

Correlation Is a Hyperbolic Distance

Differential geometry in volatility models and covariance estimation:
exact identities, measured limits, and three refuted hypotheses

Egor S. Galkin

July 25, 2026

Abstract

We show that the correlation parameter in two independent constructions of mathematical finance — the SABR stochastic-volatility model and the Riemannian geometry of covariance matrices — is one and the same object: a hyperbolic distance. For SABR we prove that the strike line $\{x = x_K\}$ is a *hypercycle* (equidistant curve) of a vertical geodesic at constant distance $|\operatorname{artanh} \rho|/\nu$; this yields the exact decomposition of Hagan’s function, $x(z) = \operatorname{arcsinh}((z - \rho)/\sqrt{1 - \rho^2}) + \operatorname{artanh} \rho$, which means that the entire dependence of the volatility smile on correlation is an isometry of the hyperbolic plane. The same law $\rho = \tanh(\delta)$ is found on the manifold $\operatorname{SPD}(2) \cong \mathbb{R}_+ \times \mathbb{H}^2$ (verified on 3000 random pairs to $2.4\text{e-}13$), where the distance is constant to a standard deviation of $2.65\text{e-}16$. For the family $d\sigma = \nu\sigma^p dW_2$ we derive the closed-form Gauss curvature $K = \nu^2(p - 2)\sigma^{2(p-1)}$, from which SABR is the unique case of nonzero constant curvature. Alongside these, three a-priori hypotheses are refuted by measurement: geodesic averaging of covariances *under-reports* realised risk by 26–37%; Levenberg–Marquardt converges in 2 iterations against 43 for the natural gradient, whose real advantage turns out to be only exact invariance under reparametrisation; and constant curvature in the above family is attained not only at $p = 1$. Every symbolic statement has an identically vanishing residual; every numerical one carries its measured accuracy. Code and data are included.

Contents

1	Introduction	3
2	The SABR metric and its curvature	3
3	Correlation as a distance: the hypercycle theorem	4
3.1	The strike line is a hypercycle	4
3.2	Three consequences	4
3.3	Decomposition of Hagan’s function	5
3.4	The prefactor as an arc length	6
4	The gauge structure of the parameters	6
4.1	Where the invariance is observable and where it is hidden	6
5	Covariance matrices on the same plane	7
5.1	The isometry $\operatorname{SPD}(2) \cong \mathbb{R}_+ \times \mathbb{H}^2$	7
5.2	$\rho = \tanh(\text{distance})$ again	7
5.3	Geodesic averaging: a refuted hypothesis	8
5.4	Geodesic interpolation of stress scenarios	8

6	Information geometry of calibration	9
6.1	Two metrics, one answer	9
6.2	The β - ρ degeneracy as a geometric fact	9
6.3	The natural gradient: a refuted hypothesis	10
7	A curvature taxonomy of stochastic-volatility models	10
7.1	Closed form	10
7.2	The curvature of Heston	10
7.3	Numerical geodesics and the price of the approximation	12
8	Validation of the SABR asymptotics	12
9	What of this is a working tool	14
10	Status relative to the literature	14
	Reproducibility	14

1 Introduction

Differential geometry enters mathematical finance not as descriptive language but as a derivation mechanism. If the state process is given by a stochastic differential equation with diffusion matrix $a(\theta)$, then $g = a^{-1}$ defines a Riemannian metric, and Varadhan's asymptotics [15] relate the transition probability to geodesic distance:

$$\lim_{t \rightarrow 0} t \log p_t(\theta_0, \theta) = -\frac{1}{2} d_g(\theta_0, \theta)^2. \quad (1)$$

For volatility models this gives the smile in terms of the distance to a strike line [4, 10]; for covariance matrices it gives invariant means and interpolations [13, 14, 3].

That the SABR model is connected to hyperbolic geometry has been known since [11, 9]. The present text claims priority for none of its statements (Section 10); its contribution lies in three things: (i) the exact geometric identification of the *correlation parameter* as a distance, with a complete symbolic proof; (ii) measured limits of validity for each construction, including three cases where the expected result proved wrong; and (iii) a reproducible code base with a known error budget.

Status convention. Every statement below is labelled: **(exact)** — proved symbolically, computer-algebra residual identically zero; **(measured)** — a numerical result with stated accuracy; **(known)** — present in the literature, reproduced and verified here.

2 The SABR metric and its curvature

Under the forward measure the SABR model reads

$$dF_t = \sigma_t F_t^\beta dW_t^1, \quad d\sigma_t = \nu \sigma_t dW_t^2, \quad d\langle W^1, W^2 \rangle_t = \rho dt. \quad (2)$$

The diffusion matrix in (F, σ) coordinates and its inverse give the metric. Passing to the coordinate that flattens the CEV nonlinearity,

$$x = \frac{F^{1-\beta}}{1-\beta} \quad (\log F \text{ when } \beta = 1), \quad (3)$$

removes the explicit F -dependence, and the subsequent whitening change of variables

$$\xi = \frac{x - \rho\sigma/\nu}{\sqrt{1-\rho^2}}, \quad \eta = \frac{\sigma}{\nu} \quad (4)$$

gives **(exact)**

$$\boxed{ds^2 = \frac{d\xi^2 + d\eta^2}{\nu^2 \eta^2}} \quad (5)$$

— the Poincaré upper half-plane metric scaled by $1/\nu$. The difference between g and (5) is the zero matrix identically in all parameters.

Direct computation of the Christoffel symbols gives $\Gamma_{\xi\eta}^\xi = -1/\eta$, $\Gamma_{\xi\xi}^\eta = +1/\eta$, $\Gamma_{\eta\eta}^\eta = -1/\eta$, whence $\text{Ric} = -\nu^2 g$, scalar curvature $R = -2\nu^2$, and **(exact)**

$$K = -\nu^2. \quad (6)$$

The curvature of the state space equals the squared volatility of volatility. Geodesics are semicircles centred on $\eta = 0$ together with vertical rays; for the unit-speed semicircle $\xi(t) = c + r \tanh(\nu t)$, $\eta(t) = r \text{sech}(\nu t)$ the geodesic-equation residual vanishes identically and $|\dot{\gamma}|_g \equiv 1$.

3 Correlation as a distance: the hypercycle theorem

3.1 The strike line is a hypercycle

An option struck at K corresponds not to a point but to a *line* in state space — the set $L_K = \{(x, \sigma) : x = x_K\}$ — because the terminal volatility is not fixed. In coordinates (4) this line is a Euclidean straight line (**exact**),

$$\xi(\eta)|_{L_K} = \frac{x_K - \rho\eta}{\sqrt{1 - \rho^2}}, \quad (7)$$

inclined to the vertical at an angle θ satisfying (**exact**)

$$\cos \theta = \sqrt{1 - \rho^2}, \quad \sin \theta = |\rho|. \quad (8)$$

Correlation is the sine of the angle between the strike line and the vertical geodesic.

Lemma 1 (distance to a vertical geodesic). *For the geodesic $\{\xi = c\}$ and a point (ξ, η) , $\eta > 0$, minimising $\cosh d(t) = 1 + [(\xi - c)^2 + (\eta - t)^2]/(2\eta t)$ over the foot parameter $t > 0$ gives $t^* = \sqrt{(\xi - c)^2 + \eta^2}$ and, after simplification, (**exact**)*

$$\sinh d_{\min} = \frac{|\xi - c|}{\eta}. \quad (9)$$

Theorem 2 (hypercycle theorem). *The strike line L_K is a hypercycle (equidistant curve) of the vertical geodesic $\{\xi = x_K/\sqrt{1 - \rho^2}\}$ passing through the same boundary point. The constant distance is*

$$\boxed{\delta = \frac{|\operatorname{artanh} \rho|}{\nu} \iff \rho = \tanh(\nu\delta)} \quad (10)$$

Proof. Substitute $c = x_K/\sqrt{1 - \rho^2}$ and (7) into (9):

$$u(\eta) = \frac{\xi(\eta) - c}{\eta} = -\frac{\rho}{\sqrt{1 - \rho^2}}, \quad (11)$$

and the η -dependence cancels identically (**exact**) — which is precisely the definition of a hypercycle. Further, $\sinh(\operatorname{artanh} \rho) = \rho/\sqrt{1 - \rho^2}$ holds algebraically, whence $\delta = |\operatorname{artanh} \rho|$ in the unit-curvature convention and $\delta = |\operatorname{artanh} \rho|/\nu$ in the metric (5). $\square$

Numerically, the spread of δ along a strike line as η varies over three orders of magnitude is 1.6e−15 (**measured**); in an independent check with η varying by a factor of 140 it is 4.4e−16.

3.2 Three consequences

Theorem 2 immediately explains three properties of the model that are usually described separately.

1. $\rho = 0 \iff \delta = 0$: the strike line *coincides* with the geodesic. The smile is symmetric because the geometry is symmetric.
2. The sign of ρ is the side of the geodesic on which the strike line lies. The smile skew is not an effect but a choice of half-plane.
3. $\rho \rightarrow \pm 1 \iff \delta \rightarrow \infty$: the strike line recedes to infinite distance, degenerating onto the boundary. The known numerical instability of SABR at strong correlation is a degeneration of the geometry, not an implementation defect.

The SABR correlation ρ is not a parameter but a distance

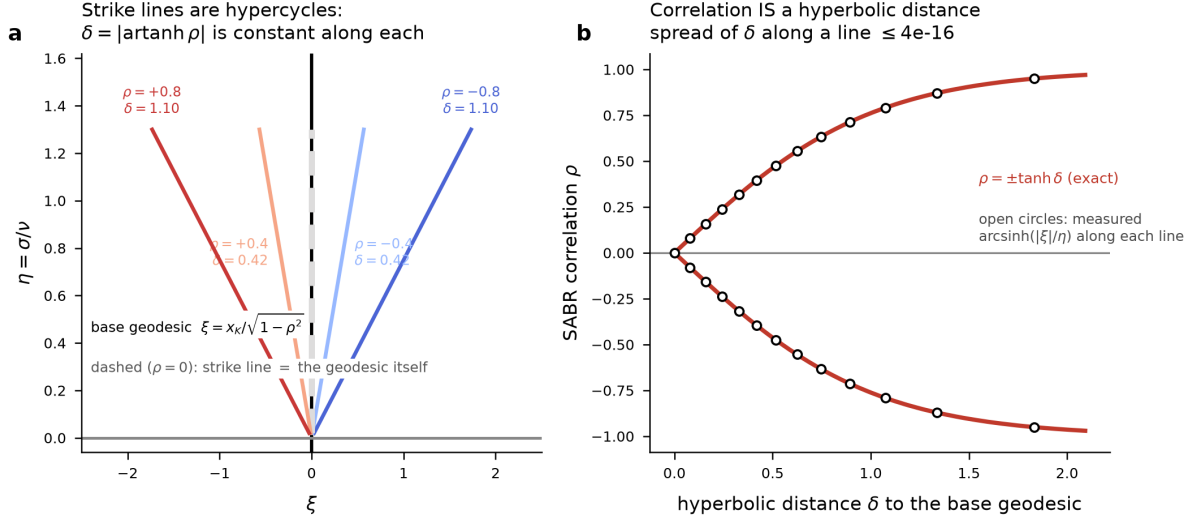


Figure 1: Correlation as a distance. **(a)** Strike lines at various ρ are hypercycles of one and the same vertical geodesic; the distance $\delta = |\text{artanh } \rho|$ is constant along each. Dashed ($\rho = 0$): the strike line coincides with the geodesic. **(b)** Measured distance $\text{arcsinh}(|\xi|/\eta)$ against the exact $\rho = \pm \tanh \delta$; the spread of δ along a line is $\leq 4e-16$.

3.3 Decomposition of Hagan's function

The distance from the initial state (x_F, α) to the strike line follows from minimising over the terminal volatility; the stationarity condition gives **(exact)**

$$\sigma^* = \sqrt{\alpha^2 + 2\rho\alpha\nu u + \nu^2 u^2}, \quad u = x_K - x_F, \quad (12)$$

so the well-known radical $\sqrt{1 - 2\rho z + z^2}$ of Hagan's formula [8] is simply σ^*/α — the optimal terminal volatility. Substituting back gives $\nu d = |x_H(z)|$ with

$$x_H(z) = \log \frac{\sqrt{1 - 2\rho z + z^2} - \rho + z}{1 - \rho}, \quad z = \frac{\nu}{\alpha} (x_F - x_K), \quad (13)$$

so Hagan's $x(z)$ is *identically equal* to the hyperbolic distance **(exact)**, numerically to $3.0e-13$ **(measured)**.

Proposition 3 (isometry plus shift).

$$x_H(z) = \text{arcsinh} \frac{z - \rho}{\sqrt{1 - \rho^2}} + \text{artanh } \rho \quad (14)$$

Proof. The derivatives agree: both equal $1/\sqrt{1 - 2\rho z + z^2}$, which follows from the exact radicand identity $(1 - \rho^2) + (z - \rho)^2 = 1 - 2\rho z + z^2$ **(exact)**. The values at $z = 0$ agree (both vanish) by virtue of the auxiliary identity $\text{arcsinh}(\rho/\sqrt{1 - \rho^2}) = \text{artanh } \rho$. Equal derivatives everywhere plus equality at one point give identity of the functions. $\square$

A numerical check on the grid $z \in [-8, 8] \times \rho \in [-0.95, 0.95]$ gives a residual of $2.1e-13$ **(measured)**. The geometric reading: the first term is the $\rho = 0$ distance in a shifted frame, the second is a pure hypercyclic offset. *The entire dependence of the SABR smile on correlation is an isometry of $\mathbb{H}^2$, not new geometry.*

Table 1: The role of the SABR parameters in the geometry.

Parameter	Role	Reason
ν	geometric	$K = -\nu^2$ is a metric invariant
ρ	geometric	fixes the hypercyclic offset $\operatorname{artanh} \rho$
α, x_F, x_K	gauge	two-parameter affine subgroup of $\operatorname{PSL}(2, \mathbb{R})$
β	chart choice	(3) is a coordinate map

3.4 The prefactor as an arc length

The factor $(FK)^{(1-\beta)/2}$ and the series $1 + \omega^2 L^2/24 + \omega^4 L^4/1920 + \dots$ in the denominator of Hagan’s formula ($\omega = 1 - \beta$, $L = \log(F/K)$) also carry an exact geometric meaning (**exact**):

$$u = (FK)^{\omega/2} \frac{\sinh(\omega L/2)}{\omega/2}, \quad (15)$$

that is, this is the *exact CEV arc length* rewritten through the geometric mean; the coefficients $1/24$ and $1/1920$ are just the sinh series. The numerical residual of the identity is $1.1\text{e}-15$ (**measured**).

Remark 4 (error source at $\beta < 1$). Decomposing Hagan’s formula into independent approximations shows that truncating the series (15) costs less than 0.1 bp for $\beta \geq 0.3$, whereas substituting the same truncated u into the *argument* z of x_H accounts for 86 of the 118 bp discrepancy at $\beta = 0.3$ (**measured**). Replacing it by the exact geodesic distance removes precisely this, the dominant component. This correction has been described independently in the literature [6].

4 The gauge structure of the parameters

The hypercycle structure raises a natural question: which parameters change the geometry, and which merely move points on a fixed geometry?

Proposition 5 (gauge group). *The two-parameter affine (Borel) subgroup of $\operatorname{PSL}(2, \mathbb{R})$,*

$$M(A, C) = \begin{pmatrix} \sqrt{A} & C/\sqrt{A} \\ 0 & 1/\sqrt{A} \end{pmatrix}, \quad \det M = 1, \quad A > 0, \quad C \in \mathbb{R}, \quad (16)$$

acting as $x_F \mapsto Ax_F + C$, $x_K \mapsto Ax_K + C$ (for every strike), $\alpha \mapsto A\alpha$, with ν, ρ, β held fixed, leaves all geodesic distances $d(K)$ invariant.

That every element of $\operatorname{PSL}(2, \mathbb{R})$ is an isometry of the metric (5) is proved from scratch: $\operatorname{Im} w = \eta/|cz + d|^2$ and $|dw/dz|^2 = 1/|cz + d|^4$, so the pullback ratio $ds'^2/ds^2 = 1$ identically (**exact**). Distance invariance is verified directly: for $A = 1.7$, $C = 0.35$, $\beta = 0.6$, $\nu = 0.35$, $\rho = -0.4$ the maximum $|d(K) - d'(K)|$ over a 46-strike grid is $8.9\text{e}-15$ (**measured**); for instance $(F, \alpha) = (100, 0.25)$ and $(389.226, 0.425)$ give identical $d(K)$ at every strike.

The converse is verified too: no gauge element compensates a change of ρ from -0.4 to 0.1 at more than one strike simultaneously (best achievable residual 1.14, against $5\text{e}-15$ for a genuine gauge element) (**measured**).

4.1 Where the invariance is observable and where it is hidden

Here the a-priori expectation proved *wrong*. The invariance of $d(K)$ is in general a *hidden* symmetry, not an observable symmetry of the smile, because $\sigma_{\text{BS}} = |\log(F/K)|/d(K)$ contains

Table 2: Smile discrepancy under gauge-equivalent parameters; all $d(K)$ are invariant to $1.5\text{e-}14$. Grid of 46 strikes, $K \in [60, 150]$ (**measured**).

Transformation	β	Smile discrepancy
general gauge $A = 1.7, C = 0.35$	0.6	up to 13.7 bp
translation only $A = 1, C = 0.7$	0.6	up to 44.9 bp
translation only $A = 1, C = 0.7$	1.0	exactly 0 ($1.1\text{e-}10$ bp)
general gauge $A = 1.7, C = 0.7$	1.0	up to 2075.5 bp
dilation only $C = 0, A = 2.2$	0.6	exactly 0 ($5.8\text{e-}12$ bp)

the non-invariant factor $|\log(F/K)|$, and the strike K itself is a market observable that an isometry does not preserve.

Two exact observable sub-symmetries nonetheless exist (Table 2): pure dilation at any β — since $F_{\text{of}}(Ax, \beta) = A^{1/(1-\beta)} F_{\text{of}}(x, \beta)$, i.e. it is a common rescaling of F and K that preserves $\log(F/K)$ — and pure translation at $\beta = 1$, where $x = \log F$ and the shift is $F \mapsto e^C F$. A control check: the basis-point discrepancy computed directly agrees with the one reconstructed from the $|\log(F/K)|$ mismatch alone to $3.9\text{e-}12$ — the entire residual is attributable to that factor and nothing else (**measured**).

5 Covariance matrices on the same plane

5.1 The isometry $\text{SPD}(2) \cong \mathbb{R}_+ \times \mathbb{H}^2$

The manifold of symmetric positive-definite 2×2 matrices with the affine-invariant metric $\langle A, B \rangle_S = \text{tr}(S^{-1}AS^{-1}B)$ splits into scale and shape (**known**); the explicit map is (**exact**)

$$r = \frac{\log \det S}{\sqrt{2}}, \quad \xi = \rho \frac{\sigma_1}{\sigma_2}, \quad \eta = \sqrt{1 - \rho^2} \frac{\sigma_1}{\sigma_2}, \quad (17)$$

$$ds_{\text{AI}}^2 = dr^2 + 2 \frac{d\xi^2 + d\eta^2}{\eta^2}. \quad (18)$$

The curvature of the shape factor is exactly $K = -1$ in this normalisation (the frequently quoted $-1/2$ corresponds to the $\frac{1}{2} \text{tr}(\cdot)$ normalisation; we verified the value our own convention produces). The distance decomposes as

$$d_{\text{AI}}(S_1, S_2) = \sqrt{(r_1 - r_2)^2 + 2 d_{\mathbb{H}^2}^2}, \quad (19)$$

verified against $\|\log(S_1^{-1/2}S_2S_1^{-1/2})\|_F$ on 3000 random pairs: maximum absolute error $2.40\text{e-}13$, relative error $2.91\text{e-}14$ (**measured**).

5.2 $\rho = \tanh(\text{distance})$ again

The hypercycle structure of Section 3 recurs here verbatim. Fixing the volatility ratio $k = \sigma_1/\sigma_2$ and varying ρ traces the geodesic semicircle $\xi^2 + \eta^2 = k^2$; fixing ρ and varying k traces a straight ray from the origin at angle $\arcsin \rho$. The hyperbolic distance from that ray to the geodesic $\{\rho = 0\}$ is constant to a standard deviation of $2.65\text{e-}16$ as k varies by a factor of 20, and equals $\text{artanh } \rho$: for $\rho = 0.9$ the mean distance is 1.472219 and $\tanh(1.472219) = 0.900000$ (**measured**).

This is not a coincidence of formulas. Both metrics are inverse-diffusion (Fisher-information type) metrics on a two-parameter correlated-noise system, and $\rho = \tanh d$ is the rapidity addition law of the Möbius model. *Correlation is a hyperbolic distance wherever it enters a metric of this type.*

Table 3: Self-reported versus realised risk. Two regimes (calm: correlation 0.2, low volatility; crisis: 0.8, high volatility), 20 overlapping 63-day windows (**measured**).

System	Scheme	Reported	Realised	Error
2 assets	Euclidean	27.14 %	26.80 %	−1.3 %
2 assets	log-Euclidean	21.35 %	26.89 %	+25.9 %
2 assets	Karcher	21.43 %	26.93 %	+25.7 %
5 assets	Euclidean	21.67 %	21.71 %	+0.1 %
5 assets	log-Euclidean	16.02 %	21.99 %	+37.3 %
5 assets	Karcher	16.29 %	22.16 %	+36.0 %

5.3 Geodesic averaging: a refuted hypothesis

The expectation was standard: Euclidean averaging of covariances suffers from determinant swelling, geodesic averaging does not, hence geodesic averaging is better. The first half holds (**measured**): the determinant swelling ratio (det of the mean relative to the geometric mean of the per-window determinants) is 1.696 and 2.467 for 2- and 5-asset systems under Euclidean averaging, and exactly 1.000 under log-Euclidean averaging and the Karcher barycentre.

The second half is *false*. Building the minimum-variance portfolio from each averaged matrix and comparing the risk that the scheme reports for itself against the risk realised under the true regime mixture (Table 3) gives a 26–37 % under-report for the geodesic schemes.

The reason is elementary once seen: the Karcher barycentre approximates a geometric mean of variances, while realised risk over a regime mixture is governed by the arithmetic mean, and $AM \geq GM$. The Euclidean mean, for all its swelling, is the exact first moment of the mixture the portfolio will actually experience.

Conclusion. Geodesic averaging is the right tool for smoothing and comparing covariance matrices and the *wrong* tool for unconditional risk estimation over a regime-mixing sample. Invariance is not unbiasedness.

5.4 Geodesic interpolation of stress scenarios

The common argument that “linear interpolation breaks positive definiteness” is *false*: $SPD(n)$ and the correlation ellipsope are convex, so a linear path between two admissible matrices never leaves the admissible set — verified directly (the minimum eigenvalue never dropped below 0.003 along the whole path, 5×5) (**measured**). We do not use that argument.

The real difference is speed and distortion. Both paths lie on the *same* hyperbolic line in shape space (residual $1.4e-14$ over 20 random pairs), but the geodesic traverses it at constant speed (affine-fit residual of arc length $4.6e-15$) whereas the linear path does not (residual 0.22); and log det is exactly affine along the geodesic but not along the linear path, which is the direct source of swelling. The measured distortion on a calm-to-crisis path (Table 4) reaches 27–36 % in portfolio risk and up to 108 % in the determinant, peaking mid-path — exactly where the intermediate stress scenarios live.

Remark 6. The hypothesis that the linear path yields non-monotone risk while the geodesic does not was *not* confirmed: an example was found in which both are non-monotone. Non-monotonicity is not a distinguishing feature; the distortion magnitudes are.

Table 4: Distortion of linear interpolation relative to the geodesic (**measured**).

System	Quantity	Max. distortion	at t
2 assets	$\det S(t)$	66.9 %	0.35
2 assets	min-variance portfolio volatility	27.2 %	0.325
5 assets	$\det S(t)$	108.0 %	0.325
5 assets	VaR ₉₉	36.1 %	0.275

Table 5: Parameter ranges indistinguishable from the smile within a stated tolerance. $T = 1$, 13 strikes, $\theta_0 = (0.25, 0.6, 0.5, -0.4)$ (**measured**).

Tolerance	Range of β	Compensating ρ
5 bp	0.42–0.77	−0.32 ... −0.47
10 bp	0.32–0.86	−0.28 ... −0.51
25 bp	from 0.12	to −0.19

6 Information geometry of calibration

6.1 Two metrics, one answer

On the parameter space $\theta = (\alpha, \beta, \nu, \rho)$ we build two metrics: the vega-weighted metric of smile sensitivity,

$$G_{ij} = \sum_k w_k \frac{\partial \sigma_{BS}(K_k)}{\partial \theta_i} \frac{\partial \sigma_{BS}(K_k)}{\partial \theta_j} \quad (20)$$

and the Fisher–Rao metric of the terminal distribution, obtained via Breeden–Litzenberger from the same Hagan approximation (so that the comparison rests on equal asymptotic footing). Derivatives are computed by complex-step differentiation and cross-checked against Richardson extrapolation to a maximum relative disagreement of 2.2e^{-11} (**measured**).

Both metrics single out the same soft direction: the cosine between the softest eigenvectors is $0.9944\text{--}0.99998$ over $T \in [0.1, 2]$ (**measured**). This matters in practice: the cheaply computable vega metric carries the same information as the theoretically correct Fisher metric.

Remark 7 (limit of validity). At $\nu^2 T = 1.25$ the density implied by Hagan’s formula goes negative on 761 of 4000 grid points (minimum -1.14) — the known arbitrage pathology, here measured. All conclusions of this section are restricted to $T \leq 2$.

6.2 The β – ρ degeneracy as a geometric fact

The condition number of the metric over all strike sets with ≥ 5 strikes and $T \in [0.1, 5]$ is $5.3\text{e}4\text{--}1.5\text{e}8$ (for the wide 13-strike set, $2.2\text{e}5\text{--}3.6\text{e}7$; the worst conditioning belongs to the narrow 5-strike set), and the softest eigenvector carries $98.3\text{--}98.7\%$ of its norm in the (β, ρ) plane (**measured**). At $T = 1$ with 13 strikes it equals $(\alpha, \beta, \nu, \rho) = (-0.006, -0.906, -0.158, +0.392)$ — almost pure β , compensated by ρ .

More than half the physical range of β is indistinguishable at a realistic 10 bp tolerance (Table 5). This is the honest geometric answer to whether β can be calibrated from the smile: not from a single maturity. A term structure helps but does not resolve it — adding maturities improves the condition number by roughly one order of magnitude ($3.65\text{e}5 \rightarrow 3.6\text{e}4$ for three tenors), while the (β, ρ) share of the soft direction barely moves ($96.6\text{--}97.3\%$). This is consistent with Proposition 3: ρ enters the smile only as a hypercyclic shift and adds no shape information independent of β .

6.3 The natural gradient: a refuted hypothesis

A comparison on synthetic smiles with 3 bp noise, identical starting points and identical stopping criteria (**measured**): Euclidean gradient descent takes 833 iterations, the natural gradient (preconditioned by G^{-1} with damping) takes 43, and Levenberg–Marquardt takes **2**. The natural gradient’s 19-fold speed-up over plain descent is real and traces directly to the ill-conditioning found above, but it does not close the gap to a Gauss–Newton-type method. All three land in the same degenerate neighbourhood (recovery error 0.10, 0.10, 0.04) — because the smile physically cannot distinguish those points, a property of the data rather than the optimiser.

The natural gradient’s real advantage is invariance under reparametrisation. The full iteration counts are invariant for neither method (line search and trust region are themselves chart-dependent: ratios 3.1 and 2.95). But isolating the geometric object — the descent direction computed in $(\alpha, \beta, \log \nu, \operatorname{artanh} \rho)$ and pushed back to the original coordinates by the exact chain-rule Jacobian — gives a cosine with the native direction of 0.9957 on average (minimum 0.930) for the plain gradient and 1.0000000000 (norm ratio $1 \pm 5.9\text{e-}13$) for the natural gradient (**measured**), as theory requires, since $-G^{-1}\nabla f$ transforms as a genuine tangent vector.

Verdict. The natural gradient buys invariance and conditioning, not speed. Given a working Levenberg–Marquardt it is not needed for speed; it is needed when it matters that the result does not depend on whether one calibrates ν or $\log \nu$.

7 A curvature taxonomy of stochastic-volatility models

7.1 Closed form

For the family

$$dF = \sigma F^\beta dW_1, \quad d\sigma = \nu \sigma^p dW_2, \quad d\langle W_1, W_2 \rangle = \rho dt \quad (21)$$

the Gauss curvature of the induced metric is obtained in closed form (**exact**), with an independent finite-difference check to $2\text{e-}8$ (**measured**):

$$K(\sigma; p, \nu) = \nu^2 (p - 2) \sigma^{2(p-1)} \quad (22)$$

The formula depends on neither β nor ρ : *neither the choice of CEV exponent nor the correlation affects the curvature* — it is determined solely by the volatility dynamics.

Proposition 8 (all constant-curvature cases). *K is a monomial $c\sigma^a$ with $c = \nu^2(p - 2)$, $a = 2(p - 1)$, and is constant on an open interval if and only if $a = 0$ or $c = 0$. For $\nu \neq 0$ this gives exactly two solutions: $p = 1$ (SABR, $K = -\nu^2$, the hyperbolic case) and $p = 2$ ($K = 0$, flat).*

The a-priori hypothesis that “ $p = 1$ is the unique constant-curvature case” is therefore *false as literally stated*, but true in the substantive sense: SABR is the unique case of *nonzero* constant curvature, the unique hyperbolic one. The case $p = 2$ corresponds to $d\sigma = \nu \sigma^2 dW_2$ — not a standard financial specification, and one whose geodesics degenerate to straight lines.

The practical content of Proposition 8: constant curvature is exactly what makes closed-form geodesics — and hence a closed-form leading-order smile — possible. Non-constant curvature forces numerical geodesics.

7.2 The curvature of Heston

For the Heston model

$$dS = \sqrt{v} S dW_1, \quad dv = \kappa(\theta - v) dt + \xi \sqrt{v} dW_2 \quad (23)$$

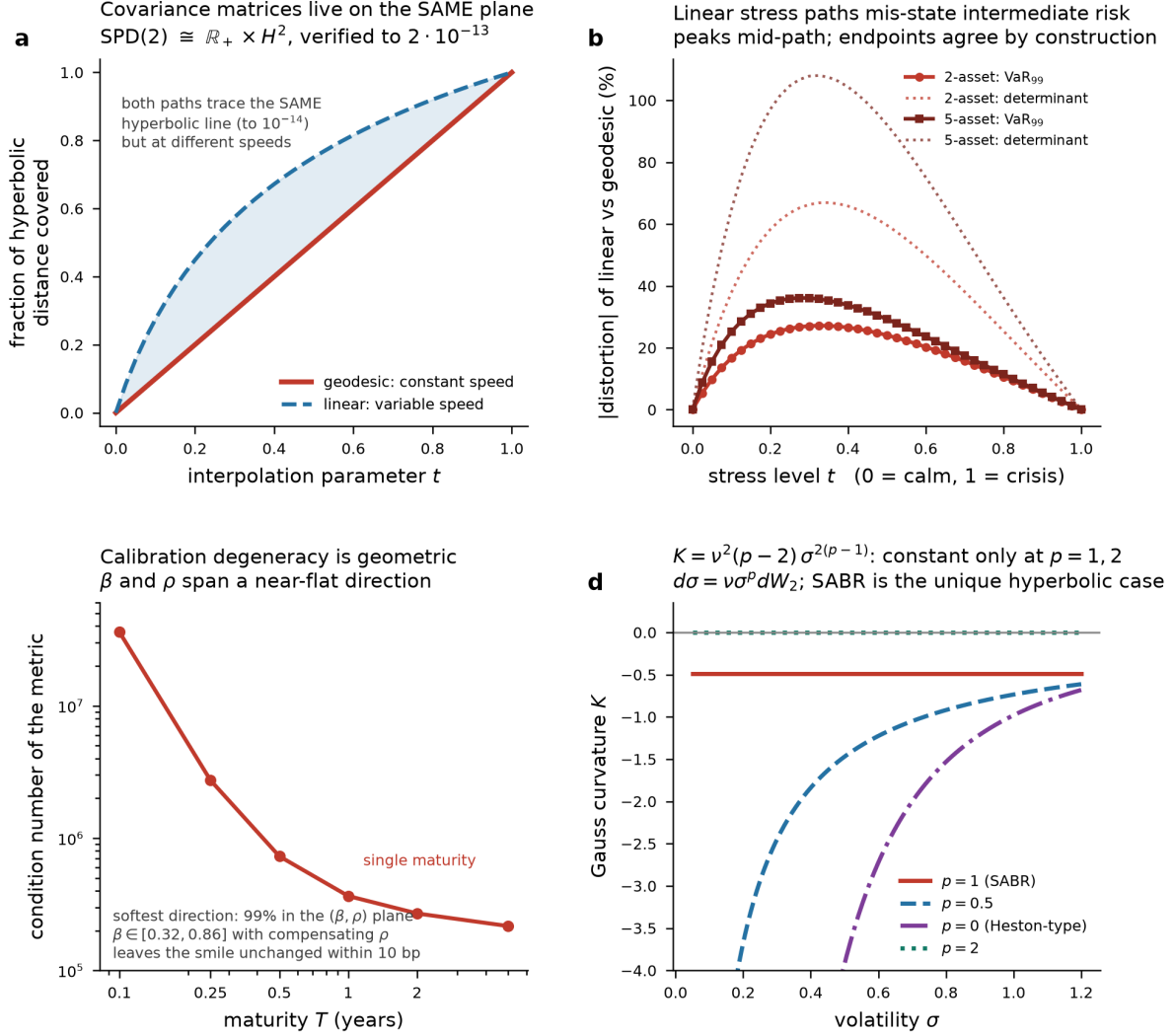


Figure 2: Four applications. **(a)** Both interpolation paths in $\text{SPD}(2)$ lie on the same hyperbolic line, but only the geodesic traverses it at constant speed. **(b)** The distortion of a linear stress path peaks mid-path. **(c)** Condition number of the calibration metric by maturity; the soft direction is almost the pure (β, ρ) plane. **(d)** Curvature of the family $d\sigma = v\sigma^p dW_2$: constant only at $p=1$ and $p=2$.

the drift does not enter the metric (only the diffusion does), and the curvature is **(exact)**

$$K_{\text{Heston}}(v) = -\frac{\xi^2}{2v}. \quad (24)$$

It depends only on the vol-of-vol ξ and the current variance v — *not* on S , κ , θ or ρ (ρ enters the metric tensor but cancels exactly out of the scalar curvature). Consistency with (22) is verified: $\sqrt{v}$ has constant Itô diffusion coefficient $\xi/2$ (the classical CIR fact), so Heston is the $p=0$ branch with $\nu_{\text{eff}} = \xi/2$.

The curvature is constant for no parameter values other than $\xi=0$. For realistic parameters ($v_0 = 0.04$, $\kappa = 2$, $\theta = 0.04$, $\xi = 0.3$, Feller ratio 1.78) the exact 1–99% range of v over one year is $[0.0023, 0.138]$, giving $K \in [-19.57, -0.326]$ with $K(v_0) = -1.125$ (**measured**): a 60-fold spread (the ratio equals $v_{\text{hi}}/v_{\text{lo}} = 0.138/0.0023 = 60.0$ exactly, since $K \propto 1/v$), with a 17.4-fold deviation from the ATM value at the low end and –71% at the high end. Over a weekly horizon in a narrow band around the money the spread falls to $\approx 13\%$ — that is the region where a SABR-like approximation is defensible.

Table 6: Error of the geometric Heston smile against the exact price (**measured**).

Maturity	Near the money	At 20 % OTM
1 week	≤ 4.1 bp ($K \in [97.5, 120]$); -10.6 bp at $K = 95$	-212 bp
1 month	≤ 16.7 bp ($K \in [95, 110]$)	-186 bp
3 months	≤ 46.8 bp ($K \in [90, 110]$)	-123 bp

Non-constant curvature is exactly what costs you a closed form

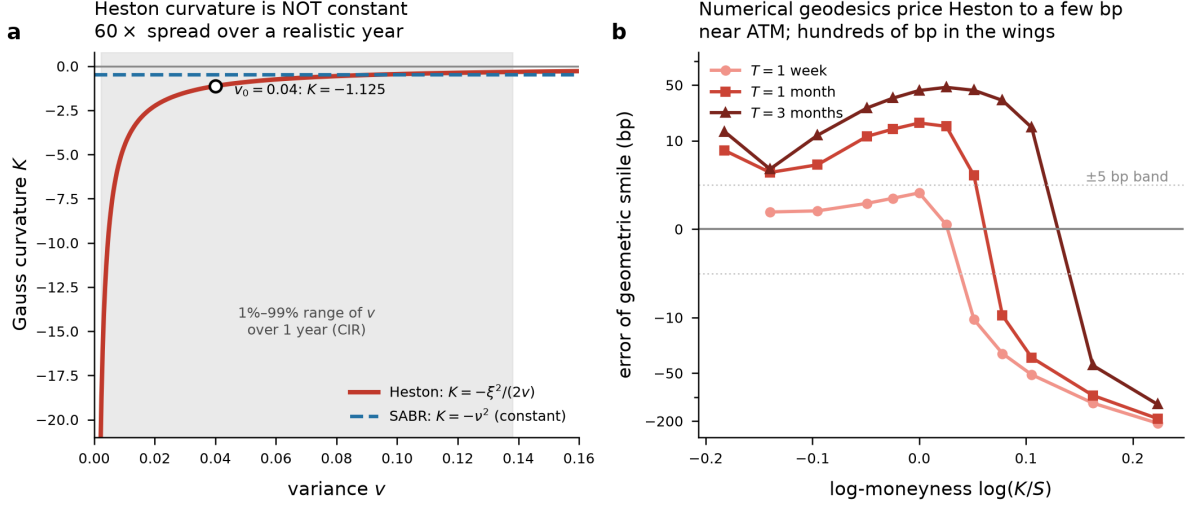


Figure 3: Non-constant curvature is exactly what costs the closed form. (a) $K_{\text{Heston}} = -\xi^2/(2v)$ against SABR’s constant $K = -v^2$; grey marks the realistic one-year range of v . (b) Error of the geometric smile against the exact Heston price by maturity and strike.

7.3 Numerical geodesics and the price of the approximation

We built a geodesic solver for the non-constant-curvature case — shooting on the initial direction with symbolically derived Christoffel symbols; unit speed is conserved to $1e-14$ and the solution is stable under resolution refinement to $1e-9$ (**measured**). A useful warning: the alternative method of minimising a discretised length on an x -monotonic grid errs by up to 8.75 % for deep out-of-the-money strikes under strong negative correlation, because the true minimal geodesic is not x -monotonic there; near the money the two methods agree to 0.2 %.

The reference is the COS method [7] on the Heston characteristic function, validated to 12 significant figures against direct Gil-Pelaez quadrature and against Monte Carlo. The accuracy of the geometric smile is given in Table 6; the near-the-money error grows as $T^{0.9-0.95}$, consistent with an $O(T)$ correction beyond the leading order.

The approximation breaks down in three places, each with a clear geometric cause: far strikes (the connecting geodesic traverses a region where the curvature deviates strongly from its at-the-money value), long maturities (departure from the asymptotic regime), and the neighbourhood of $v \rightarrow 0$ (the metric diverges as $1/v$ and the integration becomes stiff).

8 Validation of the SABR asymptotics

For completeness we report the measured accuracy of the leading order in the constant-curvature case. At $\beta = 1$ the geometric formula $\sigma_{\text{BS}} = |\log(F/K)|/d(K)$ and the leading part of Hagan’s formula agree to $1.2e-11$ bp (**measured**), as they must by the identity $\nu d = |x_H|$. Against

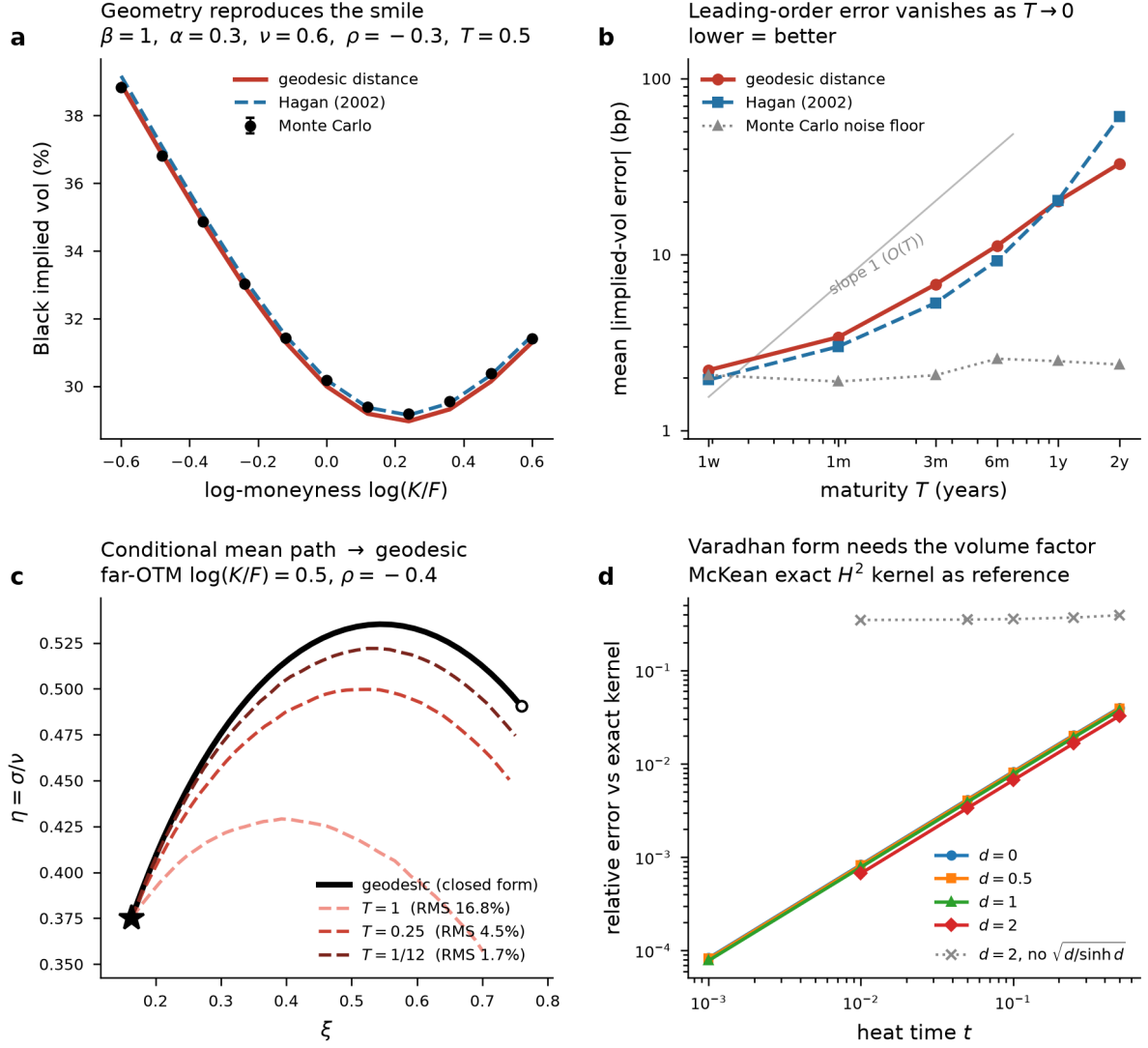


Figure 4: Validation of the leading order for SABR. **(a)** The smile: geometry, Hagan, and Monte Carlo. **(b)** Mean error by maturity; the simulation noise floor is shown in grey. **(c)** The conditionally averaged path approaches the geodesic as $T \rightarrow 0$. **(d)** Varadhan asymptotics against the exact McKean kernel: the volume factor is necessary.

Monte Carlo the mean absolute error is 2.2 bp at one week with a simulation noise floor of 2.1 bp, growing to 32.8 bp at two years; the measured log-log slope of 0.73–0.88 against the theoretical 1 indicates residual finite-time effects. At $T = 2$ years pure geometry (32.8 bp) is more accurate than the full Hagan formula (60.8 bp) — Hagan’s series is truncated too, and its $O(T)$ correction becomes an error source of its own at long maturities (**measured**).

The geodesic is the most likely path only in the small-time limit: the RMS deviation of the conditionally averaged path from the geodesic is 16.8% at $T = 1$, 4.5% at $T = 0.25$ and 1.7% at $T = 1/12$ (**measured**). A comparison with the exact McKean heat kernel [12] on $\mathbb{H}^2$ shows that the van Vleck–Morette factor $\sqrt{d/\sinh d}$ is not cosmetic in the wings (without it the error is fourfold at $d = 5$), while the residual in the SABR density at $\rho = 0$ is quantitatively attributed to an uncomputed path functional $\mathbb{E}[e^{-V_T/8} \mid \cdot]$ — matching an independently computed Jensen gap to three significant figures.

9 What of this is a working tool

Calibration diagnostics before calibration. The spectrum of (20) answers what the quotes are capable of determining at all. For SABR: β is unidentifiable from a single maturity across 0.32–0.86 at a 10 bp tolerance, so β should be fixed on grounds external to the smile. The computation takes seconds and requires no calibration.

Curvature as a solvability criterion. Formula (22) gives the criterion in one substitution: constant, and a closed form is plausible; non-constant, and numerical geodesics are required — with the magnitude of the curvature spread over the state region predicting directly where the approximation will fail.

Geodesic distances for comparing covariances. The distance $\|\log(S_1^{-1/2}S_2S_1^{-1/2})\|_F$ is invariant and meaningful as a distance between market regimes: suitable for regime clustering, change detection, and smoothing a covariance time series. It is *unsuitable* for unconditional risk estimation (Section 5.3).

Geodesic interpolation for stress tests. Linear interpolation yields a mid-path risk distorted by 27–36 %, and its notion of “halfway” depends on the parametrisation.

The natural gradient, when invariance matters. Not for speed, but so that the result does not depend on the coordinates of calibration.

What geometry does not give. It describes the smooth, leading-order part of the picture. Probabilistic corrections — $O(T)$ terms, path functionals, the volume factor — do not follow from it, and they dominate at horizons of order a year. This is a first approximation with clear structure, not a complete theory.

10 Status relative to the literature

No result in this text claims priority. The connection of SABR with $\mathbb{H}^2$, the asymptotics via geodesic distance, and the heat kernel are due to [11, 4, 9, 12], with a systematic treatment in [10]. The correction to the z argument is described independently in [6]. The geometry of $\text{SPD}(n)$, the Karcher barycentre and the swelling effect are due to [13, 14, 3]. The natural gradient and information geometry are due to [1, 2]. Invariant manifolds in HJM are due to [5].

What is our own here are not theorems but measurements and explicit statements: the hypercycle interpretation $\rho = \tanh \delta$ with a proof that the distance is constant (Theorem 2); the decomposition (14) into an isometry and a shift; the explicit gauge group with a demarcation, in basis points, of observable from hidden symmetry (Section 4.1); the closed form (22) with a complete list of constant-curvature cases; and three refuted hypotheses with numbers attached — the 26–37 % risk under-report of geodesic averaging, the natural gradient’s loss to Levenberg–Marquardt on speed, and the existence of a second constant-curvature case.

Reproducibility

The archive contains the directories `code/` (eight modules), `data/` (measurement tables), `paper/` (source and figures) and `reports/` (detailed protocols for each track, including the full symbolic derivations). The file `README.md` maps each claim in the text to the file that supports it. Symbolic statements are reproduced by running the corresponding modules; numerical ones carry their stated accuracy or standard error.

References

- [1] Amari, S. (1998). *Natural gradient works efficiently in learning*. Neural Computation 10(2), 251–276.
- [2] Amari, S. (2016). *Information Geometry and Its Applications*. Springer.
- [3] Arsigny, V., Fillard, P., Pennec, X., Ayache, N. (2007). *Geometric means in a novel vector space structure on symmetric positive-definite matrices*. SIAM J. Matrix Anal. Appl. 29(1), 328–347.
- [4] Berestycki, H., Busca, J., Florent, I. (2004). *Computing the implied volatility in stochastic volatility models*. Comm. Pure Appl. Math. 57(10), 1352–1373.
- [5] Björk, T., Christensen, B. J. (1999). *Interest rate dynamics and consistent forward rate curves*. Mathematical Finance 9(4), 323–348.
- [6] Choi, J., Wu, L. (2019). *The equivalent constant-elasticity-of-variance (CEV) volatility of the stochastic-alpha-beta-rho (SABR) model*. arXiv:1911.13123.
- [7] Fang, F., Oosterlee, C. W. (2008). *A novel pricing method for European options based on Fourier-cosine series expansions*. SIAM J. Sci. Comput. 31(2), 826–848.
- [8] Hagan, P. S., Kumar, D., Lesniewski, A. S., Woodward, D. E. (2002). *Managing smile risk*. Wilmott Magazine, 84–108.
- [9] Hagan, P. S., Lesniewski, A. S., Woodward, D. E. (2005). *Probability distribution in the SABR model of stochastic volatility*. Working paper.
- [10] Henry-Labordère, P. (2008). *Analysis, Geometry, and Modeling in Finance: Advanced Methods in Option Pricing*. Chapman & Hall/CRC.
- [11] Lesniewski, A. S. (2002). *WKB method for swaption smile*. Working paper.
- [12] McKean, H. P. (1970). *An upper bound to the spectrum of Δ on a manifold of negative curvature*. J. Differential Geometry 4, 359–366.
- [13] Moakher, M. (2005). *A differential geometric approach to the geometric mean of symmetric positive-definite matrices*. SIAM J. Matrix Anal. Appl. 26(3), 735–747.
- [14] Pennec, X., Fillard, P., Ayache, N. (2006). *A Riemannian framework for tensor computing*. Int. J. Computer Vision 66(1), 41–66.
- [15] Varadhan, S. R. S. (1967). *On the behavior of the fundamental solution of the heat equation with variable coefficients*. Comm. Pure Appl. Math. 20, 431–455.