central thesis of WP2 is that combining *rough volatility models* (for accurate pricing) with *path signature features* (for alpha generation) produces a volatility arbitrage strategy that significantly outperforms classical approaches. This primer provides all the mathematical tools needed to understand both components.

References

- [1] Bayer, C., Friz, P., Gatheral, J. (2016). *Pricing under rough volatility*. Quantitative Finance, 16(6), 887–904.
- [2] Chen, K.T. (1957). *Integration of paths, geometric invariants and a generalized Baker–Hausdorff formula*. Annals of Mathematics, 65(1), 163–178.
- [3] El Euch, O., Rosenbaum, M. (2019). *The characteristic function of rough Heston models*. Mathematical Finance, 29(1), 3–38.
- [4] Gatheral, J., Jaisson, T., Rosenbaum, M. (2018). *Volatility is rough*. Quantitative Finance, 18(6), 933–949.
- [5] Hambly, B., Lyons, T. (2010). *Uniqueness for the signature of a path of bounded variation and the reduced path group*. Annals of Mathematics, 171(1), 109–167.
- [6] Heston, S.L. (1993). *A closed-form solution for options with stochastic volatility with applications to bond and currency options*. Review of Financial Studies, 6(2), 327–343.
- [7] Kiraly, F.J., Oberhauser, H. (2019). *Kernels for sequentially ordered data*. Journal of Machine Learning Research, 20(31), 1–45.
- [8] Lewis, A. (2001). *A simple option formula for general jump-diffusion and other exponential Levy processes*. Envision Financial Syst. and OptionCity.net.
- [9] Mandelbrot, B.B., Van Ness, J.W. (1968). *Fractional Brownian motions, fractional noises and applications*. SIAM Review, 10(4), 422–437.
- [10] Ree, R. (1958). *Lie elements and an algebra associated with shuffles*. Annals of Mathematics, 68(2), 210–220.