

Arbitrage-Free Option Price Surfaces via Chebyshev Tensor Bases and a Hamiltonian Fog Post-Fit

Robert Jenkinson Álvarez

December 1, 2025

Abstract

We study the construction of arbitrage-free option price surfaces from noisy bid-ask quotes across strike and maturity. Our starting point is a Chebyshev representation of the call price surface on a warped log-moneyness/maturity rectangle, together with linear sampling and no-arbitrage operators acting on a collocation grid. Static no-arbitrage requirements are enforced as linear inequalities, while the surface is fitted directly to prices via a coverage-seeking quadratic objective that trades off squared band misfit against spectral and transport-inspired regularisation of the Chebyshev coefficients. This yields a strictly convex quadratic program in the modal coefficients, solvable at practical scales with off-the-shelf solvers (OSQP).

On top of the global backbone, we introduce a local post-fit layer based on a discrete fog of risk-neutral densities on a three-dimensional lattice (m, τ, u) and an associated Hamiltonian-type energy. On each patch of the (m, τ) plane, the fog variables are coupled to a nodal price field obtained from the baseline surface, yielding a joint convex optimisation problem that reweights noisy quotes and applies noise-aware local corrections while preserving global static no-arbitrage and locality.

The method is designed such that for equity options panels, the combined procedure achieves high inside-spread coverage in stable regimes (in calm years, 98 – 99% of quotes are priced inside the bid-ask intervals) and low rates of static no-arbitrage violations (below 1%). In stressed periods, the fog layer provides a mechanism for controlled leakage outside the band: when local quotes are mutually inconsistent or unusually noisy, the optimiser allocates fog mass outside the bid-ask tube and justifies small out-of-band deviations of the post-fit surface, while preserving a globally arbitrage-free and well-regularised description of the option surface.

CONTENTS

1	Introduction	4
2	Market coordinates, targets, and no-arbitrage axioms	6
3	Warped tensor Chebyshev basis and design matrices	7
3.1	Why normalise to $[-1, 1]^2$ and why Chebyshev?	7
3.2	Coordinate warps (endpoint preserving)	7
3.3	Chebyshev polynomials on $[-1, 1]$	8
3.4	Tensor-product basis for the surface	8
3.5	Design matrices at arbitrary points	8
3.6	Derivative design blocks via the chain rule	9
4	No-arbitrage operators on a collocation grid	12
4.1	Collocation grid and evaluation	12
4.2	Strike-space operators (monotonicity and convexity)	12
4.3	Calendar derivative at fixed strike	13
4.4	Price map and bound operators	13
4.5	Row scaling and a single no-arb weight	13
4.6	Summary (operators used in the optimiser)	16
5	Coverage-seeking data term with bid-ask geometry	17
5.1	Loss components and their roles	17
5.2	Quadratic-program form via auxiliary slacks	18
5.3	Weights, units, and invariance	20
5.4	Binned variant (optional)	21
5.5	Feasibility and the role of μ	21
5.6	What is used in this paper (precise choices)	21
6	Ridge, spectral geometry, and transport regularisation	22
6.1	Spectral ridge (modal energy control)	22
6.2	Λ -module: price-invariant reparameterisation	23
6.3	DW-module: discrete transport (H^{-1}) smoothing of density	25
6.4	RN-module: near-maturity residual and calendar flattening	27
6.5	Ω -module: high-frequency taper and commutator hook	29
6.6	Summary of fixed choices used	31

7	No-arbitrage constraints and soft penalties	32
7.1	Penalty definitions (soft versions of the shape constraints)	32
7.2	Row scaling and invariance	33
7.3	Compatibility with the Λ -module	33
7.4	Binned variant (optional)	33
8	The convex program	34
8.1	Slack QP (standard form)	34
8.2	Convexity, existence, and uniqueness	34
8.3	Invariance and scaling	37
8.4	Solution procedure (used)	37
9	Scaling, schedules, and the μ-controller	38
9.1	Row scaling (summary)	38
9.2	Short schedule for λ_{NA}	38
9.3	The μ -controller (coverage target)	38
9.4	Practical notes	39
10	Short-maturity remedy	40
10.1	Boundary anchoring and calendar flattening on Γ_{0+}	40
10.2	Frequency truncation: Ω taper near the boundary	42
10.3	Transport damping: H^{-1} smoothing of density at short maturities	43
10.4	Convexity and invariance	43
10.5	Practical choices and interaction with scaling	43
11	Diagnostics and Implementation	44
11.1	Structure monitors	44
11.2	Implementation notes	44
12	Hamiltonian Fog Post-Fit in Price Space	45
12.1	Baseline grid surface and quotes	45
12.2	Badness map and patch decomposition	46
12.3	Discrete 3D fog on (m, τ, u)	51
12.3.1	3D lattice and fog variables	51
12.3.2	2D noise marginal on the patch	52
12.3.3	Schematic geometry	52
12.4	Patch-level price field and static no-arbitrage	53

12.4.1	Interior price variables and assembly map	53
12.4.2	Global discrete static no-arbitrage on the grid	54
12.4.3	Patch-level feasible set and its geometry	60
12.5	Hamiltonian energy on the fog	62
12.5.1	Discrete 3D graph and difference operators	62
12.5.2	Kinetic energy and graph Laplacian	63
12.5.3	Potential energy and band-aware penalisation	64
12.5.4	Hamiltonian energy and basic properties	66
12.6	Noise-aware band term via the fog	67
12.6.1	Fog mass outside the band at a quote	67
12.6.2	Band misfit and noise-aware penalty	68
12.6.3	Convexity of the noise-aware band term	69
12.7	Surface energy and closeness to the baseline	72
12.7.1	Discrete density operator and patch restriction	72
12.7.2	Surface density energy	73
12.7.3	Closeness to the baseline	74
12.8	Patch-level post-fit optimisation problem	74
12.8.1	Fog feasible set and quote index set	75
12.8.2	Patch energy functional	76
12.8.3	Patch-level post-fit problem: existence and uniqueness	77
12.9	Global post-fit across patches and dates	79
12.9.1	Patch decomposition and compatibility with no-arbitrage stencils	79
12.9.2	Global post-fit surface on a fixed date	80
12.9.3	Global static no-arbitrage and locality	80
13	Conclusion and outlook	82

1. INTRODUCTION

Liquid option markets require a smooth, stable and *arbitrage-free* surface of prices or implied volatilities over strike and maturity. Such a surface underpins marking, risk management and model calibration, and feeds directly into trading and hedging decisions. In practice, the raw quote grid is sparse, noisy and often inconsistent with the static no-arbitrage conditions implied by absence of butterfly and calendar spreads. Production systems therefore interpolate and regularise the observed quotes into a dense surface subject to no-arbitrage constraints.

There is substantial literature on arbitrage-free surface construction. Parametric approaches such as SVI and its extensions impose functional forms on implied volatility and derive analytical sufficient conditions for absence of static arbitrage. Alternatively, nonparametric smoothing methods reconstruct prices or volatilities on a grid while enforcing no-arbitrage inequalities either as hard constraints or via penalisation. These methods have been successfully deployed in practice, but there remains a trade-off between fidelity to the bid-ask quotes, strict enforcement of no-arbitrage on a dense grid, and computational cost on large universes of names and dates.

This paper contributes a practical convex-programming formulation for option price surfaces that aims to balance these considerations, together with a local geometric post-fit layer that explicitly models quote noise on difficult regions of the surface. The key ingredients are:

- a global Chebyshev representation of the price surface on a warped log-moneyness / maturity rectangle, which provides high approximation power;
- linear operators that encode static no-arbitrage constraints on a dense collocation grid (monotonicity in strike, convexity in strike, calendar monotonicity and simple bounds), so that absence of butterfly and calendar arbitrage is enforced directly in price space;
- a coverage-seeking quadratic objective aligned to the bid–ask geometry, augmented by spectral and transport-inspired regularisers (ridge in the Chebyshev coefficients, discrete H^{-1} smoothing of the density, short-end anchoring and frequency tapering) that stabilise the fit while preserving convexity;
- a patch-wise post-fit in price space, built on a discrete “fog” of risk-neutral densities on a three-dimensional lattice (m, τ, u) and a Hamiltonian-type energy on that fog, which yields a convex, noise-aware local correction of the baseline surface on problematic regions while preserving global static no-arbitrage.

The resulting baseline surface is obtained as the solution of a single medium-scale QP with sparse structure, solvable reliably with off-the-shelf solvers, and tuned to reach high within-band coverage and low static no-arbitrage violation rates on a dense grid. The discrete Hamiltonian fog layer appears as a second, fully convex post-fit stage defined on local patches in (m, τ) ; it is implemented in this paper in a finite-dimensional setting (Chapter 12) and used to refine the baseline surface in stressed regimes.

The rest of the paper is organised as follows. Chapter 2 defines the market coordinates, targets and static no-arbitrage axioms. Chapters 3-4 build the warped Chebyshev tensor basis and the no-arbitrage operators on a collocation grid. Chapters 5-7 specify the coverage-seeking data term and the soft no-arbitrage penalties, and Chapter 8 assembles the global QP and discusses convexity and solution. Chapters 6 and 10 develop the spectral and transport-inspired quadratic regularisers, and Chapter 11 collects structure monitors that diagnose stability. Chapter 12 then introduces the patch-wise Hamiltonian fog post-fit in price space, formulated as a joint convex optimisation in the nodal surface and fog variables. Chapter 13 concludes and outlines a continuum version of the fog/Hamiltonian geometry, which is deferred to a separate theoretical follow-up paper.

2. MARKET COORDINATES, TARGETS, AND NO-ARBITRAGE AXIOMS

Let t be a trading date and $F_t(T)$ denote the discount adjusted forward for maturity time T . Set $\tau = T - t > 0$, the time to maturity. We work in forward discounted prices:

$$C_f(K, \tau) := e^{r(\tau)\tau} C(K, \tau) \quad \text{and} \quad m := \log \frac{K}{F_t(\tau)}.$$

Throughout, we fit a surface $C_f(m, \tau)$ from quoted calls. Puts follow from put-call parity. These are sufficient conditions for Static no-arbitrage for calls (for a.e. m, τ):

$$\partial_m C_f(m, \tau) \leq 0, \tag{2.1}$$

$$\partial_{mm} C_f(m, \tau) \geq 0, \tag{2.2}$$

$$\partial_\tau C_f(m, \tau)|_K \geq 0. \tag{2.3}$$

Bounds: $0 \leq C_f(m, \tau) \leq F_t(\tau)$ and $C_f(m, 0^+) = (F_t(0) - K)^+$.

Throughout this section we tacitly assume enough regularity for the continuum derivatives in (2.1)–(2.3) to be well defined on the compact rectangle where we approximate the surface. In particular, on the box $[m_{\min}, m_{\max}] \times [\tau_{\min}, \tau_{\max}]$ used in Section 3, we work under

$$C_f \in C^{2,1}([m_{\min}, m_{\max}] \times [\tau_{\min}, \tau_{\max}]), \quad F \in C^1([\tau_{\min}, \tau_{\max}]),$$

so that $\partial_m C_f$, $\partial_{mm} C_f$, $(\partial_\tau C_f)|_K$ and $\frac{d}{d\tau} \log F(\tau)$ all exist and are continuous. The later discrete QP only uses linear operators on a grid, but these smoothness conditions provide the natural continuum axioms behind (2.1)–(2.3).

Remark 1 (Calendar derivative at fixed strike). Since the basis uses (m, τ) , the calendar derivative at fixed K becomes $\partial_\tau C_f|_K = \partial_\tau C_f + \partial_m C_f \cdot \partial_\tau m|_K = \partial_\tau C_f - \partial_m C_f \cdot (r_{cc}(\tau) + \tau r'_{cc}(\tau))$. This is implemented exactly in the operators below.

Notation

Symbol	Meaning
K	Strike
τ	Time to maturity $T - t$
$F_t(\tau)$	Forward (discount-adjusted) underlying for τ
m	Log-moneyness $\log(K/F_t(\tau))$
C_f	Forward-discounted call price
ρ	Risk-neutral density $\partial_{KK} C_f$
A, A_m, A_{mm}, A_τ	Design/derivative blocks in coefficient space

3. WARPED TENSOR CHEBYSHEV BASIS AND DESIGN MATRICES

This chapter builds the approximation space and its derivative blocks.

3.1 Why normalise to $[-1, 1]^2$ and why Chebyshev?

The problem is to approximate a continuous surface $C_f(m, \tau)$ on a compact rectangle $[m_{\min}, m_{\max}] \times [\tau_{\min}, \tau_{\max}]$. On a compact interval, polynomials are dense (Weierstrass), and Chebyshev polynomials are numerically preferred because:

- (i) they minimise Runge oscillations on $[-1, 1]$,
- (ii) they admit stable three-term recurrences and Clenshaw evaluation,
- (iii) they possess explicit derivative identities useful for Greeks.

We therefore map each axis to $[-1, 1]$ and expand in a tensor-product Chebyshev basis.

3.2 Coordinate warps (endpoint preserving)

Let $[m_{\min}, m_{\max}]$ and $[\tau_{\min}, \tau_{\max}]$ be robust, date-adaptive intervals.

Define the warps

$$x = \Phi_m(m) \in [-1, 1], \quad y = \Phi_\tau(\tau) \in [-1, 1],$$

so that the interval endpoints map exactly to ± 1 .

Log-moneyness warp (asinh). Let c_m be a centre (e.g. the liquidity-weighted median of m), and $\lambda_m > 0$ a tail-compression parameter. Set

$$\phi_{m,\pm} := \operatorname{asinh}(\lambda_m(m_{\max/\min} - c_m)), \quad W_m := \phi_{m,+} - \phi_{m,-},$$

$$\Phi_m(m) = \frac{2}{W_m} \left(\operatorname{asinh}(\lambda_m(m - c_m)) - \phi_{m,-} \right) - 1.$$

Then $\Phi_m(m_{\min}) = -1$, $\Phi_m(m_{\max}) = +1$. Derivatives (by the chain rule) are

$$\Phi'_m(m) = \frac{2\lambda_m}{W_m} \frac{1}{\sqrt{1 + \lambda_m^2(m - c_m)^2}}, \quad \Phi''_m(m) = -\frac{2\lambda_m^3}{W_m} \frac{(m - c_m)}{(1 + \lambda_m^2(m - c_m)^2)^{3/2}}.$$

Maturity warp (square-root). Let $\Delta_\tau := \tau_{\max} - \tau_{\min} > 0$ and $s(\tau) := (\tau - \tau_{\min})/\Delta_\tau \in [0, 1]$. Set

$$\boxed{\Phi_\tau(\tau) = 2\sqrt{s(\tau)} - 1}, \quad \Phi'_\tau(\tau) = \frac{1}{\sqrt{\Delta_\tau}(\tau - \tau_{\min})} \quad (\text{finite if } \tau > \tau_{\min}).$$

The square-root allocates higher resolution near short maturities. (If $\tau_{\min}$ is very close to 0, a small positive floor avoids the endpoint singularity in Φ'_τ .)

Remark 2 (Why these warps). The asinh warp allocates more resolution near $m \approx c_m$ (ATM) while compressing deep wings; the square-root warp concentrates basis power near short maturities where curvature in τ is largest. Both preserve endpoints and expose simple chain-rule factors for derivatives.

3.3 Chebyshev polynomials on $[-1, 1]$

For $x \in [-1, 1]$, the Chebyshev polynomials of the first kind are

$$T_k(x) = \cos(k \arccos x), \quad T_0 = 1, \quad T_1 = x, \quad T_{k+1} = 2x T_k - T_{k-1}.$$

The derivatives needed later are available in closed form:

$$T'_k(x) = k U_{k-1}(x),$$

and, for $|x| < 1$,

$$(1 - x^2) T''_k(x) - x T'_k(x) + k^2 T_k(x) = 0 \Rightarrow \boxed{T''_k(x) = \frac{x k U_{k-1}(x) - k^2 T_k(x)}{1 - x^2}},$$

where U_n are Chebyshev polynomials of the second kind, which are defined recursively ($U_0(x) = 1$, $U_1(x) = 2x$, $U_{n+1}(x) = 2x \cdot U_n(x) - U_{n-1}(x)$). In practice, we evaluate T_k, U_{k-1} stably via Clenshaw recurrences.

3.4 Tensor-product basis for the surface

Let $K, L \in \mathbb{N}$ be polynomial degrees in m and τ , and define the coefficient array $a = \{a_{k\ell}\}_{k=0,\dots,K; \ell=0,\dots,L}$. We approximate

$$\boxed{C_f(m, \tau) = \sum_{k=0}^K \sum_{\ell=0}^L a_{k\ell} T_k(\Phi_m(m)) T_\ell(\Phi_\tau(\tau))}.$$

Stacking a into a vector in $\mathbb{R}^P$ with $P = (K+1)(L+1)$ yields a linear map from coefficients to prices.

3.5 Design matrices at arbitrary points

Given data points $\{(m_i, \tau_i)\}_{i=1}^N$, set $x_i = \Phi_m(m_i)$, $y_i = \Phi_\tau(\tau_i)$. Define the following row vectors

$$\begin{aligned} \mathbf{t}(x_i) &= [T_0(x_i), \dots, T_K(x_i)], \\ \mathbf{s}(y_i) &= [T_0(y_i), \dots, T_L(y_i)]. \end{aligned}$$

The pair (k, ℓ) defines the index of the column (where $k \in \{0, \dots, K\}$ and $\ell \in \{0, \dots, L\}$).

Definition 1. The (price) design matrix $A \in \mathbb{R}^{N \times P}$ is defined as

$$A_{i,(k,\ell)} = T_k(x_i) T_\ell(y_i).$$

Equivalently, if $\Phi_K \in \mathbb{R}^{N \times (K+1)}$ stacks $\mathbf{t}(x_i)$ and $\Phi_L \in \mathbb{R}^{N \times (L+1)}$ stacks $\mathbf{s}(y_i)$, then A is the row-wise Khatri–Rao product $A = \Phi_K \odot \Phi_L$; for grid evaluations, Kronecker structure ($\Phi_L \otimes \Phi_K$) can be exploited.

3.6 Derivative design blocks via the chain rule

Proposition 1 (Closed-form derivative design blocks). *Fix integers $K, L \geq 0$ and let $P = (K+1)(L+1)$. For each data point (m_i, τ_i) define*

$$x_i := \Phi_m(m_i), \quad y_i := \Phi_\tau(\tau_i),$$

where $\Phi_m \in C^2$ and $\Phi_\tau \in C^1$ on their domains. Let the price design matrix $A \in \mathbb{R}^{N \times P}$ be

$$A_{i,(k,\ell)} = T_k(x_i) T_\ell(y_i), \quad 0 \leq k \leq K, \quad 0 \leq \ell \leq L,$$

with any fixed stacking $(k, \ell) \mapsto (k, \ell)$ -column. Define the “inner-variable” derivative matrices by

$$\begin{aligned} (\partial_x A)_{i,(k,\ell)} &:= T'_k(x_i) T_\ell(y_i), & (\partial_{xx} A)_{i,(k,\ell)} &:= T''_k(x_i) T_\ell(y_i), \\ (\partial_y A)_{i,(k,\ell)} &:= T_k(x_i) T'_\ell(y_i). \end{aligned}$$

For any coefficient vector $a \in \mathbb{R}^P$, consider the model values

$$\hat{C}_i = (Aa)_i = \sum_{k,\ell} a_{k\ell} T_k(x_i) T_\ell(y_i).$$

Then the vectors of physical derivatives evaluated at the same points are linear images of a :

$$(\partial_m \hat{C})_i = (A_m a)_i, \quad (\partial_{mm} \hat{C})_i = (A_{mm} a)_i, \quad (\partial_\tau \hat{C})_i = (A_\tau a)_i,$$

where the derivative design blocks are

$$A_m = \text{diag}(\Phi'_m(m)) \partial_x A,$$

$$A_{mm} = \text{diag}((\Phi'_m)^2) \partial_{xx} A + \text{diag}(\Phi''_m(m)) \partial_x A,$$

$$A_\tau = \text{diag}(\Phi'_\tau(\tau)) \partial_y A.$$

Proof. All statements follow from linearity and the chain rule, applied row-wise.

Setup: Write the one-dimensional warped basis functions

$$\phi_k(m) := T_k(\Phi_m(m)), \quad \psi_\ell(\tau) := T_\ell(\Phi_\tau(\tau)).$$

Then the model at (m_i, τ_i) is

$$\hat{C}_i = \sum_{k=0}^K \sum_{\ell=0}^L a_{k\ell} \phi_k(m_i) \psi_\ell(\tau_i) = \sum_{k,\ell} a_{k\ell} T_k(x_i) T_\ell(y_i).$$

By construction, the i -th row of A consists of the basis values $\{T_k(x_i) T_\ell(y_i)\}_{k,\ell}$, so $\hat{C} = Aa$.

First derivative in m : Differentiating $\phi_k(m) = T_k(\Phi_m(m))$ with respect to m , by the chain rule we obtain the following:

$$\frac{d}{dm} \phi_k(m) = T'_k(\Phi_m(m)) \Phi'_m(m).$$

Fixing τ , the derivative of each product is

$$\frac{\partial}{\partial m} [\phi_k(m) \psi_\ell(\tau)] = T'_k(\Phi_m(m)) \Phi'_m(m) \cdot T_\ell(\Phi_\tau(\tau)).$$

Evaluating at (m_i, τ_i) and summing over (k, ℓ) :

$$(\partial_m \widehat{C})_i = \Phi'_m(m_i) \sum_{k, \ell} a_{k\ell} T'_k(x_i) T_\ell(y_i) = (\Phi'_m(m_i) \cdot (\partial_x A a)_i).$$

Taking the scalar formula for each i and writing as a vector-matrix equation yields

$$\partial_m \widehat{C} = \text{diag}(\Phi'_m(m)) \partial_x A a,$$

where $A_m = \text{diag}(\Phi'_m(m)) \partial_x A$.

Second derivative in m : Differentiate once more, using the product rule and chain rule:

$$\begin{aligned} \frac{d^2}{dm^2} \phi_k(m) &= \frac{d}{dm} (T'_k(\Phi_m(m)) \Phi'_m(m)) \\ &= T''_k(\Phi_m(m)) (\Phi'_m(m))^2 + T'_k(\Phi_m(m)) \Phi''_m(m). \end{aligned}$$

Therefore,

$$\frac{\partial^2}{\partial m^2} [\phi_k(m) \psi_\ell(\tau)] = (T''_k(\Phi_m(m)) (\Phi'_m(m))^2 + T'_k(\Phi_m(m)) \Phi''_m(m)) T_\ell(\Phi_\tau(\tau)).$$

Evaluating at (m_i, τ_i) and summing,

$$(\partial_{mm} \widehat{C})_i = (\Phi'_m(m_i))^2 \sum_{k, \ell} a_{k\ell} T''_k(x_i) T_\ell(y_i) + \Phi''_m(m_i) \sum_{k, \ell} a_{k\ell} T'_k(x_i) T_\ell(y_i).$$

In matrix form,

$$\partial_{mm} \widehat{C} = \text{diag}((\Phi'_m(m))^2) \partial_{xx} A a + \text{diag}(\Phi''_m(m)) \partial_x A a,$$

so $A_{mm} = \text{diag}((\Phi'_m(m))^2) \partial_{xx} A + \text{diag}(\Phi''_m(m)) \partial_x A$.

First derivative in τ : Analogously, with $y = \Phi_\tau(\tau)$,

$$\frac{d}{d\tau} \psi_\ell(\tau) = T'_\ell(\Phi_\tau(\tau)) \Phi'_\tau(\tau),$$

Fixing m ,

$$\frac{\partial}{\partial \tau} [\phi_k(m) \psi_\ell(\tau)] = T_k(\Phi_m(m)) T'_\ell(\Phi_\tau(\tau)) \Phi'_\tau(\tau).$$

Evaluating at (m_i, τ_i) and summing over (k, ℓ) yields:

$$(\partial_\tau \widehat{C})_i = \Phi'_\tau(\tau_i) \sum_{k, \ell} a_{k\ell} T_k(x_i) T'_\ell(y_i) = (\Phi'_\tau(\tau_i) \cdot (\partial_y A a)_i),$$

Taking the scalar formula for each i and writing as a vector-matrix equation yields

$$\partial_\tau \widehat{C} = \text{diag}(\Phi'_\tau(\tau)) \partial_y A a$$

where $A_\tau = \text{diag}(\Phi'_\tau(\tau)) \partial_y A$.

Conclusion. In each case the derivative vector equals a fixed matrix (depending only on the warps and basis) times a , establishing the stated formulas. $\square$

Remark 3. Precompute $\{T_k(x_i), U_{k-1}(x_i)\}_{k \leq K}$ and $\{T_\ell(y_i), U_{\ell-1}(y_i)\}_{\ell \leq L}$ via Clenshaw recurrences. Then obtain $T'_k(x_i) = k U_{k-1}(x_i)$ and, for interior points $|x_i| < 1$, $T''_k(x_i) = (x_i k U_{k-1}(x_i) - k^2 T_k(x_i)) / (1 - x_i^2)$. At the Chebyshev–Lobatto endpoints $x = \pm 1$ the denominator $1 - x^2$ vanishes, but T_k is a polynomial so $T''_k(x)$ exists and is finite there. In practice we define $T''_k(\pm 1)$ by continuity (or via the closed forms $T''_k(1) = \frac{k^2(k^2-1)}{3}$ and $T''_k(-1) = (-1)^k \frac{k^2(k^2-1)}{3}$ for $k \geq 2$) and use these values whenever $|1 - x_i^2|$ is numerically small. With this convention, all entries of $\partial_{xx}A$ are well defined and the assembly of $A, \partial_x A, \partial_{xx}A, \partial_y A$ uses only closed-form expressions, with no numerical differencing.

4. NO-ARBITRAGE OPERATORS ON A COLLOCATION GRID

From the previous chapter we have for any set of evaluation points (m, τ) , the price design matrix A and the derivative blocks A_m, A_{mm}, A_τ are defined by

$$\begin{aligned} (Aa)(m, \tau) &= C_f(m, \tau), & (A_m a)(m, \tau) &= \partial_m C_f, \\ (A_{mm} a)(m, \tau) &= \partial_{mm} C_f, & (A_\tau a)(m, \tau) &= \partial_\tau C_f. \end{aligned}$$

All maps are linear in a and are computed pointwise via the chain rule.

4.1 Collocation grid and evaluation

Let $\{(m_g, \tau_g)\}_{g=1}^G$ be a fixed collocation grid used to test the no-arbitrage shape conditions (Chebyshev nodes in m , uniform in τ are a robust choice). On this grid define the forward (no-arbitrage price of receiving one unit of the underlying at time $(t + \tau)$) and strike (rearrangement from the definition of m)

$$F_g := F_t(\tau_g), \quad K_g := F_g e^{m_g}.$$

We evaluate the same derivative blocks on the grid (rather than on $\{(m_i, \tau_i)\}_{i=1}^N$, we evaluate on $\{(m_g, \tau_g)\}_{g=1}^G$); keeping the symbols A, A_m, A_{mm}, A_τ for the $G \times P$ versions where the g -th row corresponds to (m_g, τ_g) .

Remark 4 (Why a separate grid). Quotes can be sparse or clustered. A collocation grid decouples shape testing from where data happen to lie and gives uniform control of violations over the rectangle in (m, τ) .

4.2 Strike-space operators (monotonicity and convexity)

Static no-arb for *calls* requires $\partial_K C_f \leq 0$ and $\partial_{KK} C_f \geq 0$ at fixed τ . However, our derivative blocks are in the m coordinate, where $m = \ln(K/F(\tau))$. At fixed τ ,

$$\frac{\partial m}{\partial K} = \frac{1}{K}, \quad \frac{\partial^2 m}{\partial K^2} = -\frac{1}{K^2}.$$

For any smooth $f(m, \tau)$, by the chain rule

$$\left. \frac{\partial f}{\partial K} \right|_\tau = \frac{1}{K} f_m, \quad \left. \frac{\partial^2 f}{\partial K^2} \right|_\tau = \frac{1}{K^2} (f_{mm} - f_m).$$

Apply this with $f = C_f$ row-wise on the grid, replacing f_m and f_{mm} by $A_m a$ and $A_{mm} a$.

$$\begin{aligned} \partial_K C_f(K_g, \tau_g) &= \frac{1}{K_g} \partial_m C_f(m_g, \tau_g) = \frac{1}{K_g} (A_m a)_g \\ \partial_{KK} C_f(K_g, \tau_g) &= -\frac{1}{K_g^2} \partial_m C_f(m_g, \tau_g) + \frac{1}{K_g^2} \partial_{mm} C_f(m_g, \tau_g) = \frac{(A_{mm} a)_g - (A_m a)_g}{K_g^2} \end{aligned}$$

This yields the *linear operators* that map coefficients a to strike derivatives:

$$\text{Monotonicity in strike:} \quad A_K = \text{diag}(K_g)^{-1} A_m, \quad (4.1)$$

$$\text{Convexity in strike:} \quad A_{KK} = \text{diag}(K_g)^{-2} (A_{mm} - A_m). \quad (4.2)$$

Thus, $(A_K a)_g = \partial_K C_f(K_g, \tau_g)$ and $(A_{KK} a)_g = \partial_{KK} C_f(K_g, \tau_g)$.

4.3 Calendar derivative at fixed strike

Calendar no-arbitrage requires $\partial_\tau C_f|_K \geq 0$. The block A_τ computes $\partial_\tau C_f$ at fixed m ; to switch to fixed K use the following relation

$$\left(\partial_\tau C_f\right)_K = \left(\partial_\tau C_f\right)_m + \left(\partial_m C_f\right)\left(\partial_\tau m\right)_K.$$

With $m = \log(K/F(\tau))$ and K fixed,

$$\left(\partial_\tau m\right)_K = -\frac{d}{d\tau} \log F(\tau).$$

Two equivalent parameterisations of F give:

(i) *General form.* Let $\rho(\tau) := \frac{d}{d\tau} \log F(\tau)$. Then

$$A_{\tau|K} = A_\tau - \text{diag}(\rho(\tau_g)) A_m.$$

(ii) *Report convention.* If $\log F(\tau) = \tau r_{cc}(\tau)$ (net continuously compounded carry), then $\rho(\tau) = r_{cc}(\tau) + \tau r'_{cc}(\tau)$ and

$$A_{\tau|K} = A_\tau + \text{diag}(-r_{cc}(\tau_g) - \tau_g r'_{cc}(\tau_g)) A_m. \quad (4.3)$$

Remark 5 (Sanity checks). If F is flat (zero carry), then $\rho \equiv 0$ and $A_{\tau|K} = A_\tau$. If carry is constant r , then $\rho \equiv r$ and $A_{\tau|K} = A_\tau - r A_m$.

4.4 Price map and bound operators

Price non-negativity and upper bounds by the forward read on the grid as

$$0 \leq (Aa)_g \leq F_g, \quad g = 1, \dots, G.$$

Simply write $A_{\text{price}} = A$ and use the known vector $F = (F_g)_g$ when imposing hard constraints or soft penalties for violations.

4.5 Row scaling and a single no-arb weight

The three no-arbitrage defect maps have different natural magnitudes and units:

$$\begin{aligned} A_K a & \quad (\text{“price per strike”}), & -A_{KK} a & \quad (\text{“price per strike}^2\text{”}), \\ -A_{\tau|K} a & \quad (\text{“price per time”}). \end{aligned}$$

If a single penalty weight λ_{NA} is applied to all three without normalisation, the largest magnitude block dominates and the others become numerically inert. Therefore, normalise each block by a positive scalar so that a single λ_{NA} can control them comparably.

Blocks to be normalised: On the collocation grid (size $G \times P$), set

$$B_1 := A_K, \quad B_2 := -A_{KK}, \quad B_3 := -A_{\tau|K}.$$

Robust block scales: For each $j \in \{1, 2, 3\}$, compute Euclidean row ℓ_2 norms

$$r_g^{(j)} := \|(B_j)_{g,:}\|_2 = \sqrt{\sum_{p=1}^P (B_j)_{gp}^2}, \quad g = 1, \dots, G$$

Sorting the list in ascending order, pick the value s_j below which 95% of the $r_g^{(j)}$ fall.

Definition 2. The robust scale is defined as

$$s_j := q_{0.95}(\{r_g^{(j)} : g = 1, \dots, G\}),$$

(Other robust choices are possible; $q_{0.95}$ balances outliers vs. typical rows.)

Scaled blocks and unified weight: Define the scaled operators

$$\tilde{B}_j := \frac{1}{s_j} B_j, \quad j = 1, 2, 3.$$

Using a single λ_{NA} for all three terms, the soft no-arbitrage penalty becomes

$$\frac{\lambda_{\text{NA}}}{2} \sum_{j=1}^3 \|(\tilde{B}_j a)_+\|_2^2 = \frac{1}{2} \sum_{j=1}^3 \underbrace{\left(\frac{\lambda_{\text{NA}}}{s_j^2}\right)}_{\text{effective weight for block } j} \| (B_j a)_+ \|_2^2,$$

so that the typical (p95) row magnitude of each block is ≈ 1 and one knob λ_{NA} moves all three violation shares on a comparable scale.

Proposition 2 (Invariance of hard constraints under positive scaling). *Let D be any positive diagonal matrix (in particular $D = \alpha I$ with $\alpha > 0$). Then, for any $B \in \mathbb{R}^{G \times P}$ and any $a \in \mathbb{R}^P$,*

$$Ba \leq 0 \iff (DB)a \leq 0.$$

Hence replacing B_j by $\tilde{B}_j = \frac{1}{s_j} B_j$ leaves the hard no-arbitrage feasible set unchanged; only numerical conditioning and relative penalty weights are affected.

Proof. All inequalities are understood componentwise.

Let $D = \text{diag}(d_1, \dots, d_G)$ with $d_i > 0$ for all i . For any $a \in \mathbb{R}^P$ we have

$$(DB)a = D(Ba),$$

so on the i -th component

$$((DB)a)_i = d_i (Ba)_i.$$

($\Rightarrow$) Suppose $Ba \leq 0$. Then for every i ,

$$(Ba)_i \leq 0 \implies d_i (Ba)_i \leq 0$$

because $d_i > 0$. Hence $(DB)a = D(Ba) \leq 0$.

($\Leftarrow$) Conversely, suppose $(DB)a \leq 0$. Then for every i ,

$$d_i(Ba)_i = ((DB)a)_i \leq 0.$$

Since $d_i > 0$, dividing by d_i preserves the inequality sign and yields

$$(Ba)_i \leq 0 \quad \text{for all } i,$$

i.e. $Ba \leq 0$.

Thus $\{a : Ba \leq 0\} = \{a : (DB)a \leq 0\}$, proving the equivalence $Ba \leq 0 \iff (DB)a \leq 0$.

For the final claim, take $D = \frac{1}{s_j}I$ with $s_j > 0$ and B replaced by a given block B_j . Then

$$B_j a \leq 0 \iff \left(\frac{1}{s_j}I B_j\right)a \leq 0 \iff \tilde{B}_j a \leq 0,$$

so replacing B_j by $\tilde{B}_j = \frac{1}{s_j}B_j$ leaves the hard no-arbitrage feasible set $\{a : B_j a \leq 0\}$ (and hence the intersection over all blocks j) unchanged. Only the numerical conditioning of the operators and the effective relative weights in any soft penalties involving B_j are affected. $\square$

Remark 6 (Exact recipe used in this paper).

1. *Where scaling is applied.* We first apply any coefficient reparameterisation U (price-invariant transform), i.e. replace each block by $A_\bullet U$. Scaling is computed and applied to these *post- U* blocks.
2. *Which blocks.* We scale $B_1 = A_K$, $B_2 = -A_{KK}$, $B_3 = -A_{\tau|K}$ by *one scalar per block*: s_K, s_{KK}, s_τ given by the p95 of row ℓ_2 norms on the collocation grid.
3. *How it enters the objective.* The no-arb penalty uses the scaled operators $\tilde{B}_j = B_j/s_j$ with a *single* weight λ_{NA} :

$$\mathcal{P}_{\text{NA}}(a) = \frac{\lambda_{\text{NA}}}{2} \left(\|(\tilde{A}_K a)_+\|_2^2 + \|(-\tilde{A}_{KK} a)_+\|_2^2 + \|(-\tilde{A}_{\tau|K} a)_+\|_2^2 \right).$$

4. *Bounds kept separate.* Price bounds $0 \leq Aa \leq F$ are handled with a separate weight λ_B ; we do not include A in the no-arb scaling group.
5. *Reporting.* Diagnostics/violation shares are computed from the unscaled physical operators $A_K, A_{KK}, A_{\tau|K}$.

Remark 7 (Alternative (not used): row-by-row equalisation). One may also scale *each row* to equalise row influence by taking $D_j = \text{diag}(d_g^{(j)})$ with $d_g^{(j)} = 1/\max(\|(B_j)_{g,:}\|_2, \varepsilon)$ and using $D_j B_j$. This preserves feasibility for hard constraints (Prop. 2) but reweights the grid non-uniformly. We *do not* use this in our main results; we use the block-scalar scheme of Remark 6.

4.6 Summary (operators used in the optimiser)

On the collocation grid, the no-arbitrage conditions become linear maps of a :

$$\begin{aligned} \text{strike monotonicity: } & A_K a \leq 0, \\ \text{strike convexity: } & -A_{KK} a \leq 0, \\ \text{calendar at fixed } K: & -A_{\tau|K} a \leq 0, \\ \text{bounds: } & 0 \leq Aa \leq F. \end{aligned}$$

Enforce these either as hard linear inequalities or as convex quadratic penalties on the positive parts, all while keeping the problem a single QP.

5. COVERAGE-SEEKING DATA TERM WITH BID-ASK GEOMETRY

On date t , let $\{(m_i, \tau_i)\}_{i=1}^N$ be the quote locations, and let

$$b_i := \text{bid}_i, \quad a_i := \text{ask}_i, \quad y_i := \frac{1}{2}(\text{bid}_i + \text{ask}_i),$$

be the forward-discounted band endpoints and mids ($0 \leq b_i \leq a_i$ after standard cleaning). Let $A \in \mathbb{R}^{N \times P}$ be the price design matrix so that $\hat{y}(a) := Aa$ are model prices at the quote points. Set heteroscedastic residual weights

$$w_i = \frac{\text{liq}_i}{\max(a_i - b_i, \varepsilon)^2}, \quad W := \text{diag}(w_1, \dots, w_N),$$

where $\text{liq}_i = 1 + \sqrt{\text{volume}_i} + 0.1\sqrt{\text{open_interest}_i}$ and with a small floor $\varepsilon > 0$.

5.1 Loss components and their roles

Use two convex terms:

$$\mathcal{L}_{\text{fit}}(a) = \underbrace{\frac{1}{2} \|W^{1/2}(Aa - y)\|_2^2}_{\text{within-band centre anchor}} + \underbrace{\mu \sum_{i=1}^N \ell_{\text{band}}((Aa)_i; b_i, a_i)}_{\text{coverage/Slack pricing}}, \quad (5.1)$$

where the *quadratic band hinge* for a scalar $\hat{y}$ and interval $[b, a]$ is

$$\ell_{\text{band}}(\hat{y}; b, a) = \frac{1}{2} (\max\{b - \hat{y}, 0\}^2 + \max\{\hat{y} - a, 0\}^2) = \frac{1}{2} \text{dist}(\hat{y}, [b, a])^2. \quad (5.2)$$

This means that $\ell_{\text{band}}(\hat{y}; b, a) = 0$ iff $\hat{y} \in [b, a]$, and otherwise equals one-half the squared Euclidean distance to the band. The first term in (5.1) selects a point *inside* the band (preferentially near y defined as the mid) whenever that is compatible with the other constraints; the second term is a convex surrogate that drives *coverage* by penalizing exactly the squared violation outside the band.

Lemma 1 (Convexity of band loss and fit objective). *Fix $b \leq a$ and define ℓ_{band} as in (5.2). Then $\hat{y} \mapsto \ell_{\text{band}}(\hat{y}; b, a)$ is a convex function on $\mathbb{R}$. Consequently, for any $\mu \geq 0$, any weight matrix $W \succeq 0$, design matrix $A \in \mathbb{R}^{N \times P}$ and vector $y \in \mathbb{R}^N$, the loss*

$$\mathcal{L}_{\text{fit}}(a) = \frac{1}{2} \|W^{1/2}(Aa - y)\|_2^2 + \mu \sum_{i=1}^N \ell_{\text{band}}((Aa)_i; b_i, a_i)$$

is convex in $a \in \mathbb{R}^P$.

Proof. Write

$$\ell_{\text{band}}(\hat{y}; b, a) = \frac{1}{2} \left(\max\{b - \hat{y}, 0\}^2 + \max\{\hat{y} - a, 0\}^2 \right).$$

Each map $\hat{y} \mapsto b - \hat{y}$ and $\hat{y} \mapsto \hat{y} - a$ is affine, hence convex. The hinge map $t \mapsto \max\{t, 0\}$ is convex as a pointwise maximum of two affine functions (t and 0). Therefore

$$\hat{y} \mapsto \max\{b - \hat{y}, 0\}, \quad \hat{y} \mapsto \max\{\hat{y} - a, 0\}$$

are convex functions. Moreover, both are nonnegative.

The square map $s \mapsto s^2$ is convex and nondecreasing on $[0, \infty)$. The composition of a convex, nondecreasing function with a convex, nonnegative function is convex. Hence

$$\hat{y} \mapsto \max\{b - \hat{y}, 0\}^2, \quad \hat{y} \mapsto \max\{\hat{y} - a, 0\}^2$$

are convex, and so is their sum. Multiplication by $\frac{1}{2} > 0$ preserves convexity, therefore $\ell_{\text{band}}(\cdot; b, a)$ is convex.

For the second claim, the map $a \mapsto Aa - y$ is affine, $W^{1/2}$ is linear, and $f(z) = \frac{1}{2}\|z\|_2^2$ is convex; the composition $a \mapsto f(W^{1/2}(Aa - y))$ is therefore convex. We also have just shown that $\hat{y} \mapsto \ell_{\text{band}}(\hat{y}; b_i, a_i)$ is convex for each i . Composition with the affine map $a \mapsto (Aa)_i$ preserves convexity, so $a \mapsto \ell_{\text{band}}((Aa)_i; b_i, a_i)$ is convex for all i . Summation over i and scaling by $\mu \geq 0$ preserve convexity. Adding the two convex terms yields that $\mathcal{L}_{\text{fit}}$ is convex in a . $\square$

Remark 8 (Optional dead-zone/margin). To avoid hugging the band edges, one may widen the interior by a margin $\delta_i \geq 0$ and replace $[b_i, a_i]$ with $[b_i + \delta_i, a_i - \delta_i]$ in (5.2). All results below are unchanged.

5.2 Quadratic-program form via auxiliary slacks

While (5.1) is already convex in a , it is possible to cast it as a QP with *only* a quadratic objective and linear constraints. Introduce non negative slacks (u_i, v_i) per quote:

$$\ell_{\text{band}}(\hat{y}; b, a) = \min_{u, v \geq 0} \frac{1}{2}(u^2 + v^2) \quad \text{s.t.} \quad u \geq b - \hat{y}, \quad v \geq \hat{y} - a. \quad (5.3)$$

Lemma 2 (Exact equivalence of (5.2) and (5.3)). *For any $b \leq a$ and any $\hat{y} \in \mathbb{R}$, the optimal slacks in (5.3) are $u^* = (b - \hat{y})_+$ and $v^* = (\hat{y} - a)_+$, and the optimal value equals $\frac{1}{2}[(b - \hat{y})_+^2 + (\hat{y} - a)_+^2] = \ell_{\text{band}}(\hat{y}; b, a)$.*

Proof. If $\hat{y} \in [b, a]$, feasibility with $u = v = 0$ gives value 0; nonnegativity enforces $u = v = 0$ at optimum. If $\hat{y} < b$, the constraints reduce to $u \geq b - \hat{y} > 0$ and $v \geq 0$, so the quadratic objective is minimised at $(u^*, v^*) = (b - \hat{y}, 0)$. The case $\hat{y} > a$ is symmetric. $\square$

Stacking (5.3) over quotes yields the QP

$$\min_{a, u, v} \frac{1}{2}\|W^{1/2}(Aa - y)\|_2^2 + \frac{\mu}{2}(\|u\|_2^2 + \|v\|_2^2) \quad \text{s.t.} \quad \begin{cases} u \geq b - Aa, & u \geq 0, \\ v \geq Aa - a, & v \geq 0, \end{cases} \quad (5.4)$$

where all inequalities are coordinate-wise.

Remark 9 (KKT and projection viewpoint). At the solution of (5.3) for a fixed $\hat{y}$, (u^*, v^*) is precisely the vector of signed violations projected onto the nonnegative orthant; equivalently, $\sqrt{2}\ell_{\text{band}}(\hat{y}; b, a) = \text{dist}(\hat{y}, [b, a])$. Thus the second term in (5.1) is $\frac{\mu}{2}\|\text{dist}(Aa, [b, a])\|_2^2$ (coordinatewise distance).

Strict convexity and uniqueness of the data QP

Definition 3 (Positive definiteness on the span of A). We say that the quadratic form $Q(a) = \frac{1}{2} a^\top A^\top W A a$ is *positive definite on the span of A* if

$$a \neq 0 \text{ and } Aa \neq 0 \implies a^\top A^\top W A a = (Aa)^\top W (Aa) > 0.$$

In particular, Q is strictly convex in the prediction variable $p := Aa$, and in coefficient space its only flat directions are those in $\ker(A)$: for each fixed p the restriction of Q to the affine fibre $\{a : Aa = p\}$ is constant.

Proposition 3 (Strict convexity $\implies$ uniqueness). *Assume $W \succ 0$ (symmetric positive definite, ie $x^\top W x > 0$ for every non-zero vector x) and $\mu > 0$. If $A^\top W A$ is positive definite on the span of A (in particular, if A has full column rank), then the objective of (5.4) is strictly convex in (a, u, v) , and hence (5.4) has a unique optimiser (a^*, u^*, v^*) whenever the feasible set is nonempty. Here and throughout, uniqueness is understood modulo the nullspace of A : if A is rank-deficient and (a_1, u_1, v_1) and (a_2, u_2, v_2) are both optimal solutions of (5.4), then $Aa_1 = Aa_2$, $u_1 = u_2$, $v_1 = v_2$, and $a_2 - a_1 \in \ker(A)$. Moreover, if a ridge term $\frac{\lambda}{2} \|a\|_2^2$ with $\lambda > 0$ is added, the objective is strictly convex unconditionally (regardless of $\text{rank}(A)$), yielding uniqueness of the optimiser.*

Proof. The slack QP has decision variables $z = (a, u, v)$ and objective

$$F(a, u, v) = \min_{a, u, v} \frac{1}{2} \|W^{1/2}(Aa - y)\|_2^2 + \frac{\mu}{2} (\|u\|_2^2 + \|v\|_2^2)$$

Expanding the first term

$$\frac{1}{2} (Aa - y)^\top W (Aa - y) = \frac{1}{2} a^\top A^\top W A a - y^\top W A a + \frac{1}{2} y^\top W y$$

Firstly we can see that $\frac{1}{2} y^\top W y$ does not depend on the decision variable, so it is a constant. A quadratic function can be written as:

$$q(z) = \frac{1}{2} z^\top H z + c^\top z + \text{constant}$$

with H symmetric. We can see from the expansion that $F(a, u, v)$ is a quadratic function of (a, u, v) with Hessian

$$H = \begin{bmatrix} A^\top W A & 0 & 0 \\ 0 & \mu I & 0 \\ 0 & 0 & \mu I \end{bmatrix}.$$

Also note that from the equation, the constant is $\frac{1}{2} y^\top W y$ and

$$C = \begin{bmatrix} -A^\top W y \\ 0 \\ 0 \end{bmatrix}.$$

Since $\mu > 0$, the u - and v -blocks are positive definite (namely $\mu I \succ 0$).

For the a -block, take any direction $\delta a \in \mathbb{R}^P$, and define the prediction perturbation

$$\delta p = A \delta a \in \mathbb{R}^N$$

Then

$$\delta a^\top A^\top W A \delta a = (A \delta a)^\top W (A \delta a) = (\delta p)^\top W (\delta p) = \|\delta p\|_W^2,$$

where $\|z\|_W^2 := z^\top W z$ is the weighted Euclidean norm (since $W \succ 0$ it is in fact a norm). The interpretation is that curvature in the a -block along δa equals the weighted squared change in predictions produced by that δa :

- If $A \delta a \neq 0$, predictions move and the term is > 0
- If $A \delta a = 0$, predictions don't move and the term is $= 0$

The set $\ker(A) = \{\delta a : A \delta a = 0\}$ is the nullspace (directions in coefficient space that leave predictions unchanged).

By the assumption “positive definite on the span of A ”, $\|A \delta a\|_W^2 > 0$ for every δa with $A \delta a \neq 0$; hence along any nonzero direction $(\delta a, \delta u, \delta v)$ with $(\delta u, \delta v) \neq 0$ or $A \delta a \neq 0$ we have

$$(\delta a, \delta u, \delta v)^\top H (\delta a, \delta u, \delta v) = \|A \delta a\|_W^2 + \mu \|\delta u\|_2^2 + \mu \|\delta v\|_2^2 > 0.$$

Thus the objective is strictly convex on $\mathbb{R}^P \times \mathbb{R}^N \times \mathbb{R}^N$ modulo the trivial flat directions $\delta a \in \ker(A)$ with $\delta u = \delta v = 0$. If A has full column rank, $\ker(A) = \{0\}$ and $H \succ 0$, so the objective is strictly convex in (a, u, v) . A strictly convex objective over a convex (polyhedral) feasible set admits at most one minimiser; feasibility of (5.4) then yields uniqueness.

If a ridge term $\frac{\lambda}{2} \|a\|_2^2$ with $\lambda > 0$ is added, the Hessian becomes

$$H_\lambda = \begin{bmatrix} A^\top W A + \lambda I & 0 & 0 \\ 0 & \mu I & 0 \\ 0 & 0 & \mu I \end{bmatrix} \succ 0,$$

which is positive definite regardless of $\text{rank}(A)$, hence the objective is strictly convex in (a, u, v) and the minimiser is unique. $\square$

Remark 10 (What is unique when A is rank-deficient). If A is rank-deficient and no ridge is used, the objective is strictly convex in the *predictions* $p := Aa$ and in (u, v) , but flat along $\ker(A)$. Consequently, the optimiser's predictions $p^* = Aa^*$ and slacks (u^*, v^*) are unique, while a^* is unique only up to additions by vectors in $\ker(A)$. Adding a small ridge fixes a^* uniquely.

5.3 Weights, units, and invariance

The choice $w_i \propto (a_i - b_i)^{-2}$ makes the mid-squared error scale-free with respect to the local band width; and the multiplicative factor liq_i up-weights more reliable quotes. The hinge term already measures squared band distance, so μ is dimensionless. If one rescales all prices by a factor $c > 0$, then $A \leftarrow cA$, $y \leftarrow cy$, $b \leftarrow cb$, $a \leftarrow ca$; the minimiser of (5.4) is unchanged after dividing μ by c^2 and multiplying W by c^{-2} —this is the standard homogeneity of quadratic objectives.

5.4 Binned variant (optional)

To stabilise sparse regions, let $G \in \{0, 1\}^{B \times N}$ be a selector that sums quotes in (m, τ) bins. Replacing per-quote hinge terms by binned terms yields

$$\sum_{i=1}^N \ell_{\text{band}}((Aa)_i; b_i, a_i) \rightsquigarrow \sum_{b=1}^B \ell_{\text{band}}((GAa)_b; (Gb)_b, (Ga)_b), \quad (5.5)$$

which is still a QP by Lemma 2, with slacks now attached to bins. The binned form penalises average violations in each cell and reduces sensitivity to isolated outliers.

5.5 Feasibility and the role of μ

Let $\mathcal{S}_{\text{band}} = \{a \in \mathbb{R}^P : b \leq Aa \leq a\}$ be the band-feasible set (coordinate-wise).

- If $\mathcal{S}_{\text{band}} \neq \emptyset$ and other constraints (no-arb penalties or hard inequalities) admit a feasible intersection, then taking $\mu \rightarrow \infty$ in (5.4) forces Aa into the band while the WLS term selects the point closest to y among the band-feasible reconstructions.
- If the intersection is empty, (5.4) finds the unique pair (a, u, v) that minimises the ℓ_2 -distance of Aa to the rectangle $[b, a]$ while trading off the mid anchor through W .

Remark 11 (Targeting coverage). Define $\text{coverage}(a) = \frac{1}{N} \sum_i \mathbf{1}\{b_i \leq (Aa)_i \leq a_i\}$. Increasing μ reduces hinge violations and typically (empirically) increases coverage; we adjust μ with a short controller to hit a target coverage level. Formal monotonicity in μ is not required for the optimiser or the QP structure.

5.6 What is used in this paper (precise choices)

1. **Forward-discounted bands and mids:** (b_i, a_i, y_i) constructed at each quote and robustly cleaned so $0 \leq b_i \leq a_i$.
2. **Weights:** $w_i = \text{liq}_i / \max(a_i - b_i, \varepsilon)^2$ with ε a small fixed floor; $W = \text{diag}(w)$.
3. **Band hinge:** quadratic ℓ_{band} as in (5.2); no interior margin unless stated (set $\delta_i = 0$ by default).
4. **QP form:** auxiliary slacks $(u, v) \geq 0$ with linear constraints (5.4), solved jointly with the rest of the QP (ridge, no-arb penalties, etc.).
5. **Optional binning:** G -aggregation in (5.5) enabled on sparse books; otherwise per-quote hinge.
6. **Controller for μ :** simple scheduler that increases μ until the observed coverage reaches the target (with caps); WLS weight W is held fixed across the schedule.

6. RIDGE, SPECTRAL GEOMETRY, AND TRANSPORT REGULARISATION

This chapter specifies the quadratic regularisers added to the objective, and the price invariant reparameterisation used to improve conditioning. Every term below is a fixed quadratic form in the coefficient vector $a \in \mathbb{R}^P$ (or in a linear reparameterisation $\tilde{a}$), so the overall problem remains a convex QP.

6.1 Spectral ridge (modal energy control)

Let the tensor index be (k, ℓ) with $k = 0, \dots, K$ (log-moneyness) and $\ell = 0, \dots, L$ (maturity). Define a diagonal weight

$$\Lambda_{(k,\ell),(k,\ell)} = (1 + \alpha k^2 + \beta \ell^2)^s, \quad \alpha, \beta > 0, \quad s > 0,$$

and set $\Lambda = \text{diag}(\Lambda_{(k,\ell),(k,\ell)}) \in \mathbb{R}^{P \times P}$ with $P = (K+1)(L+1)$. The spectral ridge is defined as

$$\mathcal{R}_{\text{ridge}}(a) = \frac{\lambda_{\text{ridge}}}{2} \|\Lambda^{1/2} a\|_2^2 = \frac{\lambda_{\text{ridge}}}{2} a^\top \Lambda a = \frac{\lambda_{\text{ridge}}}{2} \sum_{k=0}^K \sum_{\ell=0}^L (1 + \alpha k^2 + \beta \ell^2)^s a_{kl}^2. \quad (6.1)$$

Interpretation: Each coefficient is penalised by a weight that grows with its modal index. Low modes (small k and ℓ) get weight ≈ 1 ; higher (k, ℓ) modes carry larger weights (are expensive). α, β tune the relative penalisation across m vs. τ , and s controls the growth rate (asymptotically the weights grow like $(\alpha k^2 + \beta \ell^2)^s$, so $s = 1$ gives quadratic growth in the indices and $s = 2$ gives quartic growth). This damps high-frequency oscillations while leaving low modes essentially unchanged. In spectral methods, smooth functions have rapidly decaying coefficients and non-smooth noisy features push energy into high indices. Penalising a_{kl}^2 with a weight increasing in k, ℓ is the discrete analogue of a Sobolev H^s seminorm, suppressing high frequency components while leaving low modes mostly alone. The wraps Φ_m and Φ_τ mean smoothness is enforced in the wrapped coordinates where the basis is well-conditioned (ATM focus and short- τ density).

Unless otherwise stated, we fix $\alpha = \beta = 1$ and $s = 2$. The scalar λ_{ridge} is chosen once per date by a small-subsample generalised cross-validation (GCV) pass on the *linear* WLS subproblem. Build A_{sub} and W_{sub} on a random 8% subset of quotes. Namely $A_{\text{sub}} \in \mathbb{R}^{N_{\text{sub}} \times P}$, $W_{\text{sub}} = \text{diag}(w_{\text{sub}}) \succ 0$ and $y_{\text{sub}} \in \mathbb{R}^{N_{\text{sub}}}$. For any $\lambda > 0$, solve the ridge-regularised weighted least squares:

$$\min_a \frac{1}{2} \|W_{\text{sub}}^{1/2} (A_{\text{sub}} a - y_{\text{sub}})\|_2^2 + \frac{\lambda}{2} \|\Lambda^{1/2} a\|_2^2,$$

This has the closed form

$$a(\lambda) = (A_{\text{sub}}^\top W_{\text{sub}} A_{\text{sub}} + \lambda \Lambda)^{-1} A_{\text{sub}}^\top W_{\text{sub}} y_{\text{sub}}$$

Define the weighted residual and the hat matrix

$$r(\lambda) = W_{\text{sub}}^{1/2} (A_{\text{sub}} a(\lambda) - y_{\text{sub}}); \quad H(\lambda) = W_{\text{sub}}^{1/2} A_{\text{sub}} (A_{\text{sub}}^\top W_{\text{sub}} A_{\text{sub}} + \lambda \Lambda)^{-1} A_{\text{sub}}^\top W_{\text{sub}}^{1/2}.$$

Remark 12. A useful identity for computation is

$$\text{tr}(H(\lambda)) = \text{tr}(SG(\lambda)^{-1}), \quad S = (A_{\text{sub}}^\top W_{\text{sub}} A_{\text{sub}}), \quad G(\lambda) = S + \lambda \Lambda.$$

We compute the GCV (Generalised cross-validation) score

$$\text{GCV}(\lambda) = \frac{\|r(\lambda)\|_2^2}{(N_{\text{sub}} - \text{tr } H(\lambda))^2}.$$

Choose $\lambda_{\text{ridge}} = \arg \min_{\lambda} \text{GCV}(\lambda)$ and use it in the full QP.

Lemma 3 (Spectral ridge is a fixed quadratic form). *Fix the tensor grid (k, ℓ) with $k = 0, \dots, K$, $\ell = 0, \dots, L$, and hyperparameters $\alpha, \beta > 0$, $s > 0$. Let $\Lambda \in \mathbb{R}^{P \times P}$, $P = (K+1)(L+1)$, be the diagonal matrix defined above and let $\lambda_{\text{ridge}} \geq 0$ be fixed. Then for every $a \in \mathbb{R}^P$, the spectral ridge (6.1) can be written as*

$$\mathcal{R}_{\text{ridge}}(a) = \frac{1}{2} a^\top Q_{\text{ridge}} a, \quad Q_{\text{ridge}} := \lambda_{\text{ridge}} \Lambda,$$

with Q_{ridge} symmetric positive semidefinite and independent of a . In particular, $\mathcal{R}_{\text{ridge}}$ is a convex quadratic function of the coefficient vector a , and enters any optimisation problem as a fixed quadratic form (for given (α, β, s) and λ_{ridge}).

Proof. By definition,

$$\Lambda_{(k,\ell),(k,\ell)} = (1 + \alpha k^2 + \beta \ell^2)^s > 0 \quad \text{for all } 0 \leq k \leq K, \ 0 \leq \ell \leq L,$$

so Λ is diagonal with strictly positive diagonal entries and hence $\Lambda \succeq 0$ (indeed $\Lambda \succ 0$). For any $a \in \mathbb{R}^P$,

$$\|\Lambda^{1/2} a\|_2^2 = a^\top \Lambda^{1/2} \Lambda^{1/2} a = a^\top \Lambda a.$$

Thus (6.1) can be rewritten as

$$\mathcal{R}_{\text{ridge}}(a) = \frac{\lambda_{\text{ridge}}}{2} a^\top \Lambda a = \frac{1}{2} a^\top Q_{\text{ridge}} a, \quad Q_{\text{ridge}} := \lambda_{\text{ridge}} \Lambda.$$

The matrix Q_{ridge} is symmetric. Since $\lambda_{\text{ridge}} \geq 0$ and $\Lambda \succeq 0$, we have $Q_{\text{ridge}} \succeq 0$, so the map $a \mapsto \frac{1}{2} a^\top Q_{\text{ridge}} a$ is a convex quadratic function. For fixed hyperparameters (α, β, s) , grid sizes (K, L) , and a chosen value of λ_{ridge} , the matrix Q_{ridge} is completely determined and does not depend on a . Hence $\mathcal{R}_{\text{ridge}}$ is a fixed quadratic form in the coefficient vector. $\square$

6.2 Λ -module: price-invariant reparameterisation

Now propose a change of coordinates in the coefficient space to make the optimisation numerically well-behaved. Since the variables and matrices are changed in a consistent way, all prices and constraint values stay identical, and the conditioning of the problem improves.

Let $U \in \mathbb{R}^{P \times P}$ be a fixed, invertible linear map. Define new coefficients $\tilde{a} := U^{-1}a$ and replace every block by post-multiplication with U :

$$A \leftarrow AU, \quad A_m \leftarrow A_m U, \quad A_{mm} \leftarrow A_{mm} U, \quad A_\tau \leftarrow A_\tau U, \quad \text{etc.}$$

Predictions are unchanged: $(AU)\tilde{a} = A(U\tilde{a}) = Aa$. The ridge becomes

$$\mathcal{R}_{\text{ridge}}(a) = \frac{\lambda_{\text{ridge}}}{2} \|\Lambda^{1/2} U \tilde{a}\|_2^2 = \frac{\lambda_{\text{ridge}}}{2} \tilde{a}^\top \underbrace{U^\top \Lambda U}_{\tilde{\Lambda}} \tilde{a},$$

i.e. the same quadratic form in $\tilde{a}$ with $\tilde{\Lambda} = U^\top \Lambda U$.

Blockwise whitening and why it is safe: Firstly, partition columns by maturity slice ℓ (all m -modes for that ℓ) and define the sets $\{\mathcal{G}_\ell\}$. Then for each block form the *weighted* thin QR

$$W^{1/2} A_{[:, \mathcal{G}_\ell]} = Q_\ell R_\ell, \quad Q_\ell^\top Q_\ell = I, \quad R_\ell \text{ invertible upper-triangular},$$

and assemble a block-diagonal U with $U_{\mathcal{G}_\ell, \mathcal{G}_\ell} := R_\ell^{-1}$ (zeros off-block). Optionally, right-scale columns so $\|W^{1/2}(AU)_{[:, j]}\|_2 = 1$ by replacing $U \leftarrow UD^{-1}$ with $D = \text{diag}(d_j)$, $d_j = \|W^{1/2}(AU)_{[:, j]}\|_2$.

We implicitly require that each slice $A[:, G_\ell]$ have full column rank under the W -inner product, so that the thin QR factorisation with a square, invertible R_ℓ exists. This condition is satisfied for the Chebyshev grids used in our experiments; in degenerate cases one can replace R_ℓ^{-1} by a pseudo-inverse obtained from a rank-revealing QR or SVD, at the price of a slightly lower-dimensional reparameterisation.

Proposition 4 (W-orthonormality within slices). *With U defined above,*

$$(AU)_{[:, \mathcal{G}_\ell]} = W^{-1/2} Q_\ell \implies (AU)_{[:, \mathcal{G}_\ell]}^\top W (AU)_{[:, \mathcal{G}_\ell]} = I_{|\mathcal{G}_\ell|}.$$

In particular, the WLS normal matrix becomes block-identity within each slice (collinearity removed).

Proof. $A_{[:, \mathcal{G}_\ell]} = W^{-1/2} Q_\ell R_\ell$ and $U_{\mathcal{G}_\ell, \mathcal{G}_\ell} = R_\ell^{-1}$ give $(AU)_{[:, \mathcal{G}_\ell]} = W^{-1/2} Q_\ell$.

Hence $(AU)_{[:, \mathcal{G}_\ell]}^\top W (AU)_{[:, \mathcal{G}_\ell]} = Q_\ell^\top Q_\ell = I$. □

Proposition 5 (Price/constraint invariance). *Let $\tilde{a} = U^{-1}a$, $A' := AU$, and for any block $A_\bullet$ set $A'_\bullet := A_\bullet U$. Then*

$$A' \tilde{a} = Aa, \quad A'_\bullet \tilde{a} = A_\bullet a.$$

Consequently, hard inequalities $A_\bullet a \leq 0$ are equivalent to $(A_\bullet U) \tilde{a} \leq 0$, and soft penalties that depend on $A_\bullet a$ take the same values when written in $\tilde{a}$.

Proof. We see by definition of A' and $\tilde{a}$, $A' \tilde{a} = AUU^{-1}a = Aa$ as required. □

Proposition 6 (Ridge congruence). *For any symmetric $Q \succeq 0$,*

$$\frac{1}{2} a^\top Q a = \frac{1}{2} \tilde{a}^\top (U^\top Q U) \tilde{a}.$$

In particular, the spectral ridge becomes $\frac{\lambda_{\text{ridge}}}{2} \tilde{a}^\top \tilde{\Lambda} \tilde{a}$ with $\tilde{\Lambda} := U^\top \Lambda U \succeq 0$.

Proof. Substitute $a = U \tilde{a}$ and regroup; congruence preserves positive semidefiniteness. □

Remark 13 (Global whitening as a special case). If one QR-factorises $W^{1/2}A = QR$ once and sets $U = R^{-1}$, then $(AU)^\top W (AU) = I_P$ (full whitening); the blockwise construction above is its per-slice counterpart.

6.3 DW-module: discrete transport (H^{-1}) smoothing of density

This section penalises oscillations of the risk-neutral density ρ along m by measuring how much potential ϕ is needed so its discrete derivative matches ρ . High frequency wiggles are expensive and slowly varying shapes cost little.

Let $\rho = \partial_{KK} C_f$ denote the risk-neutral density. On each maturity slice $\tau = \tau_g$, we discretise the m -axis on the collocation nodes m_j and build:

- a diagonal mass matrix $M_m = \text{diag}(w^{(m)})$ with Gauss-Lobatto (Chebyshev) quadrature weights.
- a first-difference matrix $D_m \in \mathbb{R}^{(M_m-1) \times M_m}$ (forward differences with homogeneous Neumann boundary, i.e. zero-flux ends) with $(D_m \phi)_i = \phi_{i+1} - \phi_i$ for $i = 1, \dots, M_m - 1$.

Define the discrete Neumann Laplacian in m by

$$L_m := D_m^\top M_m^{-1} D_m \in \mathbb{R}^{M_m \times M_m}.$$

L_m is symmetric positive semidefinite and its nullspace is the span of the constant vector (along each slice). For a discrete function $f \in \mathbb{R}^{M_m}$, the discrete $H^{-1}(m)$ seminorm is defined by

$$\|f\|_{H^{-1}(m)}^2 := f^\top \underbrace{L_m^+}_{\text{Moore-Penrose pseudoinverse of } L_m} f,$$

where $^+$ is the Moore-Penrose pseudoinverse on the range. This is the standard discrete Neumann H^{-1} seminorm: L_m^+ plays the role of the inverse Laplacian, so only the mean-zero component of f is penalised and the constant/mean mode lies in the nullspace.

Assemble the full grid operator $L^+ = \text{blkdiag}(L_m^+, \dots, L_m^+)$ across slices and the sampling matrix $S : \mathbb{R}^G \rightarrow \mathbb{R}^{M_m(M_\tau+1)}$ that reshapes grid values into slice stacks. With $E := S A_{KK} \in \mathbb{R}^{M_m(M_\tau+1) \times P}$ (density map in slice-stacked order), the DW penalty is

$$\mathcal{R}_{\text{DW}}(a) = \frac{\lambda_{\text{DW}}}{2} \|\rho(a)\|_{H^{-1}}^2 = \frac{\lambda_{\text{DW}}}{2} a^\top E^\top L^+ E a, \quad (6.2)$$

a fixed quadratic form once L^+ is precomputed (e.g. Cholesky on each L_m on the mean-zero subspace, plus a rank-1 fix for the constant nullspace).

In Fourier language, $\|f\|_{H^{-1}(m)}^2 \sim \sum_k |f_k|^2 / k^2$. The high k content is amplified, so the optimiser prefers smooth densities. The constant/mean component sits in the nullspace and is not penalised. The constraint set continues to control positivity/monotonicity and DW just damps ripples that those constraints do not eliminate.

We apply (6.2) *along m only* on each slice, with Neumann boundaries and Chebyshev-Lobatto weights. The difference operator D_m only takes interior forward differences,

$$(D_m \phi)_i = \phi_{i+1} - \phi_i \quad i = 1, \dots, M_m - 1,$$

so the associated Laplacian $L_m = D_m^\top M_m^{-1} D_m$ has the constant vector in its nullspace: adding a constant to ϕ does not change $D_m \phi$ or the quadratic form. In the H^{-1} penalty, this means the mean component of ρ is left unpenalised and only fluctuations around the mean contribute to $\|\rho\|_{H^{-1}}$; this is the discrete zero-flux (Neumann) condition at the ends.

Equivalently, on each slice we can view ρ as a one-dimensional “charge distribution” along m . The matrix L_m is a discrete Neumann Laplacian on the nodes, and L_m^+ acts as its inverse on mean-zero densities. For any profile $f \in \mathbb{R}^{M_m}$ with zero M_m -mean there exists a potential ϕ (unique up to an additive constant) solving the discrete Poisson problem

$$L_m \phi = f \quad (\text{Neumann in } m).$$

Among all such potentials, the one with the smallest discrete Dirichlet energy $\phi^\top L_m \phi$ satisfies

$$\phi^\top L_m \phi = f^\top L_m^+ f = \|f\|_{H^{-1}(m)}^2.$$

So $\|f\|_{H^{-1}(m)}^2$ measures how much “bending” of the potential ϕ is needed to support the density: sharply oscillating f requires a highly curved potential and incurs a large penalty, while slowly varying f can be supported by a gentle potential and is cheap. The constant component of f generates no potential at all and is left unpenalised.

Chebyshev-Lobatto nodes cluster near the endpoints, and without quadrature weights, any discrete L^2 inner product would outweigh the ends and underweight the middle. $M_m = \text{diag}(w^{(m)})$ fixes that. For a smooth function g ,

$$\sum_{j=1}^{M_m} w_j^{(m)} g(m_j) \approx \int_{m_{\min}}^{m_{\max}} g(m) dm.$$

So $\phi^\top M_m \phi$ is a proper discretisation of $\int \phi(m)^2 dm$, independent of how densely sampled near the ends. The weights are obtained using standard Clenshaw-Curtis (Gauss-Lobatto) quadrature on the Chebyshev-Lobatto nodes in the reference variable $n \in [-1, 1]$ and then rescaled to $m \in [m_{\min}, m_{\max}]$.

If $m = \frac{m_{\max} - m_{\min}}{2} n + \frac{m_{\max} + m_{\min}}{2}$, the Jacobian is constant and

$$w_j^{(m)} = \frac{m_{\max} - m_{\min}}{2} w_j^{(n)}.$$

The weight is tapered in maturity: $\lambda_{\text{DW}}(\tau_g) = \lambda_{\text{DW}}^{(0)} \cdot \min\{1, \tau_\star / \tau_g\}$ to dampen short maturity ripples; defaults $\tau_\star = 5$ trading days. For $\tau_g \leq \tau_\star$, we use the full DW smoothing strength $\lambda_{\text{DW}}^{(0)}$ and for $\tau_g > \tau_\star$ the smoothing weight decays like $1/\tau_g$ so that the long end is not over-smoothed.

Lemma 4 (DW penalty as a fixed quadratic form). *Fix the grid operators*

$$L^+ \in \mathbb{R}^{M_m(M_\tau+1) \times M_m(M_\tau+1)}, \quad E := S A_{KK} \in \mathbb{R}^{M_m(M_\tau+1) \times P},$$

as above, and let $\lambda_{\text{DW}} \geq 0$ be given. Then for every $a \in \mathbb{R}^P$, let

$$\mathcal{R}_{\text{DW}}(a) = \frac{\lambda_{\text{DW}}}{2} \|\rho(a)\|_{H^{-1}}^2 = \frac{1}{2} a^\top Q_{\text{DW}} a, \quad Q_{\text{DW}} := \lambda_{\text{DW}} E^\top L^+ E,$$

with Q_{DW} symmetric positive semidefinite and independent of a . Then $\mathcal{R}_{\text{DW}}$ is a convex quadratic function of the coefficient vector a and enters the global problem as a fixed quadratic term, so the formulation remains a convex QP.

Proof. By construction, $\rho(a)$ is linear in a : on the slice-stacked grid, $\rho(a) = Ea$ with $E = SA_{KK}$ independent of a . The discrete H^{-1} seminorm is

$$\|\rho(a)\|_{H^{-1}}^2 = \rho(a)^\top L^+ \rho(a) = (Ea)^\top L^+ (Ea) = a^\top E^\top L^+ Ea.$$

Thus

$$\mathcal{R}_{\text{DW}}(a) = \frac{\lambda_{\text{DW}}}{2} a^\top E^\top L^+ Ea = \frac{1}{2} a^\top Q_{\text{DW}} a$$

with $Q_{\text{DW}} := \lambda_{\text{DW}} E^\top L^+ E$. The operator L^+ is symmetric positive semidefinite by construction, as the Moore–Penrose pseudoinverse of the symmetric positive semidefinite block-diagonal matrix whose blocks are $L_m = D_m^\top M_m^{-1} D_m$. Hence for any z , $z^\top L^+ z \geq 0$ (by definition of symmetric positive semidefinite), and in particular

$$a^\top Q_{\text{DW}} a = \lambda_{\text{DW}} (Ea)^\top L^+ (Ea) \geq 0$$

for all a , so $Q_{\text{DW}} \succeq 0$ whenever $\lambda_{\text{DW}} \geq 0$. All ingredients $(E, L^+, \lambda_{\text{DW}})$ are fixed once the grid, quadrature, and smoothing weight are chosen, so Q_{DW} does not depend on a . Therefore $\mathcal{R}_{\text{DW}}$ is a fixed convex quadratic function of the coefficients. $\square$

Remark 14 (Maturity taper). If the smoothing weight is tapered in maturity, with slice weights $\lambda_{\text{DW}}(\tau_g) \geq 0$ as above, one can collect them in a diagonal matrix Λ_{DW} acting on the slice-stacked density, and write

$$\mathcal{R}_{\text{DW}}(a) = \frac{1}{2} (Ea)^\top \Lambda_{\text{DW}} L^+ (Ea) = \frac{1}{2} a^\top Q_{\text{DW}} a$$

with $Q_{\text{DW}} := E^\top \Lambda_{\text{DW}} L^+ E$ (or $Q_{\text{DW}} := E^\top \Lambda_{\text{DW}}^{1/2} L^+ \Lambda_{\text{DW}}^{1/2} E$). This is again symmetric positive semidefinite and independent of a , so the DW term remains a fixed quadratic form under tapering.

6.4 RN-module: near-maturity residual and calendar flattening

The objective here is on very short maturities to pull the forward-discounted call surface towards its $\tau \downarrow 0$ limit and suppress calendar wiggles at fixed strike. Both are quadratic, so the overall QP stays convex.

On $\Gamma_{0+} = \{(m_g, \tau_g) : \tau_g \leq \tau_\star\}$, anchor the price to the intrinsic limit and penalise calendar drift at fixed strike:

$$\mathcal{R}_{\text{RN}}(a) = \frac{\lambda_{\text{RN}}}{2} \|Aa - C_{0+}\|_{2, \Gamma_{0+}}^2 + \frac{\eta_{\text{RN}}}{2} \|A_{\tau|K} a\|_{2, \Gamma_{0+}}^2, \quad (6.3)$$

where $C_{0+}(m) = F_0(1 - e^m)_+$ and $\|\cdot\|_{2, \Gamma_{0+}}$ is the ℓ_2 norm restricted to indices in Γ_{0+} . The terms mean the following:

1. Near-maturity anchor: $C_{0+}(m)$ is the intrinsic value of a forward-discounted call. As $\tau \downarrow 0$, no-arb implies that $C_f(m, \tau) \rightarrow C_{0+}(m)$. The term $\|Aa - C_{0+}\|_{2, \Gamma_{0+}}^2$ enforces this only on the short-end grid Γ_{0+} .
2. Calendar flattening at fixed strike: $\|A_{\tau|K} a\|_{2, \Gamma_{0+}}^2$ penalises the τ -slope at fixed K near $\tau = 0$. This damps spurious day-to-day oscillations that data sparsity and noise can introduce at the short end.

We must derive $A_{\tau|K}$, which was done in 4.3. Since the operator is linear in a , then it is a quadratic penalty.

To suppress odd-in- m artifacts at $\tau \downarrow 0$ we optionally project coefficients onto the even subspace along m : let P_{even} be the diagonal projector with $(P_{\text{even}})_{(k,\ell),(k,\ell)} = 1$ for even k and 0 for odd k ; replace a by $P_{\text{even}}a$ when evaluating the first term in (6.3). This is linear and preserves QP structure.

We set $\tau_\star \in [5, 10]$ trading days. Default weights: λ_{RN} chosen so that the first term's RMS on Γ_{0+} matches the median band width there; η_{RN} is set to achieve $\leq 0.5\%$ calendar violations on the shortest two slices once combined with the no-arb penalties. Parity projection is off by default ($P_{\text{even}} = I$) unless short-end butterflies appear.

Lemma 5 (RN penalty as a fixed quadratic in the coefficients). *Let Γ_{0+} be the short-maturity index set and let $\Pi_{0+} \in \mathbb{R}^{N \times N}$ be the diagonal selector with $(\Pi_{0+})_{ii} = 1$ if $i \in \Gamma_{0+}$ and 0 otherwise. Then for any $a \in \mathbb{R}^P$ the RN penalty (6.3) can be written as*

$$\mathcal{R}_{\text{RN}}(a) = \frac{1}{2} a^\top Q_{\text{RN}} a + c_{\text{RN}}^\top a + \text{const},$$

with

$$Q_{\text{RN}} := \lambda_{\text{RN}} A^\top \Pi_{0+} A + \eta_{\text{RN}} A_{\tau|K}^\top \Pi_{0+} A_{\tau|K} \succeq 0, \quad c_{\text{RN}} := -\lambda_{\text{RN}} A^\top \Pi_{0+} C_{0+}.$$

In particular, $\mathcal{R}_{\text{RN}}$ is a convex quadratic function of the coefficient vector a with fixed Hessian Q_{RN} , so adding $\mathcal{R}_{\text{RN}}$ to the objective preserves the convex QP structure.

Proof. By definition of the restricted norm, there exists a diagonal selector Π_{0+} such that for any $x \in \mathbb{R}^N$,

$$\|x\|_{2,\Gamma_{0+}}^2 = \|\Pi_{0+}x\|_2^2 = x^\top \Pi_{0+} x.$$

Therefore the two terms in (6.3) can be written as

$$\begin{aligned} \|Aa - C_{0+}\|_{2,\Gamma_{0+}}^2 &= (Aa - C_{0+})^\top \Pi_{0+} (Aa - C_{0+}), \\ \|A_{\tau|K}a\|_{2,\Gamma_{0+}}^2 &= (A_{\tau|K}a)^\top \Pi_{0+} (A_{\tau|K}a). \end{aligned}$$

Expanding the first term gives

$$(Aa - C_{0+})^\top \Pi_{0+} (Aa - C_{0+}) = a^\top A^\top \Pi_{0+} Aa - 2C_{0+}^\top \Pi_{0+} Aa + C_{0+}^\top \Pi_{0+} C_{0+},$$

while the second term is already of the form

$$(A_{\tau|K}a)^\top \Pi_{0+} (A_{\tau|K}a) = a^\top A_{\tau|K}^\top \Pi_{0+} A_{\tau|K} a.$$

Plugging into (6.3) yields

$$\mathcal{R}_{\text{RN}}(a) = \frac{1}{2} a^\top Q_{\text{RN}} a + c_{\text{RN}}^\top a + \text{const},$$

with Q_{RN} and c_{RN} as claimed, and a constant term $\frac{\lambda_{\text{RN}}}{2} C_{0+}^\top \Pi_{0+} C_{0+}$ which does not depend on a .

For any $a \in \mathbb{R}^P$,

$$a^\top Q_{\text{RN}} a = \lambda_{\text{RN}} \|Aa\|_{2,\Gamma_{0+}}^2 + \eta_{\text{RN}} \|A_{\tau|K}a\|_{2,\Gamma_{0+}}^2 \geq 0$$

whenever $\lambda_{\text{RN}}, \eta_{\text{RN}} \geq 0$, so $Q_{\text{RN}} \succeq 0$ and the Hessian of $\mathcal{R}_{\text{RN}}$ is positive semidefinite. All matrices $A, A_{\tau|K}, \Pi_{0+}$ and the vector C_{0+} are fixed once the grid, short-maturity set Γ_{0+} , and weights $\lambda_{\text{RN}}, \eta_{\text{RN}}$ are chosen; hence Q_{RN} and c_{RN} are independent of a and $\mathcal{R}_{\text{RN}}$ is a fixed convex quadratic function of the coefficients. $\square$

6.5 Ω -module: high-frequency taper and commutator hook

Let $U_\omega \in \mathbb{R}^{P \times P}$ be a fixed orthogonal change of basis to a frequency chart (separable 2D DCT aligned with the (k, ℓ) grid). Denote $\hat{a} = U_\omega a$ and let M_ω be a diagonal mask selecting high-frequency indices. The taper is

$$\mathcal{R}_\Omega(a) = \frac{\lambda_\Omega}{2} \|M_\omega \hat{a}\|_2^2 = \frac{\lambda_\Omega}{2} a^\top U_\omega^\top M_\omega^\top M_\omega U_\omega a. \quad (6.4)$$

We use a maturity-dependent mask: on slices with $\tau_g \leq \tau_\star$ only the top third of m -frequencies are penalised; for $\tau_g > 2\tau_\star$ the mask is zero.

Liouville hook (commutator residual). Let $\mathcal{L}$ denote the forward-flow generator at fixed strike, $\mathcal{L}C_f := (\partial_\tau C_f)|_K$, and let ∂_K be the strike derivative. At the continuum level, mixed derivatives commute and we can write the commutator

$$[\mathcal{L}, \partial_K]C_f := \mathcal{L}(\partial_K C_f) - \partial_K(\mathcal{L}C_f),$$

which vanishes for smooth C_f .

On the collocation grid we work with nodal price vectors $u \in \mathbb{R}^G$ and linear operators

$$\tilde{A}_K, \tilde{A}_{\tau|K} \in \mathbb{R}^{G \times G}$$

acting on u and approximating, respectively, $\partial_K C_f$ and $(\partial_\tau C_f)|_K$ at the grid nodes. They are chosen consistently with the coefficient-space design blocks $A_K, A_{\tau|K} \in \mathbb{R}^{G \times P}$ in the sense that for all coefficient vectors $a \in \mathbb{R}^P$,

$$\tilde{A}_K(Aa) \approx A_K a, \quad \tilde{A}_{\tau|K}(Aa) \approx A_{\tau|K} a,$$

where $u(a) := Aa$ denotes the nodal surface implied by a .

Define the discrete commutator on nodal prices by

$$[\tilde{A}_{\tau|K}, \tilde{A}_K]u := \tilde{A}_{\tau|K}(\tilde{A}_K u) - \tilde{A}_K(\tilde{A}_{\tau|K} u),$$

and set

$$C := \tilde{A}_{\tau|K} \tilde{A}_K - \tilde{A}_K \tilde{A}_{\tau|K} \in \mathbb{R}^{G \times G}.$$

Evaluated at the model surface $u(a) = Aa$, this yields the G -vector commutator defect

$$c(a) := [\tilde{A}_{\tau|K}, \tilde{A}_K]u(a) = C(Aa).$$

We penalise the hook residual via

$$\mathcal{R}_{\text{hook}}(a) := \frac{\lambda_{\text{hook}}}{2} \|c(a)\|_2^2 = \frac{\lambda_{\text{hook}}}{2} \|C(Aa)\|_2^2,$$

with a small stabilising weight $\lambda_{\text{hook}} \ll \lambda_{\text{NA}}$.

U_ω is the separable 2D DCT on (k, ℓ) ; M_ω masks the top 33% highest m -frequencies for $\tau \leq \tau_\star$ and is zero otherwise; default λ_Ω is picked so that the high-frequency modal energy share $\mathcal{E}_{\text{hi}} \leq 5\%$; λ_{hook} is set to a small fraction (10^{-3} – 10^{-2}) of λ_{NA} .

Lemma 6 (Ω taper and hook as fixed quadratics in the coefficients). *Let $U_\omega \in \mathbb{R}^{P \times P}$ be an orthogonal matrix ($U_\omega^\top U_\omega = I$), let $M_\omega \in \mathbb{R}^{P \times P}$ be a fixed diagonal mask, and define $\hat{a} = U_\omega a$. Let*

$$\mathcal{R}_\Omega(a) := \frac{\lambda_\Omega}{2} \|M_\omega \hat{a}\|_2^2$$

and, with $C := \tilde{A}_{\tau|K} \tilde{A}_K - \tilde{A}_K \tilde{A}_{\tau|K} \in \mathbb{R}^{G \times G}$ as above,

$$c(a) := C(Aa), \quad \mathcal{R}_{\text{hook}}(a) := \frac{\lambda_{\text{hook}}}{2} \|c(a)\|_2^2 = \frac{\lambda_{\text{hook}}}{2} \|C(Aa)\|_2^2.$$

Then both penalties are fixed convex quadratic functions of the coefficient vector a :

$$\mathcal{R}_\Omega(a) = \frac{1}{2} a^\top Q_\Omega a, \quad \mathcal{R}_{\text{hook}}(a) = \frac{1}{2} a^\top Q_{\text{hook}} a,$$

with

$$Q_\Omega := \lambda_\Omega U_\omega^\top M_\omega^\top M_\omega U_\omega \succeq 0, \quad Q_{\text{hook}} := \lambda_{\text{hook}} A^\top C^\top C A \succeq 0.$$

All matrices $U_\omega, M_\omega, A, \tilde{A}_K, \tilde{A}_{\tau|K}$, and thus $Q_\Omega, Q_{\text{hook}}$, are independent of a . In particular, adding $\mathcal{R}_\Omega$ and $\mathcal{R}_{\text{hook}}$ to the objective preserves the convex QP structure.

Proof. For the taper, write

$$\mathcal{R}_\Omega(a) = \frac{\lambda_\Omega}{2} \|M_\omega U_\omega a\|_2^2 = \frac{\lambda_\Omega}{2} (U_\omega a)^\top M_\omega^\top M_\omega (U_\omega a).$$

Set $Q_\Omega := \lambda_\Omega U_\omega^\top M_\omega^\top M_\omega U_\omega$. Then

$$\mathcal{R}_\Omega(a) = \frac{1}{2} a^\top Q_\Omega a.$$

For any a ,

$$a^\top Q_\Omega a = \lambda_\Omega \|M_\omega U_\omega a\|_2^2 \geq 0$$

whenever $\lambda_\Omega \geq 0$, so $Q_\Omega \succeq 0$. Once the grid, frequency chart, and maturity-dependent mask are chosen, both U_ω and M_ω are fixed, and therefore Q_Ω is fixed (independent of a).

For the hook, note that A is a fixed linear map from coefficients to nodal prices and C is a fixed linear operator on grid space, so the commutator residual is linear in a :

$$c(a) = C(Aa) = (CA) a.$$

Let $B := CA \in \mathbb{R}^{G \times P}$. Then

$$\mathcal{R}_{\text{hook}}(a) = \frac{\lambda_{\text{hook}}}{2} \|Ba\|_2^2 = \frac{\lambda_{\text{hook}}}{2} a^\top B^\top B a.$$

Setting $Q_{\text{hook}} := \lambda_{\text{hook}} A^\top C^\top C A = \lambda_{\text{hook}} B^\top B$ gives

$$\mathcal{R}_{\text{hook}}(a) = \frac{1}{2} a^\top Q_{\text{hook}} a.$$

For any a ,

$$a^\top Q_{\text{hook}} a = \lambda_{\text{hook}} \|Ba\|_2^2 \geq 0$$

whenever $\lambda_{\text{hook}} \geq 0$, so $Q_{\text{hook}} \succeq 0$. All ingredients are fixed once $A, \tilde{A}_K, \tilde{A}_{\tau|K}$ and λ_{hook} are chosen, so Q_{hook} is independent of a .

Thus both $\mathcal{R}_\Omega$ and $\mathcal{R}_{\text{hook}}$ are fixed convex quadratic functions of the coefficients. $\square$

6.6 Summary of fixed choices used

- **Ridge:** $\alpha = \beta = 1$, $s = 2$; λ_{ridge} by GCV on an 8% WLS subsample (no hinge, no penalties); fixed per date, reused in the full QP.
- **Λ reparameterisation:** U built by blockwise (per- τ) QR/Gram–Schmidt and column rescaling on (AU) over the quotes; all blocks post-multiplied by U ; ridge uses $\tilde{\Lambda} = U^\top \Lambda U$.
- **DW:** H^{-1} along m per slice with Neumann ends; Chebyshev–Lobatto M_m ; $\lambda_{\text{DW}}(\tau) = \lambda_{\text{DW}}^{(0)} \min\{1, \tau_\star/\tau\}$, $\tau_\star \in [5, 10]$ trading days.
- **RN:** window $\Gamma_{0+} = \{\tau \leq \tau_\star\}$; weights $(\lambda_{\text{RN}}, \eta_{\text{RN}})$ calibrated to short-end RMS and calendar share; $P_{\text{even}} = I$ by default.
- **Ω :** separable 2D DCT, mask top 33% m -frequencies for $\tau \leq \tau_\star$, off beyond $2\tau_\star$; λ_Ω chosen to cap high-frequency energy at $\leq 5\%$.
- **Hook:** commutator penalty $\frac{\lambda_{\text{hook}}}{2} \|C(Aa)\|_2^2$ with $C := \tilde{A}_{\tau|K} \tilde{A}_K - \tilde{A}_K \tilde{A}_{\tau|K} \in \mathbb{R}^{G \times G}$ and $\lambda_{\text{hook}} \in [10^{-3}, 10^{-2}] \lambda_{\text{NA}}$.

All terms above are quadratic in a (or $\tilde{a}$) and are entered additively into the QP objective. They stabilise the global fit, suppress short-maturity artifacts, and improve conditioning while preserving convexity and the solver class.

7. NO-ARBITRAGE CONSTRAINTS AND SOFT PENALTIES

We impose the three shape conditions (monotone in K , convex in K , and calendar non-negativity at fixed K) as soft penalties evaluated on the collocation grid, using the linear operators defined previously:

$$A_K, \quad A_{KK}, \quad A_{\tau|K} \in \mathbb{R}^{G \times P}, \quad \text{and the price block } A \in \mathbb{R}^{G \times P}.$$

All vectors below are understood componentwise and $(x)_+ = \max\{x, 0\}$.

7.1 Penalty definitions (soft versions of the shape constraints)

$$\mathcal{P}_{\text{mono}}(a) = \frac{\lambda_{\text{NA}}}{2} \|(A_K a)_+\|_2^2 \quad (\text{targets } \partial_K C_f \leq 0), \quad (7.1)$$

$$\mathcal{P}_{\text{conv}}(a) = \frac{\lambda_{\text{NA}}}{2} \|(-A_{KK} a)_+\|_2^2 \quad (\text{targets } \partial_{KK} C_f \geq 0), \quad (7.2)$$

$$\mathcal{P}_{\text{cal}}(a) = \frac{\lambda_{\text{NA}}}{2} \|(-A_{\tau|K} a)_+\|_2^2 \quad (\text{targets } (\partial_{\tau} C_f)|_K \geq 0), \quad (7.3)$$

$$\mathcal{P}_{\text{bnd}}(a) = \frac{\lambda_{\text{B}}}{2} \left(\|(-Aa)_+\|_2^2 + \|(Aa - F)_+\|_2^2 \right) \quad (\text{targets } 0 \leq C_f \leq F). \quad (7.4)$$

These are sums of squares of convex functions of an affine map of a , hence convex and QP-compatible.

QP form (auxiliary slacks). Exactly as in the band-hinge reformulation, each penalty admits a slack representation. For example,

$$\frac{1}{2} \|(A_K a)_+\|_2^2 = \min_{u \in \mathbb{R}^G} \frac{1}{2} \|u\|_2^2 \quad \text{s.t.} \quad u \geq A_K a, \quad u \geq 0,$$

and similarly

$$\frac{1}{2} \|(-A_{KK} a)_+\|_2^2 = \min_{v \geq 0, v \geq -A_{KK} a} \frac{1}{2} \|v\|_2^2, \quad \frac{1}{2} \|(-A_{\tau|K} a)_+\|_2^2 = \min_{w \geq 0, w \geq -A_{\tau|K} a} \frac{1}{2} \|w\|_2^2,$$

and for bounds

$$\frac{1}{2} \|(-Aa)_+\|_2^2 + \frac{1}{2} \|(Aa - F)_+\|_2^2 = \min_{s, t \geq 0, s \geq -Aa, t \geq Aa - F} \frac{1}{2} (\|s\|^2 + \|t\|^2).$$

Lemma 7. For any $x \in \mathbb{R}^G$,

$$\frac{1}{2} \|(x)_+\|_2^2 = \min_{u \in \mathbb{R}^G} \left\{ \frac{1}{2} \|(u)_+\|_2^2 : u \geq x, u \geq 0 \right\}$$

and the unique minimiser is $u^* = (x)_+$.

Proof. The problem separates across coordinates. For scalar $u \in \mathbb{R}$,

$$\min_{u \in \mathbb{R}} \frac{1}{2}u^2 \quad \text{st} \quad u \geq x, u \geq 0$$

has feasible set $u \geq \max\{x, 0\}$. The objective $\frac{1}{2}u^2$ is strictly increasing on $[0, \infty)$, so the minimum is attained at the smallest feasible point:

$$u^* = \max\{x, 0\} = x_+,$$

with value $\frac{1}{2}(x_+)^2$. Summing over coordinates gives the vector result, and strict convexity yields uniqueness. $\square$

7.2 Row scaling and invariance

As in §9, we scale each block (after any U -reparameterisation) by a positive scalar so typical row norms are comparable:

$$\tilde{A}_K = \frac{1}{s_K} A_K, \quad \tilde{A}_{KK} = \frac{1}{s_{KK}} A_{KK}, \quad \tilde{A}_{\tau|K} = \frac{1}{s_\tau} A_{\tau|K},$$

with $s_\bullet$ the empirical p95 of row ℓ_2 norms (4.5). Hard constraints are invariant under positive row/ block scaling, and with soft penalties this makes a single λ_{NA} control all three terms on a comparable numeric scale.

7.3 Compatibility with the Λ -module

Under the price-invariant reparameterisation $a = U\tilde{a}$ (see the Λ -module), all operators post-multiply by U :

$$A_\bullet \leftarrow A_\bullet U, \quad \bullet \in \{, K, KK, \tau|K\}.$$

By construction, $A_\bullet U \tilde{a} = A_\bullet a$, so the penalty values are unchanged and convexity is preserved. The spectral ridge is updated by congruence $\Lambda \mapsto \tilde{\Lambda} = U^\top \Lambda U$ as already stated.

7.4 Binned variant (optional)

To stabilise very sparse regions, aggregate quotes by a selector $G \in \{0, 1\}^{B \times N}$ (bins in (m, τ)) and replace the per-quote band-hinge term $\sum_i \ell_{\text{band}}((Aa)_i; b_i, a_i)$ by

$$\sum_{b=1}^B \ell_{\text{band}}((GAa)_b; (Gb)_b, (Ga)_b);$$

the slack QP form carries over verbatim.

After scaling, we use a single λ_{NA} for $(\tilde{A}_K, \tilde{A}_{KK}, \tilde{A}_{\tau|K})$ and select it (once per date) to reach $\leq 1\%$ violations on the evaluation grid; λ_{B} is kept separate for price bounds.

8. THE CONVEX PROGRAM

We collect all terms and write the problem as a single quadratic program (QP). When the Λ -module is active (6.2), we solve in $\tilde{a} = U^{-1}a$ with all blocks post-multiplied by U and Λ replaced by $\tilde{\Lambda} = U^\top \Lambda U$; to avoid clutter we keep the symbol a below (read as $\tilde{a}$ in that case).

8.1 Slack QP (standard form)

Let $u, v \in \mathbb{R}^N$ be the band slacks from §5.2, and let $u_K, v_{KK}, w_\tau \in \mathbb{R}^G$ and $s_{\text{lo}}, s_{\text{hi}} \in \mathbb{R}^G$ be nonnegative grid slacks for the three shape operators and price bounds, respectively (all inequalities componentwise):

$$\begin{aligned} u &\geq b - Aa, & u &\geq 0, & v &\geq Aa - a, & v &\geq 0, \\ u_K &\geq A_K a, & u_K &\geq 0, & v_{KK} &\geq -A_{KK} a, & v_{KK} &\geq 0, \\ w_\tau &\geq -A_{\tau|K} a, & w_\tau &\geq 0, & & & & \\ s_{\text{lo}} &\geq -Aa, & s_{\text{lo}} &\geq 0, & s_{\text{hi}} &\geq Aa - F, & s_{\text{hi}} &\geq 0. \end{aligned}$$

With these slacks, the objective collects the data term (§5), the quadratic regularisers (§6), and the soft no-arb penalties (§7.1):

$$\begin{aligned} \min_{a, u, v, u_K, v_{KK}, w_\tau, s_{\text{lo}}, s_{\text{hi}}} & \frac{1}{2} \|W^{1/2}(Aa - y)\|_2^2 + \frac{\mu}{2} (\|u\|_2^2 + \|v\|_2^2) \\ & + \frac{\lambda_{\text{ridge}}}{2} a^\top \Lambda a + \frac{\lambda_{\text{DW}}}{2} a^\top E^\top L^+ E a + \frac{\eta_{\text{RN}}}{2} \|A_{\tau|K} a\|_{2, \Gamma_0^+}^2 \\ & + \frac{\lambda_\Omega}{2} a^\top U_\omega^\top M_\omega^\top M_\omega U_\omega a + \frac{\lambda_{\text{hook}}}{2} \|C(Aa)\|_2^2 \\ & + \frac{\lambda_{\text{RN}}}{2} \|Aa - C_{0^+}\|_{2, \Gamma_0^+}^2 \\ & + \frac{\lambda_{\text{NA}}}{2} (\|u_K\|_2^2 + \|v_{KK}\|_2^2 + \|w_\tau\|_2^2) + \frac{\lambda_{\text{B}}}{2} (\|s_{\text{lo}}\|_2^2 + \|s_{\text{hi}}\|_2^2). \end{aligned} \tag{8.1}$$

All matrices ($A, A_K, A_{KK}, A_{\tau|K}, E, U_\omega, M_\omega, L^+, C$) are fixed from earlier sections; $\|\cdot\|_{2, \Gamma_0^+}$ denotes restriction to the short-maturity index set. Row-scaled operators (§9) may be used in place of unscaled ones.

Why (8.1) is a QP: Every term in the objective is a convex quadratic form in a or a sum of squared slacks; all constraints are linear inequalities.

8.2 Convexity, existence, and uniqueness

Proposition 7 (Convexity and global optimality). *The program (8.1) is convex. If a convex QP solver returns a feasible primal-dual point satisfying KKT, then the associated a^* is a global minimiser.*

Proof. The feasible set is a polyhedron (linear inequalities), hence convex and closed. The objective is a sum of convex quadratics, hence convex and lower semicontinuous. KKT conditions are necessary and sufficient for convex QPs; any feasible KKT point is globally optimal. $\square$

Proposition 8 (Strict convexity conditions and uniqueness). *Define $C := \tilde{A}_{\tau|K} \tilde{A}_K - \tilde{A}_K \tilde{A}_{\tau|K} \in \mathbb{R}^{G \times G}$. and Π_{Γ_0+} selects the short maturity grid. If the quadratic form in a ,*

$$Q = A^\top W A + \lambda_{\text{ridge}} \Lambda + \lambda_{\text{DW}} E^\top L^+ E + \lambda_\Omega U_\omega^\top M_\omega^\top M_\omega U_\omega + \lambda_{\text{hook}} A^\top C^\top C A + \lambda_{\text{RN}} A^\top \Pi_{0+} A + \eta_{\text{RN}} A_{\tau|K}^\top \Pi_{0+} A_{\tau|K}, \quad (8.2)$$

is positive definite, then the objective is strictly convex in $(a, u, v, \dots)$ and the minimiser is unique. In particular, it suffices that $A^\top W A$ be positive definite on $\mathcal{R}(A)$ and $\lambda_{\text{ridge}} > 0$ (cf. §5.2).

Proof. Block Hessian is $\text{blkdiag}(Q, \mu I, \mu I, \lambda_{\text{NA}} I, \dots)$; if $Q \succ 0$ and $\mu, \lambda_{\text{NA}}, \lambda_{\text{B}} > 0$, the whole Hessian is positive definite. $\square$

Definition 4 (Global coefficient-space metric). Define the symmetric positive *semidefinite* matrix

$$M_h := A^\top W A + \lambda_{\text{ridge}} \Lambda + \lambda_{\text{DW}} E^\top L^+ E + \lambda_\Omega U_\omega^\top M_\omega^\top M_\omega U_\omega + \lambda_{\text{hook}} A^\top C^\top C A + \lambda_{\text{RN}} A^\top \Pi_{0+} A + \eta_{\text{RN}} A_{\tau|K}^\top \Pi_{0+} A_{\tau|K}.$$

Assumption 1 (Strictly positive ridge shape). *The ridge shape matrix Λ is symmetric positive definite (e.g. diagonal with strictly positive entries).*

Proposition 9 (When M_h is positive definite). *Suppose Assumption 1 holds and $\lambda_{\text{ridge}} > 0$. Then the matrix M_h from Definition 4 is symmetric positive definite. In particular, M_h is invertible and induces a norm $\|a\|_{M_h}^2 := a^\top M_h a$ on $\mathbb{R}^P$.*

Proof. Each term in the definition of M_h is symmetric and positive semidefinite. Under Assumption 1, $\Lambda \succ 0$, so for any $a \neq 0$ we have

$$a^\top (\lambda_{\text{ridge}} \Lambda) a = \lambda_{\text{ridge}} a^\top \Lambda a > 0$$

whenever $\lambda_{\text{ridge}} > 0$. All the remaining terms in M_h are positive semidefinite, so

$$a^\top M_h a = a^\top (\lambda_{\text{ridge}} \Lambda) a + a^\top (\text{psd terms}) a \geq \lambda_{\text{ridge}} a^\top \Lambda a > 0$$

for all $a \neq 0$. Hence $M_h \succ 0$. $\square$

Proposition 10 (Metric projection form of the global solution). *Let $\mathcal{C}_h \subset \mathbb{R}^P$ be the polyhedron defined by the hard constraints (if any) in a (no-arbitrage, bounds, etc.). Consider the strictly convex quadratic problem*

$$\min_{a \in \mathcal{C}_h} \frac{1}{2} a^\top M_h a - b^\top a, \quad (8.3)$$

where M_h is the matrix from Definition 4 and

$$b := A^\top W y - c_{\text{RN}}, \quad (8.4)$$

with c_{RN} the linear coefficient from Lemma 5.

Set

$$\hat{a} := M_h^{-1}b. \quad (8.5)$$

Then the unique minimiser a^* of (8.3) can be written as the metric projection of $\hat{a}$ onto $\mathcal{C}_h$ in the M_h -inner product:

$$a^* = \arg \min_{a \in \mathcal{C}_h} \frac{1}{2} \|a - \hat{a}\|_{M_h}^2, \quad \|z\|_{M_h}^2 := z^\top M_h z. \quad (8.6)$$

Proof. Write the objective in (8.3) as

$$J(a) := \frac{1}{2} a^\top M_h a - b^\top a, \quad a \in \mathcal{C}_h.$$

By Definition 4 the matrix M_h is symmetric and positive semidefinite. Under the assumptions of Proposition 8 (in particular, $A^\top W A$ positive definite on the span of A and $\lambda_{\text{ridge}} > 0$), the quadratic form $a \mapsto a^\top M_h a$ is positive definite, and hence M_h is symmetric positive definite. In particular, M_h is invertible, so $\hat{a}$ is well defined by

$$\hat{a} := M_h^{-1}b \iff M_h \hat{a} = b.$$

We first rewrite J in terms of $(a - \hat{a})$. Using $b = M_h \hat{a}$ and the symmetry of M_h , for any $a \in \mathbb{R}^P$,

$$J(a) = \frac{1}{2} a^\top M_h a - b^\top a = \frac{1}{2} a^\top M_h a - (M_h \hat{a})^\top a = \frac{1}{2} a^\top M_h a - \hat{a}^\top M_h a.$$

On the other hand,

$$(a - \hat{a})^\top M_h (a - \hat{a}) = a^\top M_h a - 2 \hat{a}^\top M_h a + \hat{a}^\top M_h \hat{a},$$

again by symmetry of M_h . Hence,

$$\frac{1}{2} (a - \hat{a})^\top M_h (a - \hat{a}) - \frac{1}{2} \hat{a}^\top M_h \hat{a} = \frac{1}{2} a^\top M_h a - \hat{a}^\top M_h a = J(a).$$

Thus, for all $a \in \mathbb{R}^P$,

$$J(a) = \frac{1}{2} (a - \hat{a})^\top M_h (a - \hat{a}) - \frac{1}{2} \hat{a}^\top M_h \hat{a}.$$

The second term on the right-hand side does not depend on a . Therefore, minimising J over $a \in \mathcal{C}_h$ is equivalent to minimising

$$a \mapsto \frac{1}{2} (a - \hat{a})^\top M_h (a - \hat{a}) = \frac{1}{2} \|a - \hat{a}\|_{M_h}^2$$

over $a \in \mathcal{C}_h$. Since M_h is positive definite, this functional is strictly convex in a , so it has a unique minimiser in the closed convex set $\mathcal{C}_h$; by definition, this minimiser is the metric projection of $\hat{a}$ onto $\mathcal{C}_h$ in the M_h -inner product. This is exactly the claim. $\square$

When discussing the hard-constraint limit it will be convenient to make the global feasibility assumption explicit. Namely, we assume that the static no-arbitrage and bound inequalities admit at least one coefficient vector, i.e.

$$\mathcal{C}_{\text{NA}} := \{a \in \mathbb{R}^P : A_K a \leq 0, -A_{KK} a \leq 0, -A_{\tau|K} a \leq 0, 0 \leq Aa \leq F\} \neq \emptyset. \quad (8.7)$$

This is a modelling condition stating that the chosen Chebyshev approximation space contains at least one globally static no-arbitrage surface.

Hard-constraint limits. Hard-constraint limits. Under (8.7), letting $\lambda_{\text{NA}}, \lambda_B \rightarrow \infty$ drives the corresponding slacks to 0 and recovers the constrained solution of the remaining strictly convex quadratic objective. If the intersection is empty, the finite- λ problem yields the minimum-violation compromise (§7.1).

8.3 Invariance and scaling

If the Λ -module is used (§6.2), replace every block by its multiplied version and Λ by $\tilde{\Lambda} = U^\top \Lambda U$; the feasible set and all objective values are unchanged (Proposition 5). Row-scaling the no-arb blocks (§9) multiplies them by positive scalars and only equilibrates numeric weights; it does not alter feasibility.

8.4 Solution procedure (used)

Solve (8.1) with OSQP (warm starts). The no-arb weight λ_{NA} is set after row scaling to hit $\leq 1\%$ grid violations (§9); λ_{ridge} is fixed by GCV on a small WLS subsample (§6.1); μ is increased by a short controller until target coverage is reached (§5). All other quadratic weights follow §6.

Remark 15 (Why the 99/1 target is attainable). The data term prices *into* the bid-ask bands (hinge), while $A_K, A_{KK}, A_{\tau|K}$ are enforced densely on the grid with p95 row scaling, so a single λ_{NA} controls the violation budget. Short-maturity defects are suppressed by the RN anchoring and the DW/ Ω terms, which remove the usual butterfly/aliasing artifacts.

9. SCALING, SCHEDULES, AND THE μ -CONTROLLER

9.1 Row scaling (summary)

We use the p95 block-scalar normalisation of §4.5: on the grid, set $\tilde{A}_K = A_K/s_K$, $\tilde{A}_{KK} = A_{KK}/s_{KK}$, $\tilde{A}_{\tau|K} = A_{\tau|K}/s_\tau$ with $s_\bullet = q_{0.95}$ of row ℓ_2 norms, computed *after* the Λ -module transform. This preserves hard feasibility and allows a single λ_{NA} to control all three terms on a comparable scale.

9.2 Short schedule for λ_{NA}

We select λ_{NA} on a thinned setup to save time while preserving the target violation share.

Thinned probe: Build a reduced grid (every other Chebyshev node in m and a coarser subset in τ) and a tiny quote subset (5-10% uniformly across (m, τ)). Fix a moderate μ (the previous day's value) and all other weights.

Grid search: For a short geometric ladder $\Lambda_{\text{trial}} = \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$ solve the QP on the thinned setup and measure the violation rate:

$$\begin{aligned} \text{viol}(\lambda) = \frac{1}{3G} \sum_{g=1}^G & \left(\mathbf{1}\{(A_K a(\lambda))_g > \tau_K\} \right. \\ & \left. + \mathbf{1}\{(-A_{KK} a(\lambda))_g > \tau_{KK}\} + \mathbf{1}\{(-A_{\tau|K} a(\lambda))_g > \tau_\tau\} \right). \end{aligned} \quad (9.1)$$

with small numerical tolerances $\tau_\bullet$ (in scaled units). Pick the smallest $\lambda \in \Lambda_{\text{trial}}$ such that $\text{viol}(\lambda) \leq 1\%$, and *fix* that λ_{NA} for the full grid and book.

Explanation: Solve the QP (8.1) repeatedly, but with $\lambda_{\text{NA}} \leftarrow \lambda$ for each $\lambda \in \Lambda_{\text{trial}}$ and all other weights fixed. Each solve returns a different $a(\lambda)$ and then using this $a(\lambda)$, compute the violation rate $\text{viol}(\lambda)$. From the violation rates pick the smallest λ achieving $\leq 1\%$, and use that as λ_{NA} for the full problem.

Remark 16 (Invariance to scaling). Because each block was divided by s_B , the selected λ_{NA} is stable day-to-day and across underliers; without scaling, the same ladder would over/under-penalise whichever block happens to have the largest raw norms.

9.3 The μ -controller (coverage target)

Recall the coverage-seeking data term (§5): the mid-squared error plus μ times the quadratic band-hinge. Let

$$\text{Hinge}(a) := \sum_{i=1}^N \ell_{\text{band}}((Aa)_i; b_i, a_i) = \frac{1}{2} \|\text{dist}(Aa, [b, a])\|_2^2,$$

$$\text{Cov}(a) := \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{b_i \leq (Aa)_i \leq a_i\}.$$

We adjust μ (similar to 9.2) so that $\text{Cov}(a^*(\mu))$ reaches a target (99%). Although coverage is a discrete functional (hence may have plateaus), the hinge at the optimiser is nonincreasing in μ :

Lemma 8 (Monotonicity of optimal hinge). *Let $g(a)$ denote the full objective without the hinge weight (all terms in (8.1) except $\mu \text{Hinge}(a)$). For $\mu_1 < \mu_2$, let $a_j \in \arg \min_a \{g(a) + \mu_j \text{Hinge}(a)\}$ for $j = 1, 2$. Then $\text{Hinge}(a_2) \leq \text{Hinge}(a_1)$.*

Proof. By definition of the minimisers, for all x, y we have the following:

$$g(a_1) + \mu_1 H(a_1) \leq g(x) + \mu_1 H(x);$$

$$g(a_2) + \mu_2 H(a_2) \leq g(y) + \mu_2 H(y).$$

Taking $x = a_2$ and $y = a_1$ yields the following:

$$g(a_1) + \mu_1 H(a_1) \leq g(a_2) + \mu_1 H(a_2)$$

$$g(a_2) + \mu_2 H(a_2) \leq g(a_1) + \mu_2 H(a_1)$$

Summing and dividing by $\mu_2 - \mu_1 > 0$ yields $H(a_2) \leq H(a_1)$. □

Controller (bracket & bisection).

1. *Bracket.* Start from $(\mu_{\min}, \mu_{\max})$ (reusing prior-day values when available). If coverage at $\mu_{\max}$ is below target, expand $\mu_{\max} \leftarrow c \mu_{\max}$ (e.g. $c = 4$) until $\text{Cov}(a^*(\mu_{\max}))$ crosses the target or a cap is reached.
2. *Bisection.* While $\mu_{\max} - \mu_{\min}$ is above tolerance and coverage not yet at target, set $\mu \leftarrow \sqrt{\mu_{\min} \mu_{\max}}$ (geometric bisection), solve once, and update the endpoint whose coverage is on the wrong side of the target.

What is held fixed: All other weights ($\lambda_{\text{ridge}}, \lambda_{\text{NA}}, \lambda_{\text{DW}}, \lambda_{\Omega}, \lambda_{\text{RN}}, \lambda_{\text{B}}$) and the scaled operators are held fixed while μ is adjusted.

9.4 Practical notes

- *Warm starts.* Reuse a^* when moving along the λ_{NA} ladder and the μ bracket; OSQP converges in a few iterations from a nearby point.
- *Tolerances.* Use small positive tolerances $\tau_{\bullet}$ when counting violations to avoid flagging solver noise; report violations in unscaled operator units.
- *Stability.* If coverage oscillates near the target, accept the smallest μ in the final bracket that achieves the target.

10. SHORT-MATURITY REMEDY

Recall the calendar operator at fixed strike from §4 (see (4.3)):

$$A_{\tau|K} = A_\tau + \text{diag}(-\rho(\tau_g)) A_m, \quad \rho(\tau) \equiv \frac{d}{d\tau} \log F(\tau).$$

When τ is very small, noise in $\partial_m C_f$ is fed into $(\partial_\tau C_f)|_K$ through the $\rho(\tau)$ term, so small ripples in m can flip the calendar sign. Counter this with three convex, model-agnostic devices that act only near the boundary and vanish smoothly as maturity grows.

10.1 Boundary anchoring and calendar flattening on Γ_{0+}

Let $\Gamma_{0+} = \{g : \tau_g \leq \tau_\star\}$ be the short-maturity window (usually $\tau_\star = 5$ –10 trading days). Define the intrinsic forward-discounted limit

$$C_{0+}(m) = F_0(1 - e^m)_+ = (F_0 - K)^+, \quad m = \log(K/F_0).$$

We use the convex quadratic (see 6.3)

$$\mathcal{R}_{\text{RN}}(a) = \frac{\lambda_{\text{RN}}}{2} \|Aa - C_{0+}\|_{2, \Gamma_{0+}}^2 + \frac{\eta_{\text{RN}}}{2} \|A_{\tau|K}a\|_{2, \Gamma_{0+}}^2. \quad (10.1)$$

Lemma 9 (Consistency with the short-time limit). *Assume that for fixed $K \neq F_0$, $C_f(K, \tau) \rightarrow (F_0 - K)^+$ and $\partial_\tau C_f(K, \tau)|_K \rightarrow 0$ as $\tau \downarrow 0$. Let $\Gamma_{0+} = \{g : \tau_g \leq \tau_\star\}$ and write*

$$R_1(a) := \|Aa - C_{0+}\|_{2, \Gamma_{0+}}^2, \quad R_2(a) := \|A_{\tau|K}a\|_{2, \Gamma_{0+}}^2,$$

so that the RN penalty in (10.1) is $\frac{\lambda_{\text{RN}}}{2} R_1(a) + \frac{\eta_{\text{RN}}}{2} R_2(a)$. Let $\mathcal{F} \subset \mathbb{R}^P$ be a closed convex feasible set and let $g : \mathcal{F} \rightarrow \mathbb{R}$ collect all other (convex) terms of the objective in (8.1). Assume the boundary conditions are attainable on Γ_{0+} , i.e.

$$\mathcal{C} := \{a \in \mathcal{F} : R_1(a) = 0, R_2(a) = 0\} \neq \emptyset.$$

For $\lambda, \eta > 0$ define

$$J_{\lambda, \eta}(a) := g(a) + \frac{\lambda}{2} R_1(a) + \frac{\eta}{2} R_2(a), \quad a_{\lambda, \eta} \in \arg \min_{a \in \mathcal{F}} J_{\lambda, \eta}(a).$$

Then, as $\min\{\lambda, \eta\} \rightarrow \infty$,

$$R_1(a_{\lambda, \eta}) \rightarrow 0 \quad \text{and} \quad R_2(a_{\lambda, \eta}) \rightarrow 0,$$

so $Aa_{\lambda, \eta} \rightarrow C_{0+}$ and $A_{\tau|K}a_{\lambda, \eta} \rightarrow 0$ on Γ_{0+} . Moreover, every cluster point $a^\star$ of $\{a_{\lambda, \eta}\}$ solves the equality-constrained problem

$$\min\{g(a) : a \in \mathcal{C}\}.$$

Proof. Step 1 (residuals vanish). Pick $a^0 \in \mathcal{C}$ (exists by assumption, discussed later), so $R_1(a^0) = R_2(a^0) = 0$. By optimality of $a_{\lambda,\eta}$,

$$J_{\lambda,\eta}(a_{\lambda,\eta}) \leq J_{\lambda,\eta}(a^0) = g(a^0).$$

Hence, for all $\lambda, \eta > 0$,

$$g(a_{\lambda,\eta}) + \frac{\lambda}{2}R_1(a_{\lambda,\eta}) + \frac{\eta}{2}R_2(a_{\lambda,\eta}) \leq g(a^0).$$

Since g is bounded below on $\mathcal{F}$ (true in our QP, e.g. by the ridge term; (Remark 18)), there exists $m > -\infty$ with $g(a) \geq m$ for all $a \in \mathcal{F}$. Therefore

$$\frac{\lambda}{2}R_1(a_{\lambda,\eta}) + \frac{\eta}{2}R_2(a_{\lambda,\eta}) \leq g(a^0) - g(a_{\lambda,\eta}) \leq g(a^0) - m =: C < \infty.$$

Let $\min\{\lambda, \eta\} \rightarrow \infty$. The left-hand side is a sum of nonnegative terms with coefficients diverging to $+\infty$, so necessarily

$$R_1(a_{\lambda,\eta}) \rightarrow 0 \quad \text{and} \quad R_2(a_{\lambda,\eta}) \rightarrow 0.$$

By linearity of A and $A_{\tau|K}$, this yields $Aa_{\lambda,\eta} \rightarrow C_{0+}$ and $A_{\tau|K}a_{\lambda,\eta} \rightarrow 0$ on Γ_{0+} .

Step 2 (limit points solve the constrained problem). Because g is convex and (by the ridge) coercive on $\mathcal{F}$, the sequence $\{a_{\lambda,\eta}\}$ is bounded; thus it has cluster points. Let $a_{\lambda,\eta} \rightarrow a^*$ along some subsequence. From Step 1 and continuity of the linear maps, $a^* \in \mathcal{C}$.

For any $a \in \mathcal{C}$, we have $J_{\lambda,\eta}(a) = g(a)$, hence

$$J_{\lambda,\eta}(a_{\lambda,\eta}) \leq J_{\lambda,\eta}(a) = g(a) \quad \Rightarrow \quad g(a_{\lambda,\eta}) \leq g(a) \quad \forall \lambda, \eta.$$

Taking lim sup and using lower semicontinuity of g ,

$$g(a^*) \leq \liminf g(a_{\lambda,\eta}) \leq \limsup g(a_{\lambda,\eta}) \leq g(a) \quad \forall a \in \mathcal{C},$$

so $g(a^*) = \min_{x \in \mathcal{C}} g(x)$. This completes the proof. $\square$

Remark 17. If exact feasibility on Γ_{0+} is relaxed (e.g. excluding an ATM tube $|m| \leq c\sqrt{\tau}$), interpret R_1, R_2 with that restriction; the same argument applies. If $\mathcal{C} = \emptyset$, then $R_1(a_{\lambda,\eta})$ and $R_2(a_{\lambda,\eta})$ converge to their joint infimum and $a_{\lambda,\eta}$ approaches the minimum-violation compromise.

Remark 18 (Lower bound for g). In (8.1) the function $g(a)$ (all terms except the RN penalty) is a sum of positive-semidefinite quadratics and squared norms with positive weights: $\frac{1}{2}\|W^{1/2}(Aa - y)\|_2^2$, $\frac{\lambda_{\text{ridge}}}{2}a^\top \Lambda a$, $\frac{\lambda_{\text{DW}}}{2}a^\top E^\top L^+ E a$, $\frac{\lambda_{\Omega}}{2}a^\top U_\omega^\top M_\omega^\top M_\omega U_\omega a$, etc. After dropping the constant $\frac{1}{2}y^\top W y$ from the LS term, we have $g(a) \geq 0$ for all $a \in \mathcal{F}$. Hence $m := \inf_{a \in \mathcal{F}} g(a) \geq 0$ is finite, which is the bound used in the proof of Lemma 9.

Remark 19 (Attainability is an assumption). Lemma 9 assumes the boundary conditions are attainable on Γ_{0+} , i.e. $\mathcal{C} := \{a \in \mathcal{F} : R_1(a) = 0, R_2(a) = 0\} \neq \emptyset$. This is not automatic; it depends on the basis and the feasible set $\mathcal{F}$.

Proposition 11 (Sufficient discrete condition for attainability). *Let $A_\Gamma \in \mathbb{R}^{G_0 \times P}$ and $A_{\tau|K,\Gamma} \in \mathbb{R}^{G_0 \times P}$ denote A and $A_{\tau|K}$ restricted to the rows $g \in \Gamma_{0+}$ (with $G_0 := |\Gamma_{0+}|$). Stack the constraints into*

$$S := \begin{bmatrix} A_\Gamma \\ A_{\tau|K,\Gamma} \end{bmatrix} \in \mathbb{R}^{(2G_0) \times P}, \quad c := \begin{bmatrix} C_{0+} \\ 0 \end{bmatrix} \in \mathbb{R}^{2G_0}.$$

If $\text{rank}(S) = 2G_0$ (full row rank) and $\mathcal{F} = \mathbb{R}^P$ (or $\mathcal{F}$ is any convex set that contains a solution of $Sa = c$), then $\mathcal{C} \neq \emptyset$. In particular, when $P \geq 2G_0$ and S has full row rank, the system $Sa = c$ is solvable and any solution a satisfies $R_1(a) = R_2(a) = 0$.

Proof. If $\text{rank}(S) = 2G_0 \leq P$, then $\text{Range}(S) = \mathbb{R}^{2G_0}$, so for any right-hand side c there exists $a \in \mathbb{R}^P$ with $Sa = c$. Such an a obeys $A_\Gamma a = C_{0+}$ and $A_{\tau|K,\Gamma} a = 0$, hence $R_1(a) = R_2(a) = 0$; if $\mathcal{F}$ contains one such a , then $a \in \mathcal{C}$. $\square$

Corollary 1 (Practical sufficient conditions). *Attainability holds if (i) the coefficient space is rich enough so that $P \geq 2G_0$ and the slice-restricted design matrices have independent rows (so S has full row rank), and (ii) the feasible set $\mathcal{F}$ does not exclude these solutions. For Chebyshev tensor bases in (m, τ) , increasing degrees (K, L) makes S generically full row rank on a fixed grid Γ_{0+} ; moreover, the boundary value $C_{0+} \in [0, F_0]$ and the condition $A_{\tau|K,\Gamma} a = 0$ are compatible with usual hard constraints (monotonicity/convexity and bounds).*

Remark 20 (If attainability fails). If $\mathcal{C} = \emptyset$ (e.g. degrees too low, or additional constraints forbid exact matching), then the conclusions of Lemma 9 hold in the approximate sense: $R_1(a_{\lambda,\eta}) \rightarrow R_1^*$ and $R_2(a_{\lambda,\eta}) \rightarrow R_2^*$ where $R_1^* + R_2^* = \inf_{a \in \mathcal{F}} \{R_1(a) + R_2(a)\}$; the minimisers converge to the minimum-violation compromise on Γ_{0+} .

ATM tube (optional). To avoid over-penalizing the thin region where the $O(\sqrt{\tau})$ time value concentrates, one may exclude $|m| \leq c\sqrt{\tau}$ from Γ_{0+} (small c) or down-weight those rows.

Parity projection (optional). As a linear preprocessing that preserves QP structure, set $a \leftarrow P_{\text{even}} a$ with $(P_{\text{even}})_{(k,\ell),(k,\ell)} = 1$ for even k and 0 for odd k when evaluating the first term in (10.1); this removes odd-in- m glitches near $\tau \downarrow 0$.

10.2 Frequency truncation: Ω taper near the boundary

Use the spectral mask from §6.5, but only on short maturities. Let U_ω be the fixed frequency chart (e.g. separable 2D DCT) and define a maturity-dependent mask $M_\omega(\tau_g)$ that zeros the highest m -frequencies on slices with $\tau_g \leq \tau_*$ and ramps to 0 by $2\tau_*$:

$$\mathcal{R}_\Omega(a) = \frac{1}{2} \sum_{g: \tau_g \text{ grid}} \lambda_\Omega(\tau_g) \|M_\omega(\tau_g) U_\omega a\|_2^2,$$

$$\lambda_\Omega(\tau) = \lambda_\Omega^{(0)} \times \begin{cases} 1, & \tau \leq \tau_*, \\ 2 - \tau/\tau_*, & \tau_* < \tau \leq 2\tau_*, \\ 0, & \tau > 2\tau_*. \end{cases}$$

This convex quadratic suppresses only the high- k content where butterfly ripples originate; low modes (ATM/term structure) are left intact by construction.

10.3 Transport damping: H^{-1} smoothing of density at short maturities

Let $\rho = \partial_{KK} C_f$ and recall the discrete H^{-1} seminorm from §6.3: $\|f\|_{H^{-1}}^2 = f^\top L^+ f$ with $L^+ \succeq 0$ fixed. We weight it more at short τ :

$$\mathcal{R}_{\text{DW}}(a) = \frac{1}{2} \sum_{g: \tau_g \text{ grid}} \lambda_{\text{DW}}(\tau_g) \|(Ea)_g\|_{H^{-1}}^2, \quad \lambda_{\text{DW}}(\tau) = \lambda_{\text{DW}}^{(0)} \min\{1, \tau_\star/\tau\},$$

where Ea stacks $A_{KK}a$ slice-wise. This biases the optimiser toward *spreading* density rather than oscillating it in short strips, eliminating spurious negative lobes.

10.4 Convexity and invariance

Each addend in this chapter is a sum of squares of affine functions of a (or a congruence of a fixed SPD quadratic), hence convex and QP-compatible. Under the Λ -module (§6.2), post-multiply all blocks by U and replace Λ by $\tilde{\Lambda} = U^\top \Lambda U$; penalty values on $(Aa, A_{\tau|K}a, A_{KK}a)$ are unchanged.

10.5 Practical choices and interaction with scaling

- **Window:** $\tau_\star = 5$ –10 trading days. Apply (10.1) only on Γ_{0+} ; optionally exclude an ATM tube $|m| \leq c\sqrt{\tau}$.
- **Weights:** Choose λ_{RN} so that the RMS of $Aa - C_{0+}$ on Γ_{0+} matches the median band width there; choose η_{RN} to bring short-end calendar violations under the global 1% budget when combined with the scaled no-arb penalties (§9). Use the short-maturity ramps $\lambda_\Omega(\tau)$ and $\lambda_{\text{DW}}(\tau)$ above.
- **Row scaling:** Apply the same p95 row scaling (§9) to $A_{\tau|K}$ when used inside (10.1); this keeps the knob η_{RN} comparable to λ_{NA} .

11. DIAGNOSTICS AND IMPLEMENTATION

11.1 Structure monitors

We carry a set of diagnostics that do not enter the optimisation but certify stability.

MON1 (symplectic/volume/reversibility). If an auxiliary Hamiltonian stepper is used (e.g. to generate U or kicks), we compute the discrete symplectic defect and the map-determinant on random probes. Both are monitor scalars and must remain below pre-set tolerances.

MON2 (RN residual/commuting defect). Report *RN residual* $\|Aa - C_{0+}\|_{\Gamma_{0+}}$, and *calendar commutator* $\|C(Aa)\|_2$ on the grid, where $C := \tilde{A}_{\tau|K}\tilde{A}_K - \tilde{A}_K\tilde{A}_{\tau|K} \in \mathbb{R}^{G \times G}$ is the grid-space commutator from Section 6.5.

MON3 (aliasing). Let $\mathcal{E}_{\text{hi}}$ be the share of modal energy in the upper third of (k, ℓ) indices. Large values predict convexity noise; the Ω penalty is tuned to cap $\mathcal{E}_{\text{hi}}$.

Q (Egorov bridge). Under short-time linearisation, classical transport of observables commutes with quantum propagation up to $\mathcal{O}(\tau)$ (Egorov’s theorem). We monitor the deviation of C_f pushed through the (linearised) forward drift versus the surface rebuilt at $\tau + \delta\tau$; the resulting defect is reported as absolute/relative scalars. (Purely diagnostic; no constraints are added.)

11.2 Implementation notes

- **Basis sizes:** $K \in [28, 40]$, $L \in [22, 32]$; grid $G \approx (2K) \times (2L)$.
- **Scaling:** row p95 to 1; convexity block optionally scaled by a small boost ($\times 2$ –5).
- **Penalties:** λ_{NA} via short probe; λ_{ridge} by GCV on a random 8% subsample.
- **Short-maturity:** set $\tau_\star \approx 5$ –10 trading days; ramp $\lambda_{\text{RN}}, \lambda_\Omega, \lambda_{\text{DW}}$ to zero after $2\tau_\star$.
- **Solver:** OSQP with $\varepsilon_{\text{abs}} = \varepsilon_{\text{rel}} = 4.5 \times 10^{-7}$, adaptive ρ (default 0.1), polishing enabled, and a 900-second time cap; other options stay at their defaults.

12. HAMILTONIAN FOG POST-FIT IN PRICE SPACE

The global Chebyshev–QP fit constructed in the previous Chapters already enforces static no–arbitrage on a dense collocation grid and is tuned to achieve approximately 99% within–band coverage with at most 1% grid violations. However, some trading dates and regions of the (m, τ) –plane remain unsatisfactory even after the main fit and the short–maturity remedy.

The main QP already delivers a strong baseline surface on most dates. For *calm* years such as 2019 and 2022–23, a single choice of μ and λ_{NA} is enough to reach the target. In these regimes, the badness field ω is close to zero everywhere, so any local post-fit is minimal.

However, in 2020–21 quotes are noisier, and cross-sectional inconsistencies between strikes and maturities are more common. On these stressed dates, the baseline QP is forced into local compromises (either coverage or no-arbitrage deteriorates). Empirically, the misfit is concentrated in small regions.

In this chapter, we describe a second *local* post-fit layer that takes, for each trading date t , a forward-discounted option baseline surface $C_f^0(m, \tau)$ and returns a corrected nodal surface u_t^* on a structured (m, τ) -grid. The post-fit acts only on regions where the baseline fit is locally problematic (poor band coverage and/or fragile static no-arbitrage).

12.1 Baseline grid surface and quotes

Fix a trading date t and let

$$\mathcal{G} := \{(m_i, \tau_j) : i = 1, \dots, n_m, j = 1, \dots, n_\tau\}$$

be a structured working grid in log-moneyness m and maturity τ . Denote its cardinality by

$$G := |\mathcal{G}| = n_m n_\tau,$$

fixing any one-to-one enumeration of $\mathcal{G}$ by indices $g \in \{1, \dots, G\} \leftrightarrow (i(g), j(g)) \in \mathcal{G}$.

Let $C_f^0(m, \tau)$ be the baseline forward-discounted call surface obtained from the main QP fit on date t . The corresponding nodal values on $\mathcal{G}$ are given by

$$u_{i,j}^0 := C_f^0(m_i, \tau_j), \quad (i, j) \in \mathcal{G}.$$

Collecting them into a vector, we obtain $u^0 \in \mathbb{R}^G$.

For the same date t , consider a set of cleaned forward-discounted quote bands

$$\{(m_q, \tau_q, b_q, a_q)\}_{q=1}^Q, \quad 0 \leq b_q \leq a_q,$$

where (m_q, τ_q) denotes the quote location in (m, τ) and $[b_q, a_q]$ is the corresponding bid–ask interval in forward-discounted units.

Let $S \in \mathbb{R}^{Q \times G}$ be the (fixed) bilinear interpolation operator that maps nodal values on $\mathcal{G}$ to model prices at the quote locations. For any nodal field $u \in \mathbb{R}^G$ we write

$$C_q(u) := (Su)_q, \quad q = 1, \dots, Q,$$

so $C_q(u)$ is the model forward-discounted call price at (m_q, τ_q) implied by the nodal surface u .

12.2 Badness map and patch decomposition

We now detect where the baseline surface is locally problematic.

Definition 5 (Baseline band misfit). For each quote $q \in \{1, \dots, Q\}$, the baseline band violation is defined as

$$d_q(u^0) := \text{dist}((Su^0)_q, [b_q, a_q]) = \max\{b_q - (Su^0)_q, 0, (Su^0)_q - a_q\} \geq 0.$$

Collecting all quote-wise misfits into a vector

$$d(u^0) := (d_q(u^0))_{q=1}^Q \in \mathbb{R}_{\geq 0}^Q,$$

regard $d(u^0)$ as the baseline distance-to-band profile at the true quote locations.

We transport these quote-level misfits to the working grid $\mathcal{G}$ via a fixed linear operator

$$R^{\text{band}} \in \mathbb{R}^{G \times Q},$$

(for example by locally averaging nearby quotes around each grid node). The purpose is to observe these violations at the grid nodes rather than at the quote locations. We write

$$w^{\text{band}} := R^{\text{band}} d(u^0) \in \mathbb{R}^G, \quad w_{i,j}^{\text{band}} \geq 0, \quad (i, j) \in \mathcal{G}.$$

In practice the entries of R^{band} are chosen nonnegative and supported only on quotes (m_q, τ_q) lying in a small neighbourhood of (m_i, τ_j) , so that $w_{i,j}^{\text{band}}$ is a local aggregated band-misfit around that node.

Next, build a single scalar at each grid node to see how badly the baseline surface u^0 violates static no-arbitrage at that grid node. Let $w^{\text{noarb}} \in \mathbb{R}_{\geq 0}^G$ be any nonnegative *static no-arbitrage defect field* obtained by aggregating local violations of discrete bounds, strike monotonicity, strike convexity, and calendar monotonicity on the grid when evaluated at u^0 . Concretely, one may define at each node (i, j) :

Definition 6 (Bound violation $v_{i,j}^{\text{bnd}}$). Bounds are $0 \leq u_{i,j}^0 \leq F_{i,j}$. At each node, define the distance to the interval $[0, F_{i,j}]$

$$v_{i,j}^{\text{bnd}} = \max\{-u_{i,j}^0, 0, u_{i,j}^0 - F_{i,j}\}.$$

Therefore,

- If $u_{i,j}^0 \in [0, F_{i,j}]$, then $v_{i,j}^{\text{bnd}} = 0$.
- If $u_{i,j}^0 < 0$, then $v_{i,j}^{\text{bnd}} = -u_{i,j}^0$.

- If $u_{i,j}^0 > F_{i,j}$, then $v_{i,j}^{\text{bnd}} = u_{i,j}^0 - F_{i,j}$.

Definition 7 (Strike monotonicity violation $v_{i,j}^{\text{mono}}$). Static no-arbitrage requires call prices to be non-increasing in strike at fixed maturity. On the grid, this means that along each maturity slice j we should have

$$u_{i+1,j}^0 - u_{i,j}^0 \leq 0, \quad i = 1, \dots, n_m - 1.$$

Define the forward slope in m between nodes i and $i + 1$ by

$$\Delta_{i+\frac{1}{2},j}^{\text{mono}} := u_{i+1,j}^0 - u_{i,j}^0, \quad i = 1, \dots, n_m - 1.$$

If $\Delta_{i+\frac{1}{2},j}^{\text{mono}} \leq 0$, there is no monotonicity issue on that edge; if $\Delta_{i+\frac{1}{2},j}^{\text{mono}} > 0$, the price goes up in strike there (violation). Define the edge violation

$$d_{i+\frac{1}{2},j}^{\text{mono}} := \max \{ \Delta_{i+\frac{1}{2},j}^{\text{mono}}, 0 \}.$$

We then attach a node-based violation by aggregating the incident edges. For interior nodes $2 \leq i \leq n_m - 1$,

$$v_{i,j}^{\text{mono}} := \max \{ d_{i-\frac{1}{2},j}^{\text{mono}}, d_{i+\frac{1}{2},j}^{\text{mono}} \},$$

and at the boundaries we use the single available edge,

$$v_{1,j}^{\text{mono}} := d_{1+\frac{1}{2},j}^{\text{mono}}, \quad v_{n_m,j}^{\text{mono}} := d_{n_m-\frac{1}{2},j}^{\text{mono}}.$$

Thus,

- If all nearby slopes around (i, j) are non-positive, then $v_{i,j}^{\text{mono}} = 0$.
- If some local slope is upward, $v_{i,j}^{\text{mono}}$ records the largest upward jump touching that node.

Definition 8 (Strike convexity violation $v_{i,j}^{\text{conv}}$). Static no-arbitrage also requires convexity in strike. On a uniform m -grid, this is encoded by non-negative second differences

$$\Delta_{i,j}^2 := u_{i+1,j}^0 - 2u_{i,j}^0 + u_{i-1,j}^0 \geq 0, \quad i = 2, \dots, n_m - 1.$$

Define the convexity defect at the central node by

$$v_{i,j}^{\text{conv}} := \begin{cases} \max \{ -\Delta_{i,j}^2, 0 \}, & 2 \leq i \leq n_m - 1, \\ 0, & i = 1 \text{ or } i = n_m \text{ (no centred stencil)}. \end{cases}$$

Therefore,

- If $\Delta_{i,j}^2 \geq 0$, there is no convexity issue and $v_{i,j}^{\text{conv}} = 0$.
- If $\Delta_{i,j}^2 < 0$, the profile is locally concave in strike (violation) and $v_{i,j}^{\text{conv}} = -\Delta_{i,j}^2$ measures how badly convexity is violated.

On a non-uniform m -grid one can replace $\Delta_{i,j}^2$ by the standard three-point second-derivative formula with unequal spacings; for the purposes of constructing a badness indicator w^{noarb} the simple second difference is typically sufficient.

Definition 9 (Calendar violation $v_{i,j}^{\text{cal}}$ via $A_{\tau|K}$). Let a^* be the baseline coefficient vector and let $A_{\tau|K}$ be the calendar operator at fixed strike. Evaluate

$$h := A_{\tau|K} a^* \in \mathbb{R}^G,$$

and reshape h on the grid as $h_{i,j}$. Define the node-wise calendar defect by

$$v_{i,j}^{\text{cal}} := \max\{-h_{i,j}, 0\}.$$

Then $v_{i,j}^{\text{cal}} = 0$ whenever $(\partial_\tau C_f)(K, \tau)|_K \geq 0$ at (m_i, τ_j) , and $v_{i,j}^{\text{cal}}$ measures the local size of negative calendar slopes at fixed strike.

With the bound violation $v_{i,j}^{\text{bnd}}$, strike monotonicity violation $v_{i,j}^{\text{mono}}$, strike convexity violation $v_{i,j}^{\text{conv}}$ and calendar violation $v_{i,j}^{\text{cal}}$ define the static no-arbitrage defect field w^{noarb} .

Definition 10 (Static no-arbitrage defect field w^{noarb}). Given the node-wise violations

$$v_{i,j}^{\text{bnd}}, \quad v_{i,j}^{\text{mono}}, \quad v_{i,j}^{\text{conv}}, \quad v_{i,j}^{\text{cal}} \geq 0,$$

defined respectively for bounds, strike monotonicity, strike convexity, and calendar constraints at (m_i, τ_j) , the *static no-arbitrage defect field* is

$$w_{i,j}^{\text{noarb}} := \max\{v_{i,j}^{\text{bnd}}, v_{i,j}^{\text{mono}}, v_{i,j}^{\text{conv}}, v_{i,j}^{\text{cal}}\}, \quad (i, j) \in \mathcal{G}.$$

Thus $w_{i,j}^{\text{noarb}} \geq 0$ for all (i, j) , and $w_{i,j}^{\text{noarb}} = 0$ whenever all discrete no-arbitrage inequalities (bounds, strike monotonicity, strike convexity, and calendar monotonicity) hold without violation in a neighbourhood of (m_i, τ_j) .

Any alternative construction of a field $w^{\text{noarb}} \in \mathbb{R}_{\geq 0}^G$ with the same qualitative properties

- $w_{i,j}^{\text{noarb}} \geq 0$ for all (i, j) , and
- $w_{i,j}^{\text{noarb}} = 0$ whenever all discrete no-arbitrage inequalities are satisfied (with margin) near (m_i, τ_j) ,

is equally admissible for the purposes of the badness map construction below.

The band-misfit field w^{band} and the static no-arbitrage defect field w^{noarb} provide two complementary scalar diagnostics on the grid $\mathcal{G}$: the former reflects how hard it is for the baseline surface to respect the bid-ask bands, while the latter reflects how fragile the static shape constraints are in a neighborhood of each node. For the purposes of patch detection we now compress these two pieces of information into a single scalar badness field on $\mathcal{G}$, allowing for a tunable trade-off between band fit and no-arbitrage robustness.

Definition 11 (Raw and smoothed badness field). Fix positive scalars $\alpha_{\text{band}}, \alpha_{\text{noarb}} > 0$. The *raw badness field* on $\mathcal{G}$ is

$$\tilde{w}_{i,j} := \alpha_{\text{band}} w_{i,j}^{\text{band}} + \alpha_{\text{noarb}} w_{i,j}^{\text{noarb}}, \quad (i, j) \in \mathcal{G}.$$

Let K_σ be a fixed separable Gaussian kernel on the grid, $K_\sigma(i, j) = k_\sigma^{(m)}(i) k_\sigma^{(\tau)}(j)$, and let $*$ denote discrete convolution on $\mathcal{G}$:

$$(K_\sigma * \tilde{w})_{i,j} := \sum_{(i', j') \in \mathcal{G}} K_\sigma(i - i', j - j') \tilde{w}_{i', j'}.$$

Define the (componentwise) clipping operator

$$\text{Clip}_{[0,1]}(x)_{i,j} := \min\{1, \max\{0, x_{i,j}\}\}.$$

The *smoothed badness field* is then

$$w := \text{Clip}_{[0,1]}(K_\sigma * \tilde{w}) \in [0, 1]^G.$$

The fixed separable Gaussian kernel on the grid K_σ is a bell-shaped weight function centered at 0 and decaying as you move away. Separable means that it can factor into a product of a 1D kernel in m and a 1D kernel in τ . The convolution $(K_\sigma * \tilde{w})_{i,j}$ is a weighted average of the raw badness $\tilde{w}_{i,j}$ in a neighborhood of (i, j) with weights given by the Gaussian kernel evaluated at offsets $(i - i', j - j')$.

Example 12 (Single Spike). Consider a 1D grid with indices $i = 1, \dots, 7$ and a raw badness vector

$$\tilde{w} = (0, 0, 0, 1, 0, 0, 0),$$

so there is a single spike of badness at $i = 4$. Take a simple discrete kernel

$$K = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right),$$

interpreted as $K(-1) = \frac{1}{4}$, $K(0) = \frac{1}{2}$, $K(1) = \frac{1}{4}$, and $K(k) = 0$ for $|k| > 1$. The convolution $(K * \tilde{w})_i$ with zero-padding at the boundaries is

$$(K * \tilde{w})_i = \frac{1}{4} \tilde{w}_{i-1} + \frac{1}{2} \tilde{w}_i + \frac{1}{4} \tilde{w}_{i+1}, \quad i = 1, \dots, 7.$$

A direct computation gives

$$(K * \tilde{w}) = (0, 0, 0.25, 0.5, 0.25, 0, 0).$$

Thus the original spike of height 1 at $i = 4$ is smoothed into a smaller peak of height 0.5 at $i = 4$ with nonzero neighbours of height 0.25 at $i = 3$ and $i = 5$: the mass has been spread out and diluted.

Example 13 (Cluster). Now consider a cluster of three bad nodes

$$\tilde{w} = (0, 0, 1, 1, 1, 0, 0),$$

again on indices $i = 1, \dots, 7$, with the same kernel

$$K = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right).$$

Using the same convolution formula

$$(K * \tilde{w})_i = \frac{1}{4} \tilde{w}_{i-1} + \frac{1}{2} \tilde{w}_i + \frac{1}{4} \tilde{w}_{i+1}, \quad i = 1, \dots, 7,$$

we obtain

$$\begin{aligned} (K * \tilde{w})_2 &= \frac{1}{4} \cdot 0 + \frac{1}{2} \cdot 0 + \frac{1}{4} \cdot 1 = 0.25, \\ (K * \tilde{w})_3 &= \frac{1}{4} \cdot 0 + \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 1 = 0.75, \\ (K * \tilde{w})_4 &= \frac{1}{4} \cdot 1 + \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 1 = 1.00, \\ (K * \tilde{w})_5 &= \frac{1}{4} \cdot 1 + \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 0 = 0.75, \\ (K * \tilde{w})_6 &= \frac{1}{4} \cdot 1 + \frac{1}{2} \cdot 0 + \frac{1}{4} \cdot 0 = 0.25, \end{aligned}$$

and $(K * \tilde{w})_1 = (K * \tilde{w})_7 = 0$. Hence

$$(K * \tilde{w}) = (0, 0.25, 0.75, 1.0, 0.75, 0.25, 0).$$

In this case a contiguous cluster of bad nodes remains a single coherent bump: the central node retains height 1 and its neighbours are only slightly reduced to 0.75, while a halo of smaller values 0.25 appears around the cluster. This illustrates how Gaussian smoothing preserves genuine regions of badness while softening their edges and suppressing isolated spikes.

Therefore, a single noisy spike in $\tilde{w}$ will spread out to neighbors, clusters of large $\tilde{w}$ will be smoothed into a broader hot region rather than an isolated piece and a smooth, spatially coherent badness image is created.

The clipping operator is applied to obtain a dimensionless, bounded heatmap which is comparable across dates and assets. One can therefore choose a threshold $\theta \in (0, 1)$ and obtain patches from the connected components of $\{\omega_{i,j} > \theta\}$. Moreover, the smoothing step ensures that patches correspond to regions rather than isolated single nodes.

Graphically, w can be viewed as a heatmap on the (m, τ) -plane: nodes with $w_{i,j} \approx 0$ are locally well-behaved (good band coverage and robust static no-arbitrage), while nodes with $w_{i,j}$ close to 1 lie in fragile or hard-to-fit regions.

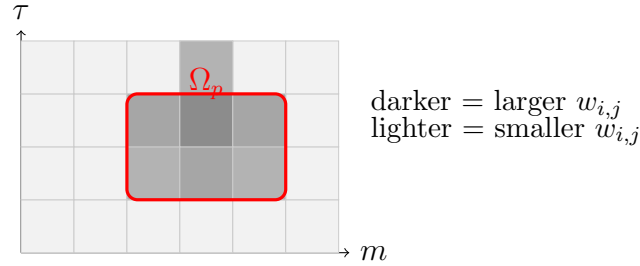


Figure 12.1: Schematic badness map $w_{i,j}$ on the (m, τ) grid $\mathcal{G}$. Darker cells indicate regions where the smoothed badness field is large; a connected high-badness region is shown as a patch Ω_p .

We now threshold w and decompose the high-badness region into connected components.

Definition 14 (Active set and patches). Fix a threshold $\theta \in (0, 1)$. The *active set* is

$$A := \{(i, j) \in \mathcal{G} : w_{i,j} > \theta\}.$$

Equip $\mathcal{G}$ with a nearest-neighbour graph structure, either *4-neighbour* (edges between (i, j) and $(i \pm 1, j)$, $(i, j \pm 1)$) or *8-neighbour* (4-neighbour plus diagonals $(i \pm 1, j \pm 1)$). Two nodes of A are said to be connected if they are joined by a path of neighbours in this graph. The connected components of A are then

$$A = \bigsqcup_{p \in \mathcal{P}} \Omega_p,$$

where each $\Omega_p \subset \mathcal{G}$ is a maximal connected subset of A (with respect to the chosen neighbourhood) and is called a *patch*.

Nodes in $\bigcup_p \Omega_p$ belong to locally difficult regions and are candidates for post-fit adjustment; nodes in $\mathcal{G} \setminus \bigcup_p \Omega_p$ are left unchanged by the post-fit. The construction is entirely local and depends only on the baseline misfit and static defect fields at the given date, not on the calendar regime; both calm and stressed dates are treated identically.

From now on we fix a single patch $\Omega \subset \mathcal{G}$ and describe the patch-level post-fit problem.

12.3 Discrete 3D fog on (m, τ, u)

On a fixed patch $\Omega \subset \mathcal{G}$ we will represent local pricing uncertainty and potential quote noise by a discretised probability density over the three-dimensional space (m, τ, u) , where u denotes forward-discounted call price.

12.3.1 3D lattice and fog variables

Definition 15 (3D fog lattice). Let $\Omega \subset \mathcal{G}$ be a patch with cardinality $N_\Omega := |\Omega|$. Fix a finite, strictly increasing sequence of price levels

$$U := \{u_k\}_{k=1}^{n_u} \subset \mathbb{R}, \quad u_1 < u_2 < \dots < u_{n_u},$$

spanning a relevant price range (for example from 0 up to a suitable multiple of the local forward). The associated three-dimensional lattice of (m, τ, u) -nodes on Ω is

$$\mathcal{L}_\Omega := \{(i, j, k) : (i, j) \in \Omega, k = 1, \dots, n_u\}.$$

Definition 16 (Fog variables and normalisation). For each $(i, j, k) \in \mathcal{L}_\Omega$ introduce a nonnegative *fog variable* $\pi_{i,j,k} \geq 0$. Collecting all such variables into a single vector

$$\pi := (\pi_{i,j,k})_{(i,j,k) \in \mathcal{L}_\Omega} \in \mathbb{R}^{N_\Omega n_u},$$

we impose the global normalisation

$$\sum_{(i,j) \in \Omega} \sum_{k=1}^{n_u} \pi_{i,j,k} = 1.$$

The associated feasible set of fog configurations on Ω is the simplex

$$\mathcal{C}_\pi(\Omega) := \left\{ \pi \in \mathbb{R}_{\geq 0}^{N_\Omega n_u} : \sum_{(i,j) \in \Omega} \sum_{k=1}^{n_u} \pi_{i,j,k} = 1 \right\}.$$

Thus π is a discrete probability measure on the finite set $\mathcal{L}_\Omega$: each $\pi_{i,j,k}$ is the probability mass (or fog mass) attached to the node (m_i, τ_j, u_k) , and $\mathcal{C}_\pi(\Omega)$ is a standard $(N_\Omega n_u - 1)$ -dimensional simplex.

At fixed $(i, j) \in \Omega$, the vertical profile

$$\{\pi_{i,j,k}\}_{k=1}^{n_u}$$

describes the distribution of fog across price levels $\{u_k\}$ at that grid location. In particular, we may interpret:

- $\pi_{i,j,k}$ large for some k as assigning high plausibility to local prices near u_k at (m_i, τ_j) ;
- a spread-out vertical profile as expressing substantial local uncertainty over u ;
- a concentrated profile as expressing locally precise information.

12.3.2 2D noise marginal on the patch

Definition 17 (2D marginal noise density). Given $\pi \in \mathcal{C}_\pi(\Omega)$, define the *2D marginal* of the fog on Ω by

$$n_{i,j} := \sum_{k=1}^{n_u} \pi_{i,j,k}, \quad (i,j) \in \Omega.$$

The marginal $n := (n_{i,j})_{(i,j) \in \Omega}$ satisfies

$$n_{i,j} \geq 0, \quad \sum_{(i,j) \in \Omega} n_{i,j} = 1,$$

and can be interpreted as a probability distribution on Ω :

$$n_{i,j} = \mathbb{P}\{(m, \tau) \text{ lies at node } (m_i, \tau_j)\}.$$

Equivalently, $n_{i,j}$ is the *total fog mass* sitting above (m_i, τ_j) when one integrates out the price dimension u .

Where $n_{i,j}$ is relatively large, the fog is thick in the (m, τ) -plane and the local order book in that region is regarded as noisy or unreliable; where $n_{i,j}$ is small, very little fog mass resides and the local book is regarded as comparatively clean.

When $n_{i,j} > 0$, it is often convenient to speak of the *conditional vertical distribution* at (i, j) ,

$$p_{i,j,k} := \frac{\pi_{i,j,k}}{n_{i,j}}, \quad k = 1, \dots, n_u,$$

which is a discrete probability distribution on $\{u_k\}$ satisfying $\sum_{k=1}^{n_u} p_{i,j,k} = 1$. In terms of this factorisation,

$$\pi_{i,j,k} = n_{i,j} p_{i,j,k},$$

so that n encodes where fog mass is located on the patch, and p encodes how that fog is distributed vertically in price at each node.

12.3.3 Schematic geometry

The geometry of the fog on a patch can be visualised as a stack of vertical columns above the (m, τ) nodes in Ω , each column sampled at levels $\{u_k\}$:

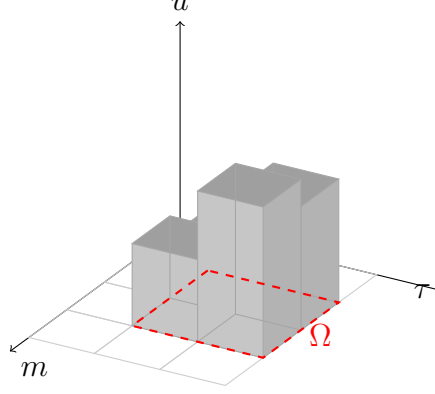


Figure 12.2: Schematic 3D fog on a patch Ω . The (m, τ) -plane is spanned by the horizontal m -axis and the in-plane τ -axis; the u -axis is vertical. Each node $(m_i, \tau_j) \in \Omega$ carries a vertical column of fog mass discretised at price levels $\{u_k\}$. The variables $\pi_{i,j,k}$ encode the mass at (m_i, τ_j, u_k) ; their 2D marginal $n_{i,j}$ is the total mass in the column above (m_i, τ_j) .

In the optimisation below, the fog configuration $\pi \in \mathcal{C}_\pi(\Omega)$ will be coupled to the nodal price field on Ω via band-based potentials and a Hamiltonian energy, with $n_{i,j}$ controlling how strongly local bid-ask information is enforced at each node.

12.4 Patch-level price field and static no-arbitrage

Fix a patch $\Omega \subset \mathcal{G}$ with cardinality $N_\Omega := |\Omega|$ and baseline nodal surface $u^0 \in \mathbb{R}^G$ as in Section 12.1. On Ω we will allow nodal prices to move, while all off-patch values are kept fixed at their baseline levels. Static no-arbitrage is imposed on the *assembled* full-grid surface.

12.4.1 Interior price variables and assembly map

Definition 18 (Interior price variables and assembled surface). Choose any one-to-one enumeration of the patch

$$\Omega = \{(i_\ell, j_\ell)\}_{\ell=1}^{N_\Omega} \subset \mathcal{G}.$$

The *interior price vector* (unknown) on Ω is

$$u_I := (u_{i_\ell, j_\ell})_{\ell=1}^{N_\Omega} \in \mathbb{R}^{N_\Omega},$$

whose entries correspond to the nodal prices on Ω . Define the *assembly map*

$$\mathcal{A}_\Omega : \mathbb{R}^{N_\Omega} \rightarrow \mathbb{R}^G, \quad u_I \mapsto u(u_I),$$

by

$$u(u_I)_{i,j} := \begin{cases} u_{i,j}, & (i,j) \in \Omega, \\ u_{i,j}^0, & (i,j) \notin \Omega. \end{cases}$$

Equivalently, if we write $u \in \mathbb{R}^G$ in the same enumeration as $\mathcal{G}$ and let $P_\Omega \in \mathbb{R}^{G \times N_\Omega}$ be the binary matrix that injects u_I into the coordinates corresponding to Ω (and zeros elsewhere), then the assembly map can be written as

$$u(u_I) = P_\Omega u_I + u^{0,\text{off}}, \tag{12.1}$$

where $u^{0,\text{off}} \in \mathbb{R}^G$ coincides with u^0 on $\mathcal{G} \setminus \Omega$ and is zero on Ω . Thus $\mathcal{A}_\Omega$ is an affine map, with linear part P_Ω .

12.4.2 Global discrete static no-arbitrage on the grid

We recall that in Chapters 4–7 the discrete static no-arbitrage conditions on the nodal grid $\mathcal{G}$ were encoded as a finite system of linear inequalities in the full nodal vector $u \in \mathbb{R}^G$. Concretely, there exists an index set $\mathcal{I} = \mathcal{I}_{\text{bnd}} \cup \mathcal{I}_{\text{mono}} \cup \mathcal{I}_{\text{conv}} \cup \mathcal{I}_{\text{cal}}$ and, for each $\alpha \in \mathcal{I}$, a row vector $\ell_\alpha^\top \in \mathbb{R}^{1 \times G}$ and a scalar $r_\alpha \in \mathbb{R}$ such that:

- for $\alpha \in \mathcal{I}_{\text{bnd}}$, the inequality $\ell_\alpha^\top u \leq r_\alpha$ encodes a bound constraint $0 \leq u_{i,j} \leq F_{i,j}$ at some node $(i, j) \in \mathcal{G}$;
- for $\alpha \in \mathcal{I}_{\text{mono}}$, the inequality encodes a discrete strike-monotonicity condition $\partial_K u \leq 0$ on a maturity slice (e.g. $u_{i+1,j} - u_{i,j} \leq 0$);
- for $\alpha \in \mathcal{I}_{\text{conv}}$, the inequality encodes a discrete strike-convexity condition $\partial_{KK} u \geq 0$ on a slice (e.g. a local second-difference inequality);
- for $\alpha \in \mathcal{I}_{\text{cal}}$, the inequality encodes a discrete calendar condition $(\partial_\tau u)|_K \geq 0$ at fixed strike, implemented via the fixed-strike calendar operator $A_{\tau|K}$ from Section 4.3.

Collecting these, the global static no-arbitrage feasible set on the full grid is

$$\mathcal{C}_{\text{glob}} := \left\{ u \in \mathbb{R}^G : \ell_\alpha^\top u \leq r_\alpha, \quad \forall \alpha \in \mathcal{I} \right\}. \quad (12.2)$$

This is precisely the intersection of finitely many closed half-spaces in $\mathbb{R}^G$.

Definition 19 (Global no-arbitrage operators). Let $\mathcal{G} = \{(i(g), j(g)) : g = 1, \dots, G\}$ be an enumeration of the nodal grid and let $e^{(g)} \in \mathbb{R}^G$ denote the g -th standard basis vector. Write $g(i, j)$ for the index such that $(i(g(i, j)), j(g(i, j))) = (i, j)$.

We define index sets and pairs (ℓ_α, r_α) as follows, where $e^{(g)}$ is the g -th standard basis vector in $\mathbb{R}^G$ (vector with 1 in position g and 0 elsewhere):

- (*Bounds*) For each $(i, j) \in \mathcal{G}$ with $g = g(i, j)$ define lower and upper indices $\alpha = (i, j, \text{lo})$, $\alpha' = (i, j, \text{up})$ and

$$\ell_{(i,j,\text{lo})}^\top := -(e^{(g)})^\top, \quad r_{(i,j,\text{lo})} := 0, \quad \ell_{(i,j,\text{up})}^\top := (e^{(g)})^\top, \quad r_{(i,j,\text{up})} := F_{i,j}.$$

Collect all such indices into $\mathcal{I}_{\text{bnd}}$.

- (*Monotonicity*) For each maturity j and $i = 1, \dots, n_m - 1$, with $g_1 = g(i, j)$, $g_2 = g(i + 1, j)$, define $\alpha = (i, j) \in \mathcal{I}_{\text{mono}}$ and

$$\ell_{(i,j)}^\top := (e^{(g_2)} - e^{(g_1)})^\top, \quad r_{(i,j)} := 0.$$

- (*Convexity*) For each maturity j and $i = 2, \dots, n_m - 1$, with $g_- = g(i - 1, j)$, $g_0 = g(i, j)$, $g_+ = g(i + 1, j)$, define $\alpha = (i, j) \in \mathcal{I}_{\text{conv}}$ and

$$\ell_{(i,j)}^\top := (-e^{(g_+)} + 2e^{(g_0)} - e^{(g_-)})^\top, \quad r_{(i,j)} := 0.$$

- (*Calendar*) Let $A_{\tau|K} \in \mathbb{R}^{G \times G}$ be the fixed-strike calendar operator. For each $g \in \{1, \dots, G\}$ define $\alpha = g \in \mathcal{I}_{\text{cal}}$ and

$$\ell_g^\top := -(A_{\tau|K})_{g,\cdot}, \quad r_g := 0.$$

Set $\mathcal{I} := \mathcal{I}_{\text{bnd}} \cup \mathcal{I}_{\text{mono}} \cup \mathcal{I}_{\text{conv}} \cup \mathcal{I}_{\text{cal}}$.

This definition can be seen as follows:

Bounds. We first explain how to encode the pointwise bounds

$$0 \leq u_{i,j} \leq F_{i,j}, \quad (i,j) \in \mathcal{G},$$

as linear inequalities of the form $\ell_\alpha^\top u \leq r_\alpha$.

Fix a node $(i,j) \in \mathcal{G}$ and let $g = g(i,j)$ be its index in the flattened nodal vector $u \in \mathbb{R}^G$, so that $u_g = u_{i,j}$. For this node we introduce *two* constraint indices:

$$\alpha = (i,j, \text{lo}), \quad \alpha' = (i,j, \text{up}).$$

Lower bound. For $\alpha = (i,j, \text{lo})$ we define

$$\ell_{(i,j,\text{lo})}^\top := -(e^{(g)})^\top, \quad r_{(i,j,\text{lo})} := 0,$$

where $e^{(g)} \in \mathbb{R}^G$ is the g -th standard basis vector. Then

$$\ell_{(i,j,\text{lo})}^\top u = -e^{(g)\top} u = -u_g = -u_{i,j},$$

so the inequality $\ell_{(i,j,\text{lo})}^\top u \leq r_{(i,j,\text{lo})}$ reads

$$-u_{i,j} \leq 0 \quad \Longleftrightarrow \quad u_{i,j} \geq 0.$$

Upper bound. For $\alpha' = (i,j, \text{up})$ we set

$$\ell_{(i,j,\text{up})}^\top := (e^{(g)})^\top, \quad r_{(i,j,\text{up})} := F_{i,j}.$$

Then

$$\ell_{(i,j,\text{up})}^\top u = e^{(g)\top} u = u_g = u_{i,j},$$

so $\ell_{(i,j,\text{up})}^\top u \leq r_{(i,j,\text{up})}$ is exactly

$$u_{i,j} \leq F_{i,j}.$$

Thus for each node (i,j) we obtain the two bounds $0 \leq u_{i,j} \leq F_{i,j}$ as the pair of inequalities

$$\ell_{(i,j,\text{lo})}^\top u \leq r_{(i,j,\text{lo})}, \quad \ell_{(i,j,\text{up})}^\top u \leq r_{(i,j,\text{up})}.$$

All such indices (i,j, lo) and (i,j, up) are collected in $\mathcal{I}_{\text{bnd}}$.

Monotonicity. We now encode discrete strike monotonicity, namely

$$u_{i+1,j} - u_{i,j} \leq 0, \quad i = 1, \dots, n_m - 1, \quad j = 1, \dots, n_\tau.$$

Fix a maturity j and $i \in \{1, \dots, n_m - 1\}$. Let

$$g_1 := g(i, j), \quad g_2 := g(i + 1, j)$$

be the indices of the adjacent nodes (i, j) and $(i + 1, j)$ in the flattened vector u . We introduce a single index $\alpha = (i, j) \in \mathcal{I}_{\text{mono}}$ and define

$$\ell_{(i,j)}^\top := (e^{(g_2)} - e^{(g_1)})^\top, \quad r_{(i,j)} := 0.$$

Then

$$\ell_{(i,j)}^\top u = e^{(g_2)\top} u - e^{(g_1)\top} u = u_{g_2} - u_{g_1} = u_{i+1,j} - u_{i,j}.$$

Thus the inequality $\ell_{(i,j)}^\top u \leq r_{(i,j)}$ is precisely

$$u_{i+1,j} - u_{i,j} \leq 0,$$

the desired monotonicity condition. Each adjacent pair of strikes at fixed j contributes one such index (i, j) to $\mathcal{I}_{\text{mono}}$.

Convexity. Discrete strike convexity requires

$$u_{i+1,j} - 2u_{i,j} + u_{i-1,j} \geq 0, \quad i = 2, \dots, n_m - 1, \quad j = 1, \dots, n_\tau.$$

Equivalently,

$$-u_{i+1,j} + 2u_{i,j} - u_{i-1,j} \leq 0.$$

Fix a maturity j and an interior strike index $i \in \{2, \dots, n_m - 1\}$. Let

$$g_- := g(i - 1, j), \quad g_0 := g(i, j), \quad g_+ := g(i + 1, j).$$

We introduce an index $\alpha = (i, j) \in \mathcal{I}_{\text{conv}}$ and set

$$\ell_{(i,j)}^\top := (-e^{(g_+)} + 2e^{(g_0)} - e^{(g_-)})^\top, \quad r_{(i,j)} := 0.$$

Then

$$\ell_{(i,j)}^\top u = -u_{g_+} + 2u_{g_0} - u_{g_-} = -u_{i+1,j} + 2u_{i,j} - u_{i-1,j},$$

so the inequality $\ell_{(i,j)}^\top u \leq r_{(i,j)}$ is exactly

$$-u_{i+1,j} + 2u_{i,j} - u_{i-1,j} \leq 0 \quad \Longleftrightarrow \quad u_{i+1,j} - 2u_{i,j} + u_{i-1,j} \geq 0.$$

All such indices (i, j) form the convexity index set $\mathcal{I}_{\text{conv}}$.

Calendar (fixed strike). Finally, we encode the discrete calendar condition at fixed strike via the operator $A_{\tau|K} \in \mathbb{R}^{G \times G}$. By construction, $(A_{\tau|K}u)_g$ is the discrete approximation of $(\partial_\tau C_f)(K, \tau)|_K$ at the grid node indexed by g . The condition

$$(A_{\tau|K}u)_g \geq 0, \quad g = 1, \dots, G,$$

can be written as

$$-(A_{\tau|K}u)_g \leq 0.$$

For each grid index $g \in \{1, \dots, G\}$ we take $\alpha = g \in \mathcal{I}_{\text{cal}}$ and define

$$\ell_g^\top := -(A_{\tau|K})_{g,\cdot}, \quad r_g := 0,$$

i.e. $\ell_g^\top$ is the negative of the g -th row of $A_{\tau|K}$. Then

$$\ell_g^\top u = -(A_{\tau|K}u)_g,$$

so the inequality $\ell_g^\top u \leq r_g$ is precisely

$$-(A_{\tau|K}u)_g \leq 0 \iff (A_{\tau|K}u)_g \geq 0.$$

Thus each row g of $A_{\tau|K}$ generates one calendar inequality, and all these indices g belong to $\mathcal{I}_{\text{cal}}$.

Definition 20 (Primitive discrete static no-arbitrage on $\mathcal{G}$). A nodal surface

$$u = (u_{i,j})_{(i,j) \in \mathcal{G}} \in \mathbb{R}^G$$

is said to satisfy *primitive discrete static no-arbitrage* on $\mathcal{G}$ if:

- *(BND)* (Bounds) For all $(i, j) \in \mathcal{G}$,

$$0 \leq u_{i,j} \leq F_{i,j}.$$

- *(MONO)* (Strike monotonicity) For each maturity j and each $i = 1, \dots, n_m - 1$,

$$u_{i+1,j} - u_{i,j} \leq 0.$$

- *(CONV)* (Strike convexity) For each maturity j and $i = 2, \dots, n_m - 1$,

$$u_{i+1,j} - 2u_{i,j} + u_{i-1,j} \geq 0.$$

- *(CAL)* (Calendar at fixed strike) Let $A_{\tau|K} \in \mathbb{R}^{G \times G}$ be the fixed-strike calendar operator from Section 4.3. Then

$$(A_{\tau|K}u)_g \geq 0 \quad \text{for all grid indices } g = 1, \dots, G.$$

Proposition 12 (Equivalence of operator encoding and primitive no-arbitrage). *Let $(\ell_\alpha, r_\alpha)_{\alpha \in \mathcal{I}}$ be defined as in Definition 19, with index sets $\mathcal{I}_{\text{bnd}}$, $\mathcal{I}_{\text{mono}}$, $\mathcal{I}_{\text{conv}}$, $\mathcal{I}_{\text{cal}}$ corresponding to bounds, strike monotonicity, strike convexity, and calendar constraints respectively, and $\mathcal{I} = \mathcal{I}_{\text{bnd}} \cup \mathcal{I}_{\text{mono}} \cup \mathcal{I}_{\text{conv}} \cup \mathcal{I}_{\text{cal}}$.*

For $u \in \mathbb{R}^G$, the following are equivalent:

(i) u satisfies all linear inequalities

$$\ell_\alpha^\top u \leq r_\alpha, \quad \forall \alpha \in \mathcal{I}.$$

(ii) u satisfies primitive discrete static no-arbitrage on $\mathcal{G}$ in the sense of Definition 20.

In particular, the global static no-arbitrage feasible set

$$\mathcal{C}_{\text{glob}} = \{u \in \mathbb{R}^G : \ell_\alpha^\top u \leq r_\alpha \ \forall \alpha \in \mathcal{I}\}$$

coincides with the set of all nodal surfaces that satisfy the primitive discrete no-arbitrage conditions (BND), (MONO), (CONV) and (CAL).

Proof. We prove (i) $\Leftrightarrow$ (ii) by decomposing $\mathcal{I}$ into its four subsets.

(ii) $\Rightarrow$ (i). Suppose u satisfies the primitive conditions (BND), (MONO), (CONV), (CAL).

Bounds. Fix $(i, j) \in \mathcal{G}$ and let $g = g(i, j)$ be its index. By Definition 19, the two bound indices $\alpha = (i, j, \text{lo})$, $\alpha' = (i, j, \text{up})$ satisfy

$$\ell_{(i,j,\text{lo})}^\top = -(e^{(g)})^\top, \quad r_{(i,j,\text{lo})} = 0, \quad \ell_{(i,j,\text{up})}^\top = (e^{(g)})^\top, \quad r_{(i,j,\text{up})} = F_{i,j},$$

where $e^{(g)}$ is the g -th standard basis vector in $\mathbb{R}^G$. Then

$$\ell_{(i,j,\text{lo})}^\top u = -u_{i,j}, \quad \ell_{(i,j,\text{up})}^\top u = u_{i,j}.$$

The bound condition $0 \leq u_{i,j} \leq F_{i,j}$ implies

$$-u_{i,j} \leq 0 = r_{(i,j,\text{lo})}, \quad u_{i,j} \leq F_{i,j} = r_{(i,j,\text{up})},$$

hence $\ell_\alpha^\top u \leq r_\alpha$ for all $\alpha \in \mathcal{I}_{\text{bnd}}$.

Strike monotonicity. Fix a maturity j and $i = 1, \dots, n_m - 1$, and let $g_1 = g(i, j)$, $g_2 = g(i + 1, j)$. By Definition 19, for $\alpha = (i, j) \in \mathcal{I}_{\text{mono}}$ we have

$$\ell_{(i,j)}^\top := (e^{(g_2)} - e^{(g_1)})^\top, \quad r_{(i,j)} := 0,$$

so that

$$\ell_{(i,j)}^\top u = u_{i+1,j} - u_{i,j}.$$

The primitive monotonicity condition (MONO) states that $u_{i+1,j} - u_{i,j} \leq 0$, hence $\ell_{(i,j)}^\top u \leq r_{(i,j)}$ for all (i, j) , i.e. for all $\alpha \in \mathcal{I}_{\text{mono}}$.

Strike convexity. Fix a maturity j and $i = 2, \dots, n_m - 1$, and let $g_- = g(i - 1, j)$, $g_0 = g(i, j)$, $g_+ = g(i + 1, j)$. For $\alpha = (i, j) \in \mathcal{I}_{\text{conv}}$, Definition 19 gives

$$\ell_{(i,j)}^\top := (-e^{(g_+)} + 2e^{(g_0)} - e^{(g_-)})^\top, \quad r_{(i,j)} := 0.$$

Then

$$\ell_{(i,j)}^\top u = -u_{i+1,j} + 2u_{i,j} - u_{i-1,j}.$$

The primitive convexity condition (CONV) is $u_{i+1,j} - 2u_{i,j} + u_{i-1,j} \geq 0$, or equivalently $-u_{i+1,j} + 2u_{i,j} - u_{i-1,j} \leq 0$, hence $\ell_{(i,j)}^\top u \leq r_{(i,j)}$ for all (i, j) , i.e. all $\alpha \in \mathcal{I}_{\text{conv}}$.

Calendar. Let $A_{\tau|K} \in \mathbb{R}^{G \times G}$ be the fixed-strike calendar operator. For each grid index $g \in \{1, \dots, G\}$, the calendar index $\alpha = g \in \mathcal{I}_{\text{cal}}$ is defined by

$$\ell_g^\top := -(A_{\tau|K})_{g,\cdot}, \quad r_g := 0,$$

so that

$$\ell_g^\top u = -(A_{\tau|K}u)_g.$$

The primitive calendar condition (CAL) states $(A_{\tau|K}u)_g \geq 0$ for all g , which is equivalent to $-(A_{\tau|K}u)_g \leq 0 = r_g$, hence $\ell_g^\top u \leq r_g$ for all $\alpha \in \mathcal{I}_{\text{cal}}$.

Combining the four families, we see that (ii) implies $\ell_\alpha^\top u \leq r_\alpha$ for all $\alpha \in \mathcal{I}$, i.e. (i) holds.

(i) $\Rightarrow$ (ii). Conversely, suppose u satisfies $\ell_\alpha^\top u \leq r_\alpha$ for all $\alpha \in \mathcal{I}$. We show that the primitive conditions (BND), (MONO), (CONV), (CAL) hold.

Bounds. Fix $(i, j) \in \mathcal{G}$ and $g = g(i, j)$. For $\alpha = (i, j, \text{lo})$ and $\alpha' = (i, j, \text{up})$ we have, by definition,

$$\ell_{(i,j,\text{lo})}^\top u = -u_{i,j} \leq r_{(i,j,\text{lo})} = 0, \quad \ell_{(i,j,\text{up})}^\top u = u_{i,j} \leq r_{(i,j,\text{up})} = F_{i,j}.$$

Thus $-u_{i,j} \leq 0$ and $u_{i,j} \leq F_{i,j}$, i.e. $0 \leq u_{i,j} \leq F_{i,j}$ for all (i, j) , which is (BND).

Strike monotonicity. For each j and $i = 1, \dots, n_m - 1$, the inequality for $\alpha = (i, j) \in \mathcal{I}_{\text{mono}}$ reads

$$\ell_{(i,j)}^\top u = u_{i+1,j} - u_{i,j} \leq 0,$$

which is exactly the discrete monotonicity condition $u_{i+1,j} - u_{i,j} \leq 0$ for all such (i, j) ; this is (MONO).

Strike convexity. For each j and $i = 2, \dots, n_m - 1$, the inequality for $\alpha = (i, j) \in \mathcal{I}_{\text{conv}}$ is

$$\ell_{(i,j)}^\top u = -u_{i+1,j} + 2u_{i,j} - u_{i-1,j} \leq 0.$$

Rearranging gives $u_{i+1,j} - 2u_{i,j} + u_{i-1,j} \geq 0$, which is (CONV).

Calendar. Finally, for each grid index $g \in \{1, \dots, G\}$ we have

$$\ell_g^\top u = -(A_{\tau|K}u)_g \leq 0,$$

or equivalently $(A_{\tau|K}u)_g \geq 0$. Thus (CAL) holds at every grid index.

Therefore all four primitive conditions (BND), (MONO), (CONV), (CAL) hold, and u satisfies primitive discrete static no-arbitrage. This proves (ii).

We have shown (ii) $\Rightarrow$ (i) and (i) $\Rightarrow$ (ii), so the two statements are equivalent. The characterisation of $\mathcal{C}_{\text{glob}}$ as the set of all primitively no-arbitrage nodal surfaces follows immediately from the definition of $\mathcal{C}_{\text{glob}}$. $\square$

12.4.3 Patch-level feasible set and its geometry

We now enforce global static no-arbitrage on the *assembled* surface $u(u_I)$ obtained from a patch interior vector u_I .

Definition 21 (No-arbitrage feasible set on a patch). The *patch-level no-arbitrage feasible set* is

$$\mathcal{C}_u(\Omega) := \{u_I \in \mathbb{R}^{N_\Omega} : u(u_I) \in \mathcal{C}_{\text{glob}}\}.$$

Equivalently, using (12.2) and the assembly map (12.1),

$$\mathcal{C}_u(\Omega) = \left\{ u_I \in \mathbb{R}^{N_\Omega} : \ell_\alpha^\top u(u_I) \leq r_\alpha, \quad \forall \alpha \in \mathcal{I} \right\}.$$

We now characterise $\mathcal{C}_u(\Omega)$ as a polyhedron and prove its basic geometric properties.

Proposition 13 (Polyhedral structure of $\mathcal{C}_u(\Omega)$). *The set $\mathcal{C}_u(\Omega) \subset \mathbb{R}^{N_\Omega}$ can be written as the finite intersection of affine half-spaces*

$$\mathcal{C}_u(\Omega) = \bigcap_{\alpha \in \mathcal{I}} H_\alpha,$$

where each H_α is of the form

$$H_\alpha := \{u_I \in \mathbb{R}^{N_\Omega} : a_\alpha^\top u_I \leq b_\alpha\}$$

for some $a_\alpha \in \mathbb{R}^{N_\Omega}$ and $b_\alpha \in \mathbb{R}$. In particular, $\mathcal{C}_u(\Omega)$ is a (possibly empty) closed convex polyhedron in $\mathbb{R}^{N_\Omega}$.

Proof. Fix $\alpha \in \mathcal{I}$. For $u_I \in \mathbb{R}^{N_\Omega}$, the corresponding assembled surface is $u(u_I) = P_\Omega u_I + u^{0,\text{off}}$ by (12.1). The α -th global no-arbitrage inequality reads

$$\ell_\alpha^\top u(u_I) \leq r_\alpha.$$

Substituting the affine expression for $u(u_I)$, we obtain

$$\ell_\alpha^\top (P_\Omega u_I + u^{0,\text{off}}) \leq r_\alpha,$$

which can be rearranged as

$$(\ell_\alpha^\top P_\Omega) u_I \leq r_\alpha - \ell_\alpha^\top u^{0,\text{off}}.$$

Now, define

$$a_\alpha := P_\Omega^\top \ell_\alpha \in \mathbb{R}^{N_\Omega}, \quad b_\alpha := r_\alpha - \ell_\alpha^\top u^{0,\text{off}} \in \mathbb{R}.$$

Note that $\ell_\alpha^\top P_\Omega = a_\alpha^\top$ by construction. Then the α -th constraint is equivalent to

$$a_\alpha^\top u_I \leq b_\alpha.$$

Therefore, the set of u_I satisfying the α -th global no-arbitrage inequality is the half-space

$$H_\alpha := \{u_I \in \mathbb{R}^{N_\Omega} : a_\alpha^\top u_I \leq b_\alpha\}.$$

Because this construction holds for every $\alpha \in \mathcal{I}$ (namely; to be patch-feasible u_I has to satisfy every constraint, so belong to every H_α), we have

$$\mathcal{C}_u(\Omega) = \{u_I \in \mathbb{R}^{N_\Omega} : a_\alpha^\top u_I \leq b_\alpha \ \forall \alpha \in \mathcal{I}\} = \bigcap_{\alpha \in \mathcal{I}} H_\alpha.$$

Recall that a half-space in $\mathbb{R}^n$ is any set that can be written as $\{x \in \mathbb{R}^n : a^\top x \leq b\}$ or $\{x \in \mathbb{R}^n : a^\top x \geq b\}$ for some fixed nonzero vector a and scalar b . Therefore; by definition, each H_α is a half-space in $\mathbb{R}^{N_\Omega}$.

A set C is convex if for all $x, y \in C$ and $\lambda \in [0, 1]$,

$$\lambda x + (1 - \lambda)y \in C.$$

Now, take $H_\alpha := \{u_I \in \mathbb{R}^{N_\Omega} : a_\alpha^\top u_I \leq b_\alpha\}$ and let $u_I^{(1)}, u_I^{(2)} \in H_\alpha$. We therefore have by definition of H_α that $a_\alpha^\top u_I^{(1)} \leq b_\alpha$ and $a_\alpha^\top u_I^{(2)} \leq b_\alpha$. Let $\lambda \in [0, 1]$ and consider $u_I^{(\lambda)} := \lambda u_I^{(1)} + (1 - \lambda)u_I^{(2)}$. Computing:

$$a_\alpha^\top u_I^{(\lambda)} = a_\alpha^\top (\lambda u_I^{(1)} + (1 - \lambda)u_I^{(2)}) = a_\alpha^\top \lambda u_I^{(1)} + a_\alpha^\top (1 - \lambda)u_I^{(2)}.$$

Since $a_\alpha^\top u_I^{(1)} \leq b_\alpha$ and $a_\alpha^\top u_I^{(2)} \leq b_\alpha$, we have

$$a_\alpha^\top \lambda u_I^{(1)} + a_\alpha^\top (1 - \lambda)u_I^{(2)} \leq \lambda b_\alpha + (1 - \lambda)b_\alpha = b_\alpha.$$

Therefore, $a_\alpha^\top u_I^{(\lambda)} \leq b_\alpha$. By definition, $u_I^{(\lambda)} \in H_\alpha$; which is the exact definition of convexity and thus H_α is convex.

We can see that each H_α is closed via sequences. Let $(u_I^{(n)})_{n \in \mathbb{N}}$ be a sequence in H_α converging to some $u_I^* \in \mathbb{R}^{N_\Omega}$. By definition of H_α we have

$$a_\alpha^\top u_I^{(n)} \leq b_\alpha, \quad \forall n \in \mathbb{N}.$$

The map $u_I \mapsto a_\alpha^\top u_I$ is linear and hence continuous on $\mathbb{R}^{N_\Omega}$, so passing to the limit $n \rightarrow \infty$ yields

$$a_\alpha^\top u_I^* = \lim_{n \rightarrow \infty} a_\alpha^\top u_I^{(n)} \leq b_\alpha.$$

Thus u_I^* also satisfies $a_\alpha^\top u_I^* \leq b_\alpha$, i.e. $u_I^* \in H_\alpha$. Therefore H_α contains the limit of every convergent sequence of its elements, and is closed.

Hence, each H_α is a closed half-space in $\mathbb{R}^{N_\Omega}$ and is convex. The intersection of any family of convex sets is convex, and the intersection of any family of closed sets is closed. Since $\mathcal{I}$ is finite, this intersection defines a closed convex polyhedron. The polyhedron may be empty or nonempty, depending on the data; we analyse feasibility separately. $\square$

Definition 22 (Patch feasibility). A patch $\Omega \subset \mathcal{G}$ is said to be *feasible* if $\mathcal{C}_u(\Omega) \neq \emptyset$.

The assumption that all patches used in the post-fit are feasible is mild and consistent with the construction of Ω from the baseline surface. Indeed, if the baseline nodal surface u^0 is globally statically no-arbitrage on $\mathcal{G}$, then $u^0 \in \mathcal{C}_{\text{glob}}$. If, in addition, the patch decomposition is chosen so that every static no-arbitrage stencil that intersects the interior of Ω is either fully contained in Ω or fully contained in $\mathcal{G} \setminus \Omega$, then the restriction of u^0 to Ω defines an interior vector $u_I^{0, \Omega}$ satisfying all constraints $a_\alpha^\top u_I \leq b_\alpha$ and hence $u_I^{0, \Omega} \in \mathcal{C}_u(\Omega)$. In practice, feasibility can be verified numerically by solving a simple linear feasibility problem for each patch; in the theoretical development of this chapter we take it as an explicit assumption that patches are chosen so that $\mathcal{C}_u(\Omega)$ is nonempty.

12.5 Hamiltonian energy on the fog

We now endow the discrete fog π on the lattice $\mathcal{L}_\Omega$ with a quadratic Hamiltonian energy. The Hamiltonian has two components: a *kinetic* (Dirichlet) term that penalises roughness of π across neighboring lattice sites in (m, τ, u) , and a *potential* term that penalises fog mass lying far outside local bid–ask tubes or basic price ranges.

Throughout this section we fix a patch $\Omega \subset \mathcal{G}$ with $N_\Omega = |\Omega|$ and a set of price levels $\{u_k\}_{k=1}^{n_u}$, and we work on the lattice

$$\mathcal{L}_\Omega := \{(i, j, k) : (i, j) \in \Omega, k = 1, \dots, n_u\},$$

as in Section 12.3.

12.5.1 Discrete 3D graph and difference operators

We begin by making the discrete graph structure of $\mathcal{L}_\Omega$ explicit and constructing the associated difference operators.

Definition 23 (Adjacency graph on $\mathcal{L}_\Omega$). Let $\mathcal{L}_\Omega$ be the 3D lattice of nodes (i, j, k) with $(i, j) \in \Omega$ and $k \in \{1, \dots, n_u\}$. We define three families of undirected edges:

- *m-edges* E_m : for every $(i, j, k) \in \mathcal{L}_\Omega$ such that $(i+1, j) \in \Omega$, we introduce an edge between (i, j, k) and $(i+1, j, k)$;
- *τ -edges* E_τ : for every $(i, j, k) \in \mathcal{L}_\Omega$ such that $(i, j+1) \in \Omega$, we introduce an edge between (i, j, k) and $(i, j+1, k)$;
- *u-edges* E_u : for every $(i, j, k) \in \mathcal{L}_\Omega$ such that $k+1 \leq n_u$, we introduce an edge between (i, j, k) and $(i, j, k+1)$.

The full edge set is

$$E := E_m \cup E_\tau \cup E_u.$$

Thus E_m connects nearest neighbours in the m -direction (at fixed (τ, u)), E_τ connects nearest neighbours in the τ -direction (at fixed (m, u)), and E_u connects nearest neighbours in the u -direction (at fixed (m, τ)). Since edges are only drawn between nodes that both belong to $\mathcal{L}_\Omega$, this corresponds to homogeneous Neumann boundary conditions at the boundary of the patch.

For each edge family we now define a discrete gradient operator as the signed incidence matrix of the corresponding graph.

Definition 24 (Discrete gradients along m, τ, u). Fix an arbitrary but fixed orientation of each edge in E_m, E_τ, E_u : for each edge $e = \{p, q\} \in E_m$, choose an ordering $(p \rightarrow q)$ (e.g. increasing in i); similarly for E_τ (increasing in j) and E_u (increasing in k).

Let $N_L := |\mathcal{L}_\Omega| = N_\Omega n_u$ be the number of lattice nodes, and enumerate $\mathcal{L}_\Omega$ as

$$\mathcal{L}_\Omega = \{\xi_\ell\}_{\ell=1}^{N_L}, \quad \xi_\ell = (i_\ell, j_\ell, k_\ell).$$

We identify fog configurations π with vectors in $\mathbb{R}^{N_L}$ via $\pi_\ell := \pi_{i_\ell, j_\ell, k_\ell}$.

- The m -gradient $D_m : \mathbb{R}^{N_L} \rightarrow \mathbb{R}^{|E_m|}$ is defined as follows: index $E_m = \{e_r\}_{r=1}^{|E_m|}$ and for each edge $e_r = (p \rightarrow q)$, set

$$(D_m \pi)_r := \pi_q - \pi_p.$$

In matrix form, D_m is the $|E_m| \times N_L$ matrix whose r -th row has entry -1 in the column corresponding to node p , entry $+1$ in the column corresponding to node q , and zeros elsewhere.

- The τ -gradient $D_\tau : \mathbb{R}^{N_L} \rightarrow \mathbb{R}^{|E_\tau|}$ is defined analogously, with one row per edge in E_τ , oriented in increasing j .
- The u -gradient $D_u : \mathbb{R}^{N_L} \rightarrow \mathbb{R}^{|E_u|}$ is defined analogously, with one row per edge in E_u , oriented in increasing k .

Thus $D_m \pi$ collects all forward differences of π along m -edges, and similarly for D_τ and D_u . We now weight these differences by nonnegative edge weights.

Definition 25 (Edge-weight matrices). Let $w_r^m \geq 0$ be a nonnegative weight associated with the r -th m -edge in E_m , and define the diagonal matrix

$$W_m := \text{diag}(w_1^m, \dots, w_{|E_m|}^m) \in \mathbb{R}^{|E_m| \times |E_m|}.$$

Similarly, let $w_r^\tau \geq 0$ and $w_r^u \geq 0$ be edge weights on E_τ and E_u , and define diagonal matrices $W_\tau \in \mathbb{R}^{|E_\tau| \times |E_\tau|}$ and $W_u \in \mathbb{R}^{|E_u| \times |E_u|}$ with these weights on the diagonal.

Typical choices include $w_r^m, w_r^\tau, w_r^u \equiv 1$ (unweighted differences), or weights that depend on grid spacings and/or the 2D marginal $n_{i,j}$; the only property needed here is nonnegativity.

For any such diagonal matrix $W \succeq 0$ and vector x , we write

$$\|x\|_W^2 := x^\top W x$$

for the weighted squared norm.

12.5.2 Kinetic energy and graph Laplacian

We now define the Dirichlet kinetic energy of the fog in terms of these discrete gradients.

Definition 26 (Kinetic energy of the fog). Let $\kappa_m, \kappa_\tau, \kappa_u \geq 0$ be fixed nonnegative parameters. The *kinetic energy* (Dirichlet energy) of a fog configuration $\pi \in \mathbb{R}^{N_L}$ is

$$\mathcal{E}_{\text{kin}}(\pi) := \frac{\kappa_m}{2} \|D_m \pi\|_{W_m}^2 + \frac{\kappa_\tau}{2} \|D_\tau \pi\|_{W_\tau}^2 + \frac{\kappa_u}{2} \|D_u \pi\|_{W_u}^2.$$

Explicitly,

$$\mathcal{E}_{\text{kin}}(\pi) = \frac{\kappa_m}{2} (D_m \pi)^\top W_m (D_m \pi) + \frac{\kappa_\tau}{2} (D_\tau \pi)^\top W_\tau (D_\tau \pi) + \frac{\kappa_u}{2} (D_u \pi)^\top W_u (D_u \pi).$$

The Dirichlet energy is a nonnegative quadratic form in π , and it can be written in the standard graph-Laplacian form.

Proposition 14 (Matrix form and positive semidefiniteness of L_π). *Define*

$$L_m := D_m^\top W_m D_m, \quad L_\tau := D_\tau^\top W_\tau D_\tau, \quad L_u := D_u^\top W_u D_u$$

and

$$L_\pi := \kappa_m L_m + \kappa_\tau L_\tau + \kappa_u L_u.$$

Then:

- (i) L_m, L_τ, L_u and L_π are symmetric positive semidefinite matrices in $\mathbb{R}^{N_L \times N_L}$;
- (ii) for all $\pi \in \mathbb{R}^{N_L}$,

$$\mathcal{E}_{\text{kin}}(\pi) = \frac{1}{2} \pi^\top L_\pi \pi.$$

Proof. (i) For any matrix D and diagonal matrix $W \succeq 0$, the matrix $L := D^\top W D$ is symmetric and positive semidefinite:

$$L^\top = (D^\top W D)^\top = D^\top W^\top D = D^\top W D = L,$$

and for any x ,

$$x^\top L x = x^\top D^\top W D x = (D x)^\top W (D x) \geq 0,$$

since W has nonnegative diagonal entries. Applying this with $(D, W) = (D_m, W_m)$, (D_τ, W_τ) and (D_u, W_u) yields the claimed properties for L_m, L_τ, L_u . A nonnegative linear combination of symmetric positive semidefinite matrices is again symmetric positive semidefinite, so L_π is symmetric positive semidefinite.

(ii) By definition of L_m, L_τ, L_u ,

$$\begin{aligned} (D_m \pi)^\top W_m (D_m \pi) &= \pi^\top L_m \pi, & (D_\tau \pi)^\top W_\tau (D_\tau \pi) &= \pi^\top L_\tau \pi, \\ (D_u \pi)^\top W_u (D_u \pi) &= \pi^\top L_u \pi. \end{aligned}$$

Therefore

$$\mathcal{E}_{\text{kin}}(\pi) = \frac{1}{2} (\kappa_m \pi^\top L_m \pi + \kappa_\tau \pi^\top L_\tau \pi + \kappa_u \pi^\top L_u \pi) = \frac{1}{2} \pi^\top L_\pi \pi,$$

which proves the claim. $\square$

Consequently, $\mathcal{E}_{\text{kin}}$ is a convex quadratic functional on $\mathbb{R}^{N_L}$, with flat directions corresponding to fog configurations that are constant along connected components of the underlying graph (if all $\kappa_m, \kappa_\tau, \kappa_u > 0$ and the graph is connected with Neumann boundary, the constant vector lies in the kernel of L_π).

12.5.3 Potential energy and band-aware penalisation

We now introduce a nonnegative potential field V on $\mathcal{L}_\Omega$ that penalises fog mass far from the local bid-ask tubes and from basic price bounds.

Definition 27 (Band and range potential). For each quote q lying on the patch, let (m_q, τ_q) be its location and $[b_q, a_q]$ its cleaned forward-discounted bid–ask band, and choose a representative grid node $(i_q, j_q) \in \Omega$ (e.g. the nearest neighbour in Ω).

Fix parameters $\alpha_{\text{band}} \geq 0$ and $\alpha_{\text{range}} \geq 0$. Define the *band potential* as

$$V_{i,j,k}^{\text{band}} := \begin{cases} \alpha_{\text{band}} \text{dist}(u_k, [b_q, a_q])^2, & \text{if } (i, j) = (i_q, j_q) \text{ for some quote } q, \\ 0, & \text{otherwise,} \end{cases}$$

for all $(i, j, k) \in \mathcal{L}_\Omega$, where $\text{dist}(u, [b, a]) := \max\{b - u, 0, u - a\}$ is the Euclidean distance from u to the interval $[b, a]$.

Define the *range potential* by

$$V_{i,j,k}^{\text{range}} := \alpha_{\text{range}} \left(\mathbf{1}_{\{u_k < 0\}} + \mathbf{1}_{\{u_k > F_{i,j}\}} \right),$$

where $F_{i,j}$ is the forward at node (i, j) and $\mathbf{1}_A$ is the indicator of the event A .

Finally, set

$$V_{i,j,k} := V_{i,j,k}^{\text{band}} + V_{i,j,k}^{\text{range}}, \quad (i, j, k) \in \mathcal{L}_\Omega.$$

By construction, $V_{i,j,k} \geq 0$ for all (i, j, k) . The band potential is small when u_k lies inside the bid–ask interval associated with the quote at (i_q, j_q) , and grows quadratically as u_k moves away from that interval; it is zero at grid nodes that are not directly associated with quotes. The range potential imposes a hard penalty α_{range} whenever u_k lies below zero or above the local forward $F_{i,j}$, discouraging fog from sitting at obviously unreasonable price levels.

It is convenient to collect the potential values into a vector $V \in \mathbb{R}^{N_L}$ by setting $V_\ell := V_{i_\ell, j_\ell, k_\ell}$ for each lattice index $\ell = 1, \dots, N_L$, and to define the diagonal matrix $\text{diag}(V) \in \mathbb{R}^{N_L \times N_L}$ with entries $(\text{diag}(V))_{\ell\ell} = V_\ell$.

Definition 28 (Potential energy of the fog). The *potential energy* of a fog configuration $\pi \in \mathbb{R}^{N_L}$ is

$$\mathcal{E}_{\text{pot}}(\pi) := \frac{1}{2} \sum_{(i,j) \in \Omega} \sum_{k=1}^{n_u} V_{i,j,k} \pi_{i,j,k}^2.$$

Equivalently, in vector notation,

$$\mathcal{E}_{\text{pot}}(\pi) = \frac{1}{2} \pi^\top \text{diag}(V) \pi.$$

Because $V_{i,j,k} \geq 0$ for all (i, j, k) , the matrix $\text{diag}(V)$ is symmetric positive semidefinite, and $\mathcal{E}_{\text{pot}}$ is a convex quadratic functional. Note that $\mathcal{E}_{\text{pot}}(\pi)$ penalises large values of $\pi_{i,j,k}$ at lattice sites where $V_{i,j,k}$ is large, i.e. far outside the band tube or the basic price range; it is indifferent to the sign of $\pi_{i,j,k}$ as a quadratic form, but in our optimisation the fog variables are constrained to be nonnegative and to lie on the simplex $\mathcal{C}_\pi(\Omega)$.

12.5.4 Hamiltonian energy and basic properties

We now combine kinetic and potential contributions into a single Hamiltonian energy.

Definition 29 (Hamiltonian matrix and energy). Define the *Hamiltonian matrix* by

$$H_\pi := L_\pi + \text{diag}(V) \in \mathbb{R}^{N_L \times N_L},$$

where L_π is as in Proposition 14 and V is from Definition 27. The *Hamiltonian energy* of a fog configuration $\pi \in \mathbb{R}^{N_L}$ is the quadratic functional

$$\mathcal{E}_{\text{Ham}}(\pi) := \frac{1}{2} \pi^\top H_\pi \pi.$$

By construction,

$$\mathcal{E}_{\text{Ham}}(\pi) = \mathcal{E}_{\text{kin}}(\pi) + \mathcal{E}_{\text{pot}}(\pi).$$

Proposition 15 (Symmetry, positive semidefiniteness, and convexity). *The Hamiltonian matrix H_π is symmetric positive semidefinite. Consequently, $\mathcal{E}_{\text{Ham}} : \mathbb{R}^{N_L} \rightarrow \mathbb{R}_+$ is a convex quadratic functional. Moreover, if at least one of the following holds:*

- *the graph underlying L_π is connected and $\kappa_m + \kappa_\tau + \kappa_u > 0$, and $V_{i,j,k} > 0$ at least at one lattice site; or*
- *more generally, H_π is positive definite on the affine subspace $\{\pi \in \mathbb{R}^{N_L} : \sum_\ell \pi_\ell = 1\}$,*

then $\mathcal{E}_{\text{Ham}}$ is strictly convex on the simplex $\mathcal{C}_\pi(\Omega)$, and has a unique minimiser on $\mathcal{C}_\pi(\Omega)$.

Proof. By Proposition 14, L_π is symmetric positive semidefinite. The matrix $\text{diag}(V)$ is diagonal with nonnegative entries and hence symmetric positive semidefinite. Therefore their sum $H_\pi = L_\pi + \text{diag}(V)$ is symmetric positive semidefinite. For any $\pi \in \mathbb{R}^{N_L}$,

$$\mathcal{E}_{\text{Ham}}(\pi) = \frac{1}{2} \pi^\top H_\pi \pi \geq 0.$$

A quadratic form with positive semidefinite matrix is convex, so $\mathcal{E}_{\text{Ham}}$ is convex.

If H_π is positive definite on a subspace $S \subset \mathbb{R}^{N_L}$ (in particular, on the subspace tangent to the simplex), then the restriction of $\mathcal{E}_{\text{Ham}}$ to S is strictly convex. The simplex $\mathcal{C}_\pi(\Omega)$ lies in the affine hyperplane $\{\pi : \sum_\ell \pi_\ell = 1\}$, and the tangent space at any point of the simplex is the subspace $\{\delta\pi : \sum_\ell \delta\pi_\ell = 0\}$. If H_π is positive definite on this subspace, then $\mathcal{E}_{\text{Ham}}$ is strictly convex on $\mathcal{C}_\pi(\Omega)$, and a strictly convex continuous function on a compact convex set has a unique minimiser. The sufficient condition stated in the proposition ensures this property in typical settings. The detailed proof of positive definiteness on the tangent space depends on the connectivity of the graph and the support of V and is standard in the theory of weighted graph Laplacians plus diagonal potentials. $\square$

To justify the strict convexity statement in the “moreover” part, we record the standard argument that under the connectivity and positivity assumptions in the first bullet H_π is positive definite on the simplex tangent

$$T := \{\delta\pi \in \mathbb{R}^{N_L} : \mathbf{1}^\top \delta\pi = 0\}.$$

Assume that the underlying graph on L_Ω is connected and that $\kappa_m + \kappa_\tau + \kappa_u > 0$, so that L_π is a weighted graph Laplacian with $\ker L_\pi = \text{span}\{\mathbf{1}\}$. Suppose in addition that there exists at least one lattice site $\ell^\star$ with $V_{\ell^\star} > 0$. If $H_\pi \delta\pi = 0$, then

$$0 = \delta\pi^\top H_\pi \delta\pi = \delta\pi^\top L_\pi \delta\pi + \delta\pi^\top \text{diag}(V) \delta\pi,$$

and both terms on the right-hand side are nonnegative. Hence $\delta\pi \in \ker L_\pi \cap \ker \text{diag}(V)$. The first condition implies $\delta\pi = c\mathbf{1}$ for some $c \in \mathbb{R}$, while the second forces $\delta\pi_{\ell^\star} = 0$ and therefore $c = 0$. Thus $\delta\pi = 0$ is the only vector with $\delta\pi^\top H_\pi \delta\pi = 0$, so H_π has trivial kernel and is positive definite. In particular there is no nonzero $\delta\pi \in T$ with $\delta\pi^\top H_\pi \delta\pi = 0$, and E_{Ham} is strictly convex on $C_\pi(\Omega)$ and on its tangent space.

Remark 21 (Interpretation of the Hamiltonian energy). The kinetic energy $\mathcal{E}_{\text{kin}}(\pi)$ penalises large discrete gradients of the fog in the (m, τ, u) directions: it is large when π varies rapidly across neighbouring lattice sites and small when π is smooth. The potential energy $\mathcal{E}_{\text{pot}}(\pi)$ penalises fog mass located at lattice sites with large $V_{i,j,k}$, i.e. far outside bid-ask tubes or basic price ranges.

On a calm patch with tight bands and reasonable baseline fit, the minimum-energy fog tends to concentrate its mass at price levels u_k inside the local bands and within $[0, F_{i,j}]$, while remaining smooth across neighbouring nodes. On a stressed patch with conflicting quotes or strong local misfit, a portion of the fog may be forced to reside outside the bands; in that case $\mathcal{E}_{\text{Ham}}$ balances the cost of leaking mass out of the band against the cost of introducing sharp gradients in (m, τ, u) .

12.6 Noise-aware band term via the fog

We now couple the 3D fog π on $\mathcal{L}_\Omega$ with the nodal surface $u(u_I)$ at each quote on the patch. The aim is to obtain, for each quote q , a band penalty whose effective strength is modulated by the local fog mass outside the corresponding bid-ask band.

Throughout this section we fix a patch $\Omega \subset \mathcal{G}$, an interior price vector $u_I \in \mathbb{R}^{N_\Omega}$ with associated full nodal surface $u(u_I) \in \mathbb{R}^G$, and a fog configuration $\pi = (\pi_{i,j,k})_{(i,j,k) \in \mathcal{L}_\Omega} \in \mathbb{R}^{N_\Omega n_u}$.

12.6.1 Fog mass outside the band at a quote

We first define, for each quote q , the fraction of fog mass that lies on price levels outside the corresponding bid-ask interval.

Definition 30 (Index set of out-of-band levels at a quote). Let q be a quote associated with location (m_q, τ_q) and cleaned forward-discounted band $[b_q, a_q]$. Let $(i_q, j_q) \in \Omega$ be a fixed representative of (m_q, τ_q) on the patch grid. Recall that $\{u_k\}_{k=1}^{n_u}$ are the discrete price levels. The *out-of-band index set* at quote q is

$$\mathcal{K}_q^{\text{out}} := \{k \in \{1, \dots, n_u\} : u_k < b_q \text{ or } u_k > a_q\}.$$

Thus $\mathcal{K}_q^{\text{out}}$ collects exactly those vertical levels u_k which lie strictly below the bid or strictly above the ask at quote q .

Definition 31 (Local fog mass outside the band at a quote). For a fog configuration π , the *fog mass outside the band at quote q* is defined by

$$M_q(\pi) := \sum_{k \in \mathcal{K}_q^{\text{out}}} \pi_{i_q, j_q, k}.$$

By construction $\pi_{i,j,k} \geq 0$ for all $(i, j, k) \in \mathcal{L}_\Omega$ on the feasible set $\mathcal{C}_\pi(\Omega)$, hence $M_q(\pi) \geq 0$ for all quotes q . The quantity $M_q(\pi)$ should be interpreted as the local probability mass (or “fog thickness”) allocated by π to out-of-band price levels at quote q .

Remark 22 (Linearity of M_q). For each fixed q , the map $\pi \mapsto M_q(\pi)$ is linear: there exists a vector $c_q \in \mathbb{R}^{N_\Omega n_u}$ with entries

$$(c_q)_{i,j,k} = \begin{cases} 1, & \text{if } (i, j) = (i_q, j_q) \text{ and } k \in \mathcal{K}_q^{\text{out}}, \\ 0, & \text{otherwise,} \end{cases}$$

such that $M_q(\pi) = c_q^\top \pi$ for all π . In particular, M_q is both linear and continuous.

12.6.2 Band misfit and noise-aware penalty

We now recall the band misfit at a quote and introduce the noise-aware band penalty, whose strength is modulated by the local fog mass outside the band.

Definition 32 (Band misfit at a quote). Let $S \in \mathbb{R}^{Q \times G}$ be the fixed sampling operator mapping nodal values $u \in \mathbb{R}^G$ to model prices at quote locations. For a given nodal surface $u(u_I)$, the model price at quote q is

$$C_q(u) := (Su)_q.$$

The corresponding *band violation* is

$$d_q(u) := \text{dist}(C_q(u), [b_q, a_q]) = \max\{b_q - C_q(u), 0, C_q(u) - a_q\} \geq 0.$$

Since $u \mapsto C_q(u)$ is affine and $\text{dist}(\cdot, [b_q, a_q])$ is the pointwise maximum of three affine functions (see Definition 5), the composition $u \mapsto d_q(u)$ is a convex function on $\mathbb{R}^G$. Therefore $u_I \mapsto d_q(u(u_I))$ is also convex on $\mathbb{R}^{N_\Omega}$ because $u(u_I)$ depends affinely on u_I .

We now define the noise-aware band penalty, which couples the misfit $d_q(u)$ and the fog mass outside the band $M_q(\pi)$.

Definition 33 (Fog simplex on a patch). The fog feasible set on Ω is the probability simplex

$$\mathcal{C}_\pi(\Omega) := \left\{ \pi \in \mathbb{R}^{N_\Omega n_u} : \pi_{i,j,k} \geq 0 \ \forall (i, j, k) \in \mathcal{L}_\Omega, \quad \sum_{(i,j) \in \Omega} \sum_{k=1}^{n_u} \pi_{i,j,k} = 1 \right\}.$$

On $\mathcal{C}_\pi(\Omega)$, the quantity $M_q(\pi)$ defined in Definition 31 satisfies $0 \leq M_q(\pi) \leq 1$ for each q .

Definition 34 (Noise-aware band penalty at a quote). Fix parameters $\lambda_{\text{noise}} \geq 0$ and $\varepsilon > 0$. For a given fog π and quote q , define

$$\nu_q(\pi) := \varepsilon + M_q(\pi) = \varepsilon + \sum_{k \in \mathcal{K}_q^{\text{out}}} \pi_{i_q, j_q, k}.$$

Then $\nu_q(\pi) \in [\varepsilon, 1 + \varepsilon]$ for all $\pi \in \mathcal{C}_\pi(\Omega)$. Given an interior price vector $u_I \in \mathbb{R}^{N_\Omega}$, with associated nodal surface $u = u(u_I)$ and band violation $d_q(u)$, the *noise-aware band penalty* at quote q is

$$\phi_q(u_I, \pi) := \frac{d_q(u)^2}{\nu_q(\pi)} + \lambda_{\text{noise}} \nu_q(\pi).$$

Intuitively, $\nu_q(\pi)$ is a local “noise scale” at quote q : if the fog is almost entirely inside the band, then $M_q(\pi)$ is small and $\nu_q(\pi) \approx \varepsilon$, so any nonzero violation $d_q(u) > 0$ is heavily penalised by the term $d_q(u)^2/\nu_q(\pi)$. Conversely, if a significant fraction of the local fog mass lies outside the band, then $M_q(\pi)$ and hence $\nu_q(\pi)$ are larger, making violations $d_q(u) > 0$ cheaper; however, large $\nu_q(\pi)$ is itself penalised linearly through $\lambda_{\text{noise}}\nu_q(\pi)$.

12.6.3 Convexity of the noise-aware band term

We now establish joint convexity of the noise-aware band term in its two arguments (u_I, π) , which is crucial for the global convexity of the patch objective.

The key tool is the *perspective* of a convex function.

Definition 35 (Perspective of a convex function). Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function with $g(x) \geq 0$ for all $x \in \mathbb{R}$. The *perspective* of g is the function $\tilde{g} : \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}$ defined by

$$\tilde{g}(d, \nu) := \nu g\left(\frac{d}{\nu}\right), \quad \nu > 0.$$

Lemma 10 (Convexity of the perspective). *Let $g : \mathbb{R} \rightarrow [0, \infty)$ be convex. Then its perspective $\tilde{g}(d, \nu) = \nu g(d/\nu)$ is convex on $\mathbb{R} \times (0, \infty)$.*

Proof. This is a standard result in convex analysis; we recall the argument for completeness. Let (d_1, ν_1) and (d_2, ν_2) be in $\mathbb{R} \times (0, \infty)$ and let $\theta \in [0, 1]$. Set

$$(d, \nu) := \theta(d_1, \nu_1) + (1 - \theta)(d_2, \nu_2) = (\theta d_1 + (1 - \theta)d_2, \theta \nu_1 + (1 - \theta)\nu_2),$$

with $\nu > 0$ by convexity of $(0, \infty)$. Then

$$\frac{d}{\nu} = \frac{\theta \nu_1}{\nu} \frac{d_1}{\nu_1} + \frac{(1 - \theta) \nu_2}{\nu} \frac{d_2}{\nu_2},$$

where the coefficients

$$\alpha_1 := \frac{\theta \nu_1}{\nu}, \quad \alpha_2 := \frac{(1 - \theta) \nu_2}{\nu}$$

are nonnegative and satisfy $\alpha_1 + \alpha_2 = 1$. By convexity of g ,

$$g\left(\frac{d}{\nu}\right) \leq \alpha_1 g\left(\frac{d_1}{\nu_1}\right) + \alpha_2 g\left(\frac{d_2}{\nu_2}\right).$$

Multiplying both sides by $\nu > 0$ yields

$$\tilde{g}(d, \nu) = \nu g\left(\frac{d}{\nu}\right) \leq \theta \nu_1 g\left(\frac{d_1}{\nu_1}\right) + (1 - \theta) \nu_2 g\left(\frac{d_2}{\nu_2}\right) = \theta \tilde{g}(d_1, \nu_1) + (1 - \theta) \tilde{g}(d_2, \nu_2).$$

Thus $\tilde{g}$ is convex on $\mathbb{R} \times (0, \infty)$. □

We now apply this lemma with $g(x) = x^2$.

Proposition 16 (Convexity of the noise-aware band term). *Let q be a quote on the patch. The map*

$$(u_I, \pi) \mapsto \phi_q(u_I, \pi)$$

defined in Definition 34 is jointly convex on $\mathcal{C}_u(\Omega) \times \mathcal{C}_\pi(\Omega)$, where $\mathcal{C}_u(\Omega)$ and $\mathcal{C}_\pi(\Omega)$ are as in Definitions 21 and 33.

Proof. We proceed in steps.

(1) *Convexity of $d_q(u_I)$ in u_I .* The map $u_I \mapsto u(u_I)$ is affine by construction of the assembly map (equation (12.1)). The band misfit $d_q(u) = \text{dist}(C_q(u), [b_q, a_q])$ can be written as

$$d_q(u) = \max\{b_q - C_q(u), 0, C_q(u) - a_q\},$$

where $u \mapsto C_q(u)$ is affine. A pointwise maximum of finitely many affine functions is convex, hence $u \mapsto d_q(u)$ is convex on $\mathbb{R}^G$. Composing with the affine map $u(u_I)$, we obtain that

$$u_I \mapsto d_q(u(u_I))$$

is convex on $\mathbb{R}^{N_\Omega}$ and, in particular, on $\mathcal{C}_u(\Omega)$.

(2) *Affinity and positivity of $\nu_q(\pi)$ in π .* By Remark 22, $M_q(\pi)$ is a linear functional of π and is therefore affine. Adding the constant $\varepsilon > 0$, we obtain

$$\nu_q(\pi) = \varepsilon + M_q(\pi),$$

which is an affine function of π . On the simplex $\mathcal{C}_\pi(\Omega)$ we have $M_q(\pi) \geq 0$, hence

$$\nu_q(\pi) \geq \varepsilon > 0$$

for all $\pi \in \mathcal{C}_\pi(\Omega)$. Thus the pair $(d_q(u_I), \nu_q(\pi))$ always lies in $\mathbb{R} \times (0, \infty)$ on the feasible domain.

(3) *Convexity of $(d, \nu) \mapsto \frac{d^2}{\nu}$.* Consider the function $g : \mathbb{R} \rightarrow [0, \infty)$ defined by $g(x) = x^2$. It is convex and nonnegative. Its perspective is

$$\tilde{g}(d, \nu) := \nu g(d/\nu) = \frac{d^2}{\nu}, \quad \nu > 0.$$

By Lemma 10, $\tilde{g}$ is convex on $\mathbb{R} \times (0, \infty)$. Hence the map

$$(d, \nu) \mapsto \frac{d^2}{\nu}$$

is convex on $\mathbb{R} \times (0, \infty)$.

(4) *Convexity of $(d, \nu) \mapsto \frac{d^2}{\nu} + \lambda_{\text{noise}}\nu$.* The function

$$(d, \nu) \mapsto \lambda_{\text{noise}}\nu$$

is affine (hence convex) on $\mathbb{R} \times (0, \infty)$. The sum of a convex function and an affine function is convex, so the map

$$h(d, \nu) := \frac{d^2}{\nu} + \lambda_{\text{noise}}\nu$$

is convex on $\mathbb{R} \times (0, \infty)$. Moreover, for each fixed $\nu > 0$, the function $d \mapsto h(d, \nu)$ is nondecreasing on $[0, \infty)$, since

$$\frac{\partial}{\partial d} h(d, \nu) = \frac{2d}{\nu} \geq 0 \quad \text{for all } d \geq 0.$$

(5) Joint convexity of $\varphi_q(u_I, \pi)$. Let (u_I^1, π^1) and (u_I^2, π^2) be arbitrary points in $\mathcal{C}_u(\Omega) \times \mathcal{C}_\pi(\Omega)$ and let $\theta \in [0, 1]$. Define

$$(u_I^\theta, \pi^\theta) := \theta(u_I^1, \pi^1) + (1 - \theta)(u_I^2, \pi^2) \in \mathcal{C}_u(\Omega) \times \mathcal{C}_\pi(\Omega).$$

For $i = 1, 2$ set

$$d_i := d_q(u(u_I^i)), \quad \nu_i := \nu_q(\pi^i),$$

and similarly

$$d_\theta := d_q(u(u_I^\theta)), \quad \nu_\theta := \nu_q(\pi^\theta).$$

By construction $d_q(\cdot)$ is a distance to the interval $[b_q, a_q]$, hence $d_q(u) \geq 0$ for all u , and therefore $d_i, d_\theta \geq 0$. From (1), the map $u_I \mapsto d_q(u(u_I))$ is convex, so

$$d_\theta = d_q(u(u_I^\theta)) \leq \theta d_q(u(u_I^1)) + (1 - \theta) d_q(u(u_I^2)) = \theta d_1 + (1 - \theta) d_2.$$

From (2), $\nu_q(\pi)$ is affine in π , hence

$$\nu_\theta = \nu_q(\pi^\theta) = \theta \nu_q(\pi^1) + (1 - \theta) \nu_q(\pi^2) = \theta \nu_1 + (1 - \theta) \nu_2.$$

Recall that $\varphi_q(u_I, \pi) = h(d_q(u(u_I)), \nu_q(\pi))$ with h convex on $\mathbb{R} \times (0, \infty)$ by (4), and that for each $\nu > 0$ the map $d \mapsto h(d, \nu)$ is nondecreasing on $[0, \infty)$. Using $d_\theta \geq 0$ and $d_\theta \leq \theta d_1 + (1 - \theta) d_2$, monotonicity in the first argument yields

$$h(d_\theta, \nu_\theta) \leq h(\theta d_1 + (1 - \theta) d_2, \nu_\theta) = h(\theta d_1 + (1 - \theta) d_2, \theta \nu_1 + (1 - \theta) \nu_2).$$

The pair on the right-hand side is exactly the convex combination

$$(\theta d_1 + (1 - \theta) d_2, \theta \nu_1 + (1 - \theta) \nu_2) = \theta(d_1, \nu_1) + (1 - \theta)(d_2, \nu_2).$$

By convexity of h we therefore have

$$h(\theta d_1 + (1 - \theta) d_2, \theta \nu_1 + (1 - \theta) \nu_2) \leq \theta h(d_1, \nu_1) + (1 - \theta) h(d_2, \nu_2).$$

Combining the two inequalities gives

$$h(d_\theta, \nu_\theta) \leq \theta h(d_1, \nu_1) + (1 - \theta) h(d_2, \nu_2).$$

Rewriting in terms of φ_q ,

$$\varphi_q(u_I^\theta, \pi^\theta) = h(d_\theta, \nu_\theta) \leq \theta h(d_1, \nu_1) + (1 - \theta) h(d_2, \nu_2) = \theta \varphi_q(u_I^1, \pi^1) + (1 - \theta) \varphi_q(u_I^2, \pi^2).$$

This is exactly the defining inequality for joint convexity of $(u_I, \pi) \mapsto \varphi_q(u_I, \pi)$ on $\mathcal{C}_u(\Omega) \times \mathcal{C}_\pi(\Omega)$. $\square$

Remark 23 (Interpretation of the noise-aware band penalty). The penalty $\phi_q(u_I, \pi)$ can be seen as an adaptive band penalty whose effective stiffness is controlled by the fog. When the local fog mass outside the band is small ($M_q(\pi) \approx 0$, thus $\nu_q(\pi) \approx \varepsilon$), any nonzero violation $d_q(u)$ incurs a large cost $d_q(u)^2/\nu_q(\pi)$, forcing the surface u to stay tightly inside the band. When the fog assigns significant mass to out-of-band price levels ($M_q(\pi)$ large), violations become cheaper but increase the “noise budget” $\lambda_{\text{noise}} \nu_q(\pi)$. The optimiser can therefore treat a subset of quotes as noisy (outliers) by allowing the fog to populate out-of-band regions, but must pay a linear cost for doing so, while still operating within a globally convex framework.

12.7 Surface energy and closeness to the baseline

On each patch Ω we regularise the surface in two complementary ways:

- (i) by penalising roughness of the implied risk-neutral density in a patch-level influence region,
- (ii) by penalising deviations from the baseline nodal values on Ω .

Both terms are quadratic in the interior vector u_I and yield convex contributions to the patch objective.

12.7.1 Discrete density operator and patch restriction

Recall that the (continuum) risk-neutral density associated with the forward- discounted call surface $C_f(K, \tau)$ is

$$\rho(K, \tau) := \partial_{KK} C_f(K, \tau).$$

On the nodal grid $\mathcal{G}$, and on any additional grid used to represent densities (for instance a collocation grid in (K, τ)), the Breeden-Litzenberger relation is implemented by a fixed linear operator that maps nodal prices to discretised densities.

Assumption 2 (Global discrete density operator). *There exists a finite set of density evaluation points*

$$\mathcal{G}_\rho = \{(K_r, \tau_r)\}_{r=1}^{N_\rho},$$

and a matrix $D_\rho \in \mathbb{R}^{N_\rho \times G}$ such that, for every nodal vector $u \in \mathbb{R}^G$, the vector

$$\rho(u) := D_\rho u \in \mathbb{R}^{N_\rho}$$

represents the discrete risk-neutral density evaluated at the points in $\mathcal{G}_\rho$.

We are interested only in those density points that are influenced by the patch Ω , namely points whose density values depend (possibly together with off-patch values) on at least one interior node in Ω .

Definition 36 (Patch influence region in density space). Let $P_\Omega \in \mathbb{R}^{G \times N_\Omega}$ be the patch assembly matrix from (12.1), which injects an interior vector $u_I \in \mathbb{R}^{N_\Omega}$ into the full nodal vector. We write the full surface as

$$u(u_I) = P_\Omega u_I + u^{0, \text{off}},$$

where $u^{0, \text{off}} \in \mathbb{R}^G$ is the off-patch baseline contribution ($u_{i,j}^{0, \text{off}} = 0$ for $(i, j) \in \Omega$ and $u_{i,j}^{0, \text{off}} = u_{i,j}^0$ otherwise).

Define the *patch influence index set* in density space by

$$\mathcal{I}_\rho(\Omega) := \left\{ r \in \{1, \dots, N_\rho\} : (D_\rho P_\Omega)_{r,\cdot} \neq 0 \right\},$$

namely those density rows whose value depends on at least one interior node in Ω . Let $N_{\rho, \Omega} := |\mathcal{I}_\rho(\Omega)|$, and define the restriction operator $R_\Omega \in \mathbb{R}^{N_{\rho, \Omega} \times N_\rho}$ that extracts the components with indices in $\mathcal{I}_\rho(\Omega)$.

Thus, given u_I , the vector $R_\Omega \rho(u(u_I))$ collects precisely those density values that are affected by the patch Ω .

Definition 37 (Patch-level density map). With D_ρ and R_Ω as above, define

$$\rho_\Omega(u_I) := R_\Omega \rho(u(u_I)) = R_\Omega D_\rho (P_\Omega u_I + u^{0,\text{off}}).$$

We write this as an affine map

$$\rho_\Omega(u_I) = B_\Omega u_I + \rho_{\text{off}},$$

where

$$B_\Omega := R_\Omega D_\rho P_\Omega \in \mathbb{R}^{N_{\rho,\Omega} \times N_\Omega}, \quad \rho_{\text{off}} := R_\Omega D_\rho u^{0,\text{off}} \in \mathbb{R}^{N_{\rho,\Omega}}.$$

By construction, $\rho_\Omega(u_I)$ collects exactly the density values on the patch influence region, and depends affinely on the interior vector u_I .

12.7.2 Surface density energy

We now penalise rough or irregular density configurations on the patch influence region via a quadratic form in $\rho_\Omega(u_I)$.

Definition 38 (Surface density energy). Let $H_\rho \in \mathbb{R}^{N_{\rho,\Omega} \times N_{\rho,\Omega}}$ be a fixed symmetric positive semidefinite matrix, $H_\rho \succeq 0$. For example, H_ρ may encode a discrete H^{-1} -type smoothing operator or a weighted graph Laplacian on $\mathcal{G}_\rho$ restricted to $\mathcal{I}_\rho(\Omega)$. The *surface density energy* associated with an interior vector u_I is

$$E_{\text{surf}}(u_I) := \frac{1}{2} \rho_\Omega(u_I)^\top H_\rho \rho_\Omega(u_I).$$

Because $\rho_\Omega(u_I) = B_\Omega u_I + \rho_{\text{off}}$ is affine in u_I , E_{surf} is a quadratic functional of u_I . We make this explicit.

Proposition 17 (Quadratic form and convexity of E_{surf}). *The surface density energy can be written as*

$$E_{\text{surf}}(u_I) = \frac{1}{2} u_I^\top Q_\rho u_I + c_\rho^\top u_I + c_0,$$

where

$$Q_\rho := B_\Omega^\top H_\rho B_\Omega \succeq 0, \quad c_\rho := B_\Omega^\top H_\rho \rho_{\text{off}}, \quad c_0 := \frac{1}{2} \rho_{\text{off}}^\top H_\rho \rho_{\text{off}}.$$

In particular, E_{surf} is a convex quadratic function of u_I with Hessian Q_ρ .

Proof. Substituting the affine form $\rho_\Omega(u_I) = B_\Omega u_I + \rho_{\text{off}}$ into Definition 38, we obtain

$$E_{\text{surf}}(u_I) = \frac{1}{2} (B_\Omega u_I + \rho_{\text{off}})^\top H_\rho (B_\Omega u_I + \rho_{\text{off}}).$$

Expanding the quadratic form yields

$$E_{\text{surf}}(u_I) = \frac{1}{2} u_I^\top B_\Omega^\top H_\rho B_\Omega u_I + u_I^\top B_\Omega^\top H_\rho \rho_{\text{off}} + \frac{1}{2} \rho_{\text{off}}^\top H_\rho \rho_{\text{off}}.$$

Identifying

$$Q_\rho := B_\Omega^\top H_\rho B_\Omega, \quad c_\rho := B_\Omega^\top H_\rho \rho_{\text{off}}, \quad c_0 := \frac{1}{2} \rho_{\text{off}}^\top H_\rho \rho_{\text{off}},$$

we obtain the claimed quadratic representation. Since $H_\rho \succeq 0$, we have for any $x \in \mathbb{R}^{N_\Omega}$,

$$x^\top Q_\rho x = x^\top B_\Omega^\top H_\rho B_\Omega x = (B_\Omega x)^\top H_\rho (B_\Omega x) \geq 0.$$

Thus $Q_\rho \succeq 0$, and the Hessian of E_{surf} with respect to u_I is positive semidefinite. A quadratic function with positive semidefinite Hessian is convex, hence E_{surf} is convex in u_I . $\square$

In particular, E_{surf} penalises interior configurations u_I that produce “rough” or oscillatory risk-neutral densities in the patch influence region, with the exact notion of roughness encoded by H_ρ .

12.7.3 Closeness to the baseline

We also penalise departures of the patch interior from the baseline nodal values, in order to avoid gratuitous changes that are not required by the data and no-arbitrage constraints.

Definition 39 (Closeness to the baseline). Let $u_I^0 \in \mathbb{R}^{N_\Omega}$ be the vector of baseline nodal values on Ω , extracted from u^0 , and fix a parameter $\lambda_{\text{cl}} > 0$. The *closeness* (or Tikhonov) term on Ω is

$$E_{\text{cl}}(u_I) := \frac{\lambda_{\text{cl}}}{2} \|u_I - u_I^0\|_2^2 = \frac{\lambda_{\text{cl}}}{2} (u_I - u_I^0)^\top (u_I - u_I^0).$$

This is a standard ℓ^2 -type regulariser that penalises deviations from the baseline. Its convexity and strict positive definiteness are immediate.

Proposition 18 (Strict convexity of E_{cl}). *The functional $E_{\text{cl}} : \mathbb{R}^{N_\Omega} \rightarrow \mathbb{R}_+$ is a strictly convex quadratic function of u_I with Hessian $\lambda_{\text{cl}} I_{N_\Omega} \succ 0$.*

Proof. Expanding the square, we have

$$E_{\text{cl}}(u_I) = \frac{\lambda_{\text{cl}}}{2} (u_I^\top u_I - 2u_I^\top u_I^0 + u_I^{0\top} u_I^0),$$

so

$$E_{\text{cl}}(u_I) = \frac{\lambda_{\text{cl}}}{2} u_I^\top u_I - \lambda_{\text{cl}} u_I^{0\top} u_I + \frac{\lambda_{\text{cl}}}{2} u_I^{0\top} u_I^0.$$

The Hessian with respect to u_I is $\lambda_{\text{cl}} I_{N_\Omega}$, which is positive definite since $\lambda_{\text{cl}} > 0$. A quadratic functional with positive definite Hessian is strictly convex, so E_{cl} is strictly convex on $\mathbb{R}^{N_\Omega}$. $\square$

Remark 24 (Combined surface regularisation). Both E_{surf} and E_{cl} are convex quadratic functionals of u_I . The combined surface regulariser

$$u_I \mapsto E_{\text{cl}}(u_I) + \lambda_{\text{surf}} E_{\text{surf}}(u_I), \quad \lambda_{\text{surf}} \geq 0,$$

is therefore convex. If either $\lambda_{\text{cl}} > 0$ or the matrix $Q_\rho = B_\Omega^\top H_\rho B_\Omega$ is positive definite on the relevant subspace, the combined regulariser is strictly convex, which contributes to uniqueness of the patch-level minimiser.

12.8 Patch-level post-fit optimisation problem

We now assemble the various ingredients introduced above into a single patch-level objective and formulate the post-fit optimisation problem on a patch $\Omega \in \mathcal{G}$.

12.8.1 Fog feasible set and quote index set

Recall that on Ω the fog variables are

$$\pi = (\pi_{i,j,k})_{(i,j,k) \in \mathcal{L}_\Omega} \in \mathbb{R}^{N_\Omega n_u},$$

where $\pi_{i,j,k}$ represents the fog mass at (m_i, τ_j, u_k) and $\mathcal{L}_\Omega = \{(i, j, k) : (i, j) \in \Omega, k = 1, \dots, n_u\}$.

Definition 40 (Fog feasible set on a patch). The *fog feasible set* on Ω is the probability simplex

$$\mathcal{C}_\pi(\Omega) := \left\{ \pi \in \mathbb{R}^{N_\Omega n_u} : \pi_{i,j,k} \geq 0 \ \forall (i, j, k) \in \mathcal{L}_\Omega, \quad \sum_{(i,j) \in \Omega} \sum_{k=1}^{n_u} \pi_{i,j,k} = 1 \right\}.$$

Lemma 11 (Geometry of $\mathcal{C}_\pi(\Omega)$). *The set $\mathcal{C}_\pi(\Omega)$ is a nonempty, compact, convex polytope in $\mathbb{R}^{N_\Omega n_u}$.*

Proof. Nonemptiness is obvious, for example the uniform vector $\pi_{i,j,k} = 1/(N_\Omega n_u)$ belongs to $\mathcal{C}_\pi(\Omega)$. The constraints defining $\mathcal{C}_\pi(\Omega)$ consist of finitely many linear equalities and inequalities:

$$\pi_{i,j,k} \geq 0, \quad \sum_{(i,j) \in \Omega} \sum_{k=1}^{n_u} \pi_{i,j,k} = 1.$$

Thus $\mathcal{C}_\pi(\Omega)$ is the intersection of a finite number of closed half-spaces (one per inequality) and a hyperplane (the equality constraint), so it is a closed convex polyhedron. The additional equality fixing the total mass to 1 implies boundedness: all coordinates are nonnegative and sum to 1, so $0 \leq \pi_{i,j,k} \leq 1$ for every (i, j, k) . A closed and bounded subset of $\mathbb{R}^{N_\Omega n_u}$ is compact. Being a bounded polyhedron, it is in fact a polytope. $\square$

We also need to know which quotes interact with a given patch.

Definition 41 (Quote index set attached to a patch). Let $S \in \mathbb{R}^{Q \times G}$ be the sampling operator mapping nodal prices to quote locations, so that $C_q(u) = (Su)_q$ for $q = 1, \dots, Q$. Each row of S has finite support (the interpolation stencil of that quote). We define

$$Q_\Omega := \left\{ q \in \{1, \dots, Q\} : \text{the stencil of row } q \text{ of } S \text{ intersects } \Omega \right\}.$$

Equivalently, $q \in Q_\Omega$ if and only if there exists $(i, j) \in \Omega$ such that the nodal value $u_{i,j}$ enters $(Su)_q$ with nonzero weight.

Thus Q_Ω collects exactly those quotes whose model prices depend on at least one interior node of Ω ; the remaining quotes are insensitive to changes on Ω and need not appear in the patch objective.

12.8.2 Patch energy functional

We now define the patch-level energy as a sum of four components: noise-aware band penalties, closeness-to-baseline, density regularisation, and Hamiltonian energy of the fog. Recall:

- $\phi_q(u_I, \pi)$ is the noise-aware band penalty at quote q , defined in Definition 34, with $u = u(u_I)$ assembled from u_I ;
- $E_{\text{cl}}(u_I)$ is the closeness energy from Definition 39;
- $E_{\text{surf}}(u_I)$ is the surface density energy from Definition 38;
- $\mathcal{E}_{\text{Ham}}(\pi)$ is the Hamiltonian energy of the fog from Definition 29.

Definition 42 (Patch energy functional). Fix nonnegative weights

$$\lambda_{\text{noise}}, \lambda_{\text{surf}}, \lambda_{\pi} \geq 0,$$

and recall that $\lambda_{\text{cl}} > 0$ is part of the definition of E_{cl} . For an interior vector $u_I \in \mathcal{C}_u(\Omega)$ and a fog $\pi \in \mathcal{C}_{\pi}(\Omega)$, the *patch energy* is the functional $J_{\Omega} : \mathcal{C}_u(\Omega) \times \mathcal{C}_{\pi}(\Omega) \rightarrow \mathbb{R}$ defined by

$$\begin{aligned} J_{\Omega}(u_I, \pi) &:= \sum_{q \in Q_{\Omega}} \phi_q(u_I, \pi) + E_{\text{cl}}(u_I) + \lambda_{\text{surf}} E_{\text{surf}}(u_I) + \lambda_{\pi} \mathcal{E}_{\text{Ham}}(\pi) \\ &= \sum_{q \in Q_{\Omega}} \left(\frac{d_q(u)^2}{\varepsilon + M_q(\pi)} + \lambda_{\text{noise}}(\varepsilon + M_q(\pi)) \right) \\ &\quad + \frac{\lambda_{\text{cl}}}{2} \|u_I - u_I^0\|_2^2 + \frac{\lambda_{\text{surf}}}{2} \rho_{\Omega}(u_I)^{\top} H_{\rho} \rho_{\Omega}(u_I) + \frac{\lambda_{\pi}}{2} \pi^{\top} H_{\pi} \pi, \end{aligned} \tag{12.3}$$

where $u = u(u_I)$, $d_q(u)$ is the band violation at quote q (Definition 32) and $M_q(\pi)$ is the fog mass outside the band at quote q (Definition 31).

Remark 25 (Well-definedness and continuity of J_{Ω}). Since $\pi \in \mathcal{C}_{\pi}(\Omega)$ implies $M_q(\pi) \geq 0$ and $\varepsilon > 0$ by construction, the denominators $\varepsilon + M_q(\pi)$ are bounded away from zero and all terms in (12.3) are finite. Each component $\phi_q(u_I, \pi)$, $E_{\text{cl}}(u_I)$, $E_{\text{surf}}(u_I)$, and $\mathcal{E}_{\text{Ham}}(\pi)$ is continuous in its arguments. Therefore J_{Ω} is a continuous real-valued function on $\mathcal{C}_u(\Omega) \times \mathcal{C}_{\pi}(\Omega)$.

We now establish the basic convexity property of J_{Ω} .

Proposition 19 (Convexity of the patch energy). *For any feasible patch Ω (i.e. $\mathcal{C}_u(\Omega) \neq \emptyset$), the patch energy functional J_{Ω} is jointly convex in (u_I, π) on $\mathcal{C}_u(\Omega) \times \mathcal{C}_{\pi}(\Omega)$.*

Proof. By Proposition 16, each individual noise-aware band term $\phi_q(u_I, \pi)$ is jointly convex in (u_I, π) on $\mathcal{C}_u(\Omega) \times \mathcal{C}_{\pi}(\Omega)$. By Proposition 17, $E_{\text{surf}}(u_I)$ is a convex quadratic functional of u_I . By Proposition 18, $E_{\text{cl}}(u_I)$ is a strictly convex quadratic functional of u_I (hence convex). By Proposition 15, $\mathcal{E}_{\text{Ham}}(\pi)$ is a convex quadratic functional of π .

Multiplying convex functionals by nonnegative scalars preserves convexity, and summing finitely many convex functionals yields a convex functional. Therefore, for each (u_I, π) in the convex set $\mathcal{C}_u(\Omega) \times \mathcal{C}_{\pi}(\Omega)$, the map

$$(u_I, \pi) \mapsto J_{\Omega}(u_I, \pi)$$

is convex. This proves the claim. $\square$

12.8.3 Patch-level post-fit problem: existence and uniqueness

We can now formulate the patch-level convex optimisation problem.

Definition 43 (Patch-level post-fit problem). Let $\Omega \subset \mathcal{G}$ be a feasible patch (i.e. $\mathcal{C}_u(\Omega) \neq \emptyset$). The *patch-level post-fit problem* is the constrained optimisation problem

$$\min \{ J_\Omega(u_I, \pi) : u_I \in \mathcal{C}_u(\Omega), \pi \in \mathcal{C}_\pi(\Omega) \}.$$

Any pair $(u_I^*, \pi^*) \in \mathcal{C}_u(\Omega) \times \mathcal{C}_\pi(\Omega)$ achieving this minimum is called a *patch minimiser*.

We now show that at least one minimiser exists under our standing assumptions.

Proposition 20 (Existence of patch-level minimisers). *Assume that $\mathcal{C}_u(\Omega) \neq \emptyset$ and that $\lambda_{\text{cl}} > 0$ (as in Definition 39). Then the patch problem admits at least one minimiser (u_I^*, π^*) .*

Proof. The feasible set

$$\mathcal{F}_\Omega := \mathcal{C}_u(\Omega) \times \mathcal{C}_\pi(\Omega) \subset \mathbb{R}^{N_\Omega} \times \mathbb{R}^{N_\Omega n_u}$$

is nonempty by assumption on $\mathcal{C}_u(\Omega)$ and Lemma 11. By Proposition 13, $\mathcal{C}_u(\Omega)$ is a closed convex polyhedron in $\mathbb{R}^{N_\Omega}$ and may be unbounded. By Lemma 11, $\mathcal{C}_\pi(\Omega)$ is a compact convex polytope in $\mathbb{R}^{N_\Omega n_u}$. Hence $\mathcal{F}_\Omega$ is closed, convex, and nonempty, but not necessarily bounded.

To apply the Weierstrass theorem, we consider sublevel sets of J_Ω . From (12.3), using $E_{\text{cl}}(u_I) = \frac{\lambda_{\text{cl}}}{2} \|u_I - u_I^0\|_2^2$ and nonnegativity of all other terms, we have

$$J_\Omega(u_I, \pi) \geq E_{\text{cl}}(u_I) = \frac{\lambda_{\text{cl}}}{2} \|u_I - u_I^0\|_2^2.$$

Let $m := \inf_{\mathcal{F}_\Omega} J_\Omega$ denote the infimum of J_Ω on the feasible set, which is finite because $J_\Omega \geq 0$ and the baseline pair $(u_I^0, \pi^{\text{ref}})$ (with any fixed $\pi^{\text{ref}} \in \mathcal{C}_\pi(\Omega)$) belongs to $\mathcal{F}_\Omega$. For any $\alpha > m$, consider the sublevel set

$$\mathcal{F}_\Omega(\alpha) := \{(u_I, \pi) \in \mathcal{F}_\Omega : J_\Omega(u_I, \pi) \leq \alpha\}.$$

By the inequality above,

$$\frac{\lambda_{\text{cl}}}{2} \|u_I - u_I^0\|_2^2 \leq J_\Omega(u_I, \pi) \leq \alpha \implies \|u_I - u_I^0\|_2^2 \leq \frac{2\alpha}{\lambda_{\text{cl}}}.$$

Thus, for any $(u_I, \pi) \in \mathcal{F}_\Omega(\alpha)$, the interior vector u_I lies in the closed Euclidean ball of radius $\sqrt{2\alpha/\lambda_{\text{cl}}}$ centred at u_I^0 . The fog variable π always lies in the compact set $\mathcal{C}_\pi(\Omega)$ by definition of $\mathcal{F}_\Omega$. It follows that $\mathcal{F}_\Omega(\alpha)$ is bounded.

Moreover, $\mathcal{F}_\Omega(\alpha)$ is closed: it is the intersection of the closed set $\mathcal{F}_\Omega$ with the closed inverse image $\{(u_I, \pi) : J_\Omega(u_I, \pi) \leq \alpha\}$ of $(-\infty, \alpha]$ under the continuous map $(u_I, \pi) \mapsto J_\Omega(u_I, \pi)$. Hence $\mathcal{F}_\Omega(\alpha)$ is compact.

By construction m is the infimum of J_Ω over $\mathcal{F}_\Omega$, so there exists a sequence $(u_I^{(n)}, \pi^{(n)})$ in $\mathcal{F}_\Omega$ such that $J_\Omega(u_I^{(n)}, \pi^{(n)}) \downarrow m$ as $n \rightarrow \infty$. All but finitely many of these points lie in $\mathcal{F}_\Omega(\alpha)$ for any fixed $\alpha > m$. By compactness of $\mathcal{F}_\Omega(\alpha)$, the sequence has a convergent subsequence $(u_I^{(n_k)}, \pi^{(n_k)})$ with limit $(u_I^*, \pi^*) \in \mathcal{F}_\Omega(\alpha) \subset \mathcal{F}_\Omega$. Continuity of J_Ω implies

$$J_\Omega(u_I^*, \pi^*) = \lim_{k \rightarrow \infty} J_\Omega(u_I^{(n_k)}, \pi^{(n_k)}) = m.$$

Thus (u_I^*, π^*) attains the infimum and is a minimiser of J_Ω on $\mathcal{F}_\Omega$. $\square$

We now provide sufficient conditions for uniqueness of the patch minimiser.

Proposition 21 (Uniqueness under strict convexity). *Suppose, in addition to the assumptions of Proposition 20, that:*

1. *The quadratic form in u_I given by*

$$u_I \mapsto E_{\text{cl}}(u_I) + \lambda_{\text{surf}} E_{\text{surf}}(u_I)$$

is strictly convex on the affine hull of $\mathcal{C}_u(\Omega)$; equivalently, its Hessian $\lambda_{\text{cl}} I_{N_\Omega} + \lambda_{\text{surf}} Q_\Omega$ is positive definite on the tangent cone of $\mathcal{C}_u(\Omega)$.

2. *$\lambda_\pi > 0$ and the Hamiltonian matrix H_π is positive definite on the affine hull of $\mathcal{C}_\pi(\Omega)$; equivalently, the quadratic form $\pi \mapsto \pi^\top H_\pi \pi$ is strictly convex on $\mathcal{C}_\pi(\Omega)$.*

Then the minimiser (u_I^*, π^*) of J_Ω on $\mathcal{C}_u(\Omega) \times \mathcal{C}_\pi(\Omega)$ is unique.

Proof. By assumption (1), the map

$$u_I \mapsto E_{\text{cl}}(u_I) + \lambda_{\text{surf}} E_{\text{surf}}(u_I)$$

is strictly convex on $\mathcal{C}_u(\Omega)$. By assumption (2), the map

$$\pi \mapsto \lambda_\pi \mathcal{E}_{\text{Ham}}(\pi) = \frac{\lambda_\pi}{2} \pi^\top H_\pi \pi$$

is strictly convex on $\mathcal{C}_\pi(\Omega)$. The remaining contribution to J_Ω is the sum of noise-aware band terms

$$(u_I, \pi) \mapsto \sum_{q \in Q_\Omega} \phi_q(u_I, \pi),$$

which is convex by Proposition 19 (and does not affect strict convexity, since adding a convex function to a strictly convex one preserves strict convexity).

To see that J_Ω is strictly convex on the product set $\mathcal{C}_u(\Omega) \times \mathcal{C}_\pi(\Omega)$, let $(u_I^{(1)}, \pi^{(1)})$ and $(u_I^{(2)}, \pi^{(2)})$ be two distinct feasible points, and let $\theta \in (0, 1)$. Then at least one of the components $u_I^{(1)}$ and $u_I^{(2)}$ differs, or $\pi^{(1)}$ and $\pi^{(2)}$ differ. If $u_I^{(1)} \neq u_I^{(2)}$, strict convexity of the u_I -quadratic implies

$$\begin{aligned} & E_{\text{cl}}(\theta u_I^{(1)} + (1 - \theta) u_I^{(2)}) + \lambda_{\text{surf}} E_{\text{surf}}(\theta u_I^{(1)} + (1 - \theta) u_I^{(2)}) \\ & < \theta (E_{\text{cl}}(u_I^{(1)}) + \lambda_{\text{surf}} E_{\text{surf}}(u_I^{(1)})) + (1 - \theta) (E_{\text{cl}}(u_I^{(2)}) + \lambda_{\text{surf}} E_{\text{surf}}(u_I^{(2)})). \end{aligned}$$

If instead $u_I^{(1)} = u_I^{(2)}$ but $\pi^{(1)} \neq \pi^{(2)}$, strict convexity of $\lambda_\pi \mathcal{E}_{\text{Ham}}$ on $\mathcal{C}_\pi(\Omega)$ implies a strict inequality in the π -component. In either case, adding the convex sum of band terms preserves strict inequality:

$$\begin{aligned} & J_\Omega(\theta u_I^{(1)} + (1 - \theta) u_I^{(2)}, \theta \pi^{(1)} + (1 - \theta) \pi^{(2)}) \\ & < \theta J_\Omega(u_I^{(1)}, \pi^{(1)}) + (1 - \theta) J_\Omega(u_I^{(2)}, \pi^{(2)}). \end{aligned}$$

Thus J_Ω is strictly convex on the convex feasible set $\mathcal{C}_u(\Omega) \times \mathcal{C}_\pi(\Omega)$. A strictly convex function on a convex set has at most one minimiser. Combined with existence (Proposition 20), this implies that the minimiser of J_Ω is unique. $\square$

Remark 26 (Non-quadratic but convex structure). The patch energy J_Ω is convex but not quadratic in the joint variables (u_I, π) . The non-quadratic structure arises from the perspective-type terms $d_q(u)^2/(\varepsilon + M_q(\pi))$ in the noise-aware band penalties ϕ_q , which couple the surface misfit and the fog mass in a nonlinear way. Introducing additional slack variables to eliminate the perspective structure would break the natural probabilistic interpretation of π and $\nu_q(\pi)$, and is not pursued here. Consequently, the patch-level post-fit is formulated and solved as a general convex optimisation problem, rather than as a quadratic program.

12.9 Global post-fit across patches and dates

We now describe how the patch-level post-fit is assembled into a global arbitrage-free surface on each date, and state conditions under which global static no-arbitrage is preserved.

Throughout this section we fix a calendar date t and suppress explicit t -dependence in the notation when no ambiguity arises. All objects (quotes, bands, forwards, baseline u^0 , operators ℓ_α, r_α , etc.) are understood to be associated with this fixed date.

12.9.1 Patch decomposition and compatibility with no-arbitrage stencils

Recall that the global discrete static no-arbitrage constraints on the nodal grid $\mathcal{G}$ are encoded by the index set $\mathcal{I}$ and linear inequalities

$$\ell_\alpha^\top u \leq r_\alpha, \quad \alpha \in \mathcal{I}, \quad (12.4)$$

as in Definition 19 and equation (12.2). For each $\alpha \in \mathcal{I}$, the *support* of the stencil is

$$\text{supp}(\ell_\alpha) := \{g \in \{1, \dots, G\} : (\ell_\alpha)_g \neq 0\}.$$

Equivalently, $\text{supp}(\ell_\alpha)$ is the set of nodal indices at which u enters the α -th constraint with nonzero coefficient.

Let $\{\Omega_p\}_{p \in \mathcal{P}}$ be a finite family of pairwise disjoint patches in $\mathcal{G}$, i.e.

$$\Omega_p \subset \mathcal{G}, \quad \Omega_p \cap \Omega_{p'} = \emptyset \text{ for } p \neq p'.$$

Define their union and complement by

$$\Omega_{\text{all}} := \bigcup_{p \in \mathcal{P}} \Omega_p, \quad \Omega_{\text{off}} := \mathcal{G} \setminus \Omega_{\text{all}}.$$

We explicitly assume that the patch decomposition is compatible with the global no-arbitrage stencils in the following sense.

Assumption 3 (Stencil compatibility of the patch decomposition). *For every $\alpha \in \mathcal{I}$, the support of ℓ_α is either contained entirely in one patch or entirely outside all patches; that is, for each $\alpha \in \mathcal{I}$ there exists either:*

- *a patch index $p \in \mathcal{P}$ such that $\text{supp}(\ell_\alpha) \subset \Omega_p$, or*
- *no patch index with this property, in which case $\text{supp}(\ell_\alpha) \subset \Omega_{\text{off}}$.*

Equivalently, there is no $\alpha \in \mathcal{I}$ such that $\text{supp}(\ell_\alpha)$ intersects both Ω_p and $\mathcal{G} \setminus \Omega_p$ for some p .

In words, no-arbitrage stencils do not “straddle” patch boundaries: each discrete bound, monotonicity, convexity, or calendar constraint is supported either entirely on a single patch, or entirely outside the union of patches. This is a slightly stronger version of the patch feasibility condition discussed after Definition 22, and is natural in view of the local construction of patches from the badness field.

12.9.2 Global post-fit surface on a fixed date

For a fixed date t , the patch-level post-fit yields, for each patch Ω_p , a pair $(u_{I,p}^*, \pi_p^*)$ solving the patch problem (Definition 43) on that patch, i.e.

$$(u_{I,p}^*, \pi_p^*) \in \arg \min \{ J_{\Omega_p}(u_I, \pi) : u_I \in \mathcal{C}_u(\Omega_p), \pi \in \mathcal{C}_\pi(\Omega_p) \}.$$

By construction, $u_{I,p}^* \in \mathcal{C}_u(\Omega_p)$, so the assembled surface $u^{(p)} := u(u_{I,p}^*)$ (defined as in Section 12.4) satisfies all global no-arbitrage inequalities (12.4) with off-patch nodes fixed to their baseline values.

We now combine all patch-level interior solutions into a single global nodal surface u^* for date t .

Definition 44 (Global post-fit surface on a date). Let $u^0 \in \mathbb{R}^G$ be the baseline nodal surface for date t , and let $\{\Omega_p\}_{p \in \mathcal{P}}$ be a stencil-compatible patch decomposition (Assumption 3) with corresponding interior solutions $\{u_{I,p}^*\}_{p \in \mathcal{P}}$. The *global post-fit nodal surface* $u^* \in \mathbb{R}^G$ for date t is defined componentwise by

$$u_{i,j}^* := \begin{cases} (u_{I,p}^*)_{i,j}, & \text{if } (i,j) \in \Omega_p \text{ for some } p \in \mathcal{P}, \\ u_{i,j}^0, & \text{if } (i,j) \in \Omega_{\text{off}}. \end{cases}$$

Equivalently, u^* coincides with the patch-level interior solutions on each Ω_p and with the baseline on all nodes outside the union of patches.

We emphasise that the fog fields π_p^* remain patch-local; they are not assembled into a single global fog, since only u^* is used in further pricing and calibration.

12.9.3 Global static no-arbitrage and locality

We now show that under Assumption 3, the global post-fit surface u^* is statically no-arbitrage on $\mathcal{G}$, and that nodes outside the patches are unchanged.

Proposition 22 (Global static no-arbitrage and locality). *Fix a date t and suppose:*

- (i) *the patch decomposition $\{\Omega_p\}_{p \in \mathcal{P}}$ satisfies Assumption 3;*
- (ii) *each patch Ω_p is feasible in the sense of Definition 22 and admits a patch-level solution $(u_{I,p}^*, \pi_p^*)$ as in Definition 43;*
- (iii) *the baseline nodal surface u^0 is globally statically no-arbitrage, i.e. $u^0 \in \mathcal{C}_{\text{glob}}$.*

Let u^* be the global post-fit surface defined in Definition 44. Then:

1. u^* is statically no-arbitrage on $\mathcal{G}$, i.e. $u^* \in \mathcal{C}_{\text{glob}}$; and
2. locality holds: $u_{i,j}^* = u_{i,j}^0$ for all $(i,j) \in \Omega_{\text{off}}$.

Proof. Part (b) (locality) is immediate from the definition of u^* : by Definition 44, for $(i,j) \in \Omega_{\text{off}}$ we set $u_{i,j}^* := u_{i,j}^0$. Hence u^* agrees with the baseline on all off-patch nodes.

We now prove (a). It suffices to show that all global no-arbitrage inequalities (12.4) hold for u^* ; that is, we must verify

$$\ell_\alpha^\top u^* \leq r_\alpha, \quad \forall \alpha \in \mathcal{I}.$$

Fix $\alpha \in \mathcal{I}$ and consider the support $\text{supp}(\ell_\alpha)$. By Assumption 3, there are two mutually exclusive cases:

Case 1: $\text{supp}(\ell_\alpha) \subset \Omega_{\text{off}}$.

In this case, the α -th inequality involves only off-patch nodes. On Ω_{off} , we have $u_{i,j}^* = u_{i,j}^0$, so the α -th constraint evaluated at u^* is identical to that evaluated at u^0 :

$$\ell_\alpha^\top u^* = \ell_\alpha^\top u^0.$$

By assumption (iii), $u^0 \in \mathcal{C}_{\text{glob}}$, so $\ell_\alpha^\top u^0 \leq r_\alpha$. Hence $\ell_\alpha^\top u^* \leq r_\alpha$ in Case 1.

Case 2: There exists $p \in \mathcal{P}$ such that $\text{supp}(\ell_\alpha) \subset \Omega_p$.

In this case, the α -th constraint involves only nodes inside the single patch Ω_p . Let $u^{(p)}$ denote the full nodal surface corresponding to the patch-level interior solution $u_{I,p}^*$, i.e. the assembled surface obtained by replacing u^0 by $u_{I,p}^*$ on Ω_p and keeping all other nodes at their baseline values. By definition of $\mathcal{C}_u(\Omega_p)$ (Definition 21), we have $u_{I,p}^* \in \mathcal{C}_u(\Omega_p)$, hence

$$\ell_\alpha^\top u^{(p)} \leq r_\alpha, \quad \forall \alpha \in \mathcal{I}.$$

In particular, for the specific index α under consideration,

$$\ell_\alpha^\top u^{(p)} \leq r_\alpha.$$

We now compare $u^{(p)}$ and u^* on the support of ℓ_α . On Ω_p , both $u^{(p)}$ and u^* take the same nodal values, namely $(u_{I,p}^*)_{i,j}$; on $\mathcal{G} \setminus \Omega_p$, the α -th constraint has zero coefficients (since $\text{supp}(\ell_\alpha) \subset \Omega_p$). Therefore

$$\ell_\alpha^\top u^* = \ell_\alpha^\top u^{(p)}.$$

Hence

$$\ell_\alpha^\top u^* = \ell_\alpha^\top u^{(p)} \leq r_\alpha.$$

In both cases we have shown $\ell_\alpha^\top u^* \leq r_\alpha$. Since $\alpha \in \mathcal{I}$ was arbitrary, it follows that $\ell_\alpha^\top u^* \leq r_\alpha$ for all $\alpha \in \mathcal{I}$, i.e. $u^* \in \mathcal{C}_{\text{glob}}$. This proves (a). $\square$

Remark 27 (Independence across dates). The above argument is purely cross-sectional and is applied separately on each date t . There is no coupling in the static no-arbitrage constraints between different dates, so the global post-fit surfaces $\{u_t^*\}_t$ across all dates are obtained by applying the per-date patch decomposition and assembly independently. Provided that the assumptions of Proposition 22 hold for each date, the family $\{u_t^*\}_t$ is statically no-arbitrage on every date, and coincides with the baseline surfaces outside the union of patches on each date.

13. CONCLUSION AND OUTLOOK

We have presented a convex-programming framework for constructing arbitrage-free option price surfaces based on a global Chebyshev representation on a warped log-moneyness domain. By encoding static no-arbitrage inequalities as linear constraints on a dense collocation grid, and fitting directly to prices via a coverage-seeking quadratic objective, the method yields a surface that is both smooth and internally consistent.

On the empirical side, our implementation attains high inside-spread coverage and low rates of static no-arbitrage violations across a multi-year panel of equity options. These results suggest that Chebyshev/QP formulations, combined with spectral-geometry and transport-type regularisers, are a viable and competitive alternative to more widely used parametric and spline-based approaches, particularly when tight control over arbitrage metrics is required.

Beyond the global QP backbone, we have formulated a local post-fit layer in which a discrete fog of risk-neutral densities on (m, τ, u) is endowed with a Hamiltonian-type energy. On each problematic patch of the (m, τ) -plane, this fog is coupled convexly to a nodal price field that remains globally arbitrage-free. The resulting patch problems are jointly convex in the surface and fog variables and yield noise-aware corrections that improve local band coverage in stressed regions while preserving static no-arbitrage and locality.

Several limitations and directions for further work remain. First, our study focuses on a particular choice of warping, regularisation and grid design; different markets or underlyings may benefit from alternative configurations, and a more systematic comparison against SVI-type and deep-learning-based surfaces would be informative. This will be tackled in a separate paper. Second, we have evaluated performance primarily through static diagnostics (spread coverage, violation rates, smoothness) and local band metrics on patches. A natural next step is to examine the impact on hedging performance and risk measures, for example via delta-hedging backtests or scenario analysis of risk-neutral densities, both for the baseline QP and for the fog-corrected surface.

Finally, the Hamiltonian fog layer is implemented here in a finite-dimensional, patch-wise discretisation. From a mathematical standpoint, it suggests a continuous framework in which a fog density $\pi(m, \tau, u)$ on a three-dimensional manifold evolves under a Hamiltonian or transport-type metric, with the option surface appearing as a constrained “sheet” inside this geometry. Developing this continuous theory—including PDE and variational formulations, existence and uniqueness questions, and connections to optimal transport on the space of risk-neutral measures is beyond the scope of the present paper and will be pursued in separate work. We view the discrete constructions in this article as a practical, convex realisation of that program: the global Chebyshev/QP fit provides a transparent arbitrage-free backbone, and the patch-wise Hamiltonian fog post-fit offers a local, noise-aware refinement that remains compatible with production-style constraints and solvers.

Acknowledgments

The author made limited use of an AI language model (ChatGPT by OpenAI) as a writing and brainstorming aid; all models, proofs and numerical results presented are the author's own work and have been independently verified.

BIBLIOGRAPHY

- [1] Y. Achdou and O. Pironneau. Volatility smile by multilevel least square. *International Journal of Theoretical and Applied Finance*, 5, 2002.
- [2] J. Andreasen and B. N. Høge. Volatility interpolation. *Risk*, 24, 2010.
- [3] J. Andreasen and B. H. D. Markov. Volatility interpolation. *SSRN*, 2011.
- [4] A. Antonov, M. Konikov, and M. Spector. A new arbitrage-free parametric volatility surface. *SSRN Electronic Journal*, 2019.
- [5] M. Avellaneda, C. A. Friedman, R. Holmes, and D. J. Samperi. Calibrating volatility surfaces via relative-entropy minimization. *Taylor and Francis Journals*, 4, 1997.
- [6] Y. Aït-Sahalia and J. Duarte. Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116, 2003.
- [7] J. Bodurtha and M. Jermakyan. Nonparametric estimation of an implied volatility surface. *The Journal of Computational Finance*, 2, 1999.
- [8] S. Boyd and L. Vandenberghe. Convex optimization. *Cambridge University Press*, 2004.
- [9] D. T. Breeden and R. H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51, 1978.
- [10] S. Crépey. Calibration of the local volatility in a generalized black–scholes model using tikhonov regularization. *SIAM Journal on Mathematical Analysis*, 34, 2003.
- [11] E. Derman and I. Kani. Riding on a smile. *Risk*, 7, 1994.
- [12] F. Deschâtres. Convex volatility interpolation. *SSRN Electronic Journal*, 2024.
- [13] B. Dupire. Pricing with a smile. *Risk*, 7, 1994.
- [14] R. L. Dykstra. An iterative procedure for obtaining i-projections onto the intersection of convex sets. *Institute of Mathematical Statistics*, 1985.
- [15] M. Fengler. Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance*, 9, 2009.
- [16] F. L. Floc'h and C. W. Oosterlee. Model-free stochastic collocation for an arbitrage-free implied volatility (part ii). *Risks*, 7, 2019.
- [17] J. Gatheral. The volatility surface: A practitioner’s guide. *Wiley Finance*, 2006.
- [18] J. Gatheral and A. J. Jacquier. Arbitrage-free svi volatility surfaces. *Quantitative Finance*, 14, 2012.
- [19] P. Hagan, D. Kumar, , A. Lesniewski, and D. Woodward. Managing smile risk. *Wilmott Magazine*, 1, 2002.

- [20] N. Kahale. An arbitrage-free interpolation of volatilities. *Risk*, 17, 2004.
- [21] R. Lagnado and S. Osher. A technique for calibrating derivative security pricing models: numerical solution of an inverse problem. *The Journal of Computational Finance*, 1, 1997.
- [22] H. D. March and P. Henry-Labordere. Building arbitrage-free implied volatility: Sinkhorn’s algorithm and variants. *ArXiv*, 2023.
- [23] J. Mason and D. C. Handscomb. Chebyshev polynomials. *Chapman and Hall/CRC*, 2002.
- [24] M. Rubenstein. Implied binomial trees. *Journal of Finance*, 49, 1994.
- [25] H. Yin, L. Qi, and Y. Wang. Shape-preserving interpolation and smoothing for options market implied volatility. *Journal of Optimization Theory and Applications*, 142, 2009.
- [26] V. Zetocha. Volatility transformers: an optimal transport-inspired approach to arbitrage-free shaping of implied volatility surfaces. *SSRN Electronic Journal*, 2023.