CAN THERE BE AN EXPLICIT FORMULA FOR IMPLIED VOLATILITY?

STEFAN GERHOLD

ABSTRACT. It is "well known" that there is no explicit expression for the Black-Scholes implied volatility. We *prove* that, as a function of underlying, strike, and call price, implied volatility does not belong to the class of D-finite functions. This does not rule out all explicit expressions, but shows that implied volatility does not belong to a certain large class, which contains many elementary functions and classical special functions.

1. INTRODUCTION

The Black-Scholes formula is the most popular way to price European stock options. Assuming a frictionless market and an underlying modelled by geometric Brownian motion with volatility σ, the arbitrage free price of a European call option with strike K, maturity T, and initial stock price S is

$$C_{\mathrm{BS}}(S, K, T, \sigma) = SN(d_1) - KN(d_2),$$

where N is the cdf of the standard normal distribution, and

$$d_{1,2} = \frac{\log(S/K)}{\sigma\sqrt{T}} \pm \frac{\sigma\sqrt{T}}{2}.$$

(We assume a zero interest rate for simplicity.) Given a call price surface C that satisfies the no-arbitrage bounds

$$(S - K)^+ \leq C(S, K, T) < S, \tag{1}$$

the Black-Scholes formula can be inverted w.r.t. σ to obtain the *implied volatility* σ_{imp}:

$$C_{\mathrm{BS}}(S, K, T, \sigma_{\mathrm{imp}}(S, K, T)) = C(S, K, T). \tag{2}$$

Implied volatility allows to compare option prices for different strikes, maturities, underlyings, and valuation times. For entry points to the literature on implied volatility, see, e.g., [9, 10, 11, 16, 21].

The compute the implied volatility from a call price, equation (2) has to be solved numerically; many authors have added that the inversion cannot be done in closed form. The latter statement has been made so often that one might ask whether it can be turned into something tangible. First, note that it is hard to say anything about explicit expressions for the function

$$(S, K, T) \mapsto \sigma_{\mathrm{imp}}(S, K, T). \tag{3}$$

This function actually depends on $C(\cdot, \cdot, \cdot)$, and it would be unnatural to make assumptions on the generic call price surface C that would allow to refute a closed

form for (3). Instead, we consider the function I which satisfies

$$(4) \qquad C_{\mathrm{BS}}(S, K, I(S, K, c)) = c, \quad S, K > 0, \ (S - K)^+ < c < S,$$

and is defined on the open set

$$\mathcal{D}_I := \{(S, K, c) \in \mathbb{R}^3 : S, K > 0, \ (S - K)^+ < c < S\}.$$

The maturity T has been omitted here, and we will continue to do so, assuming a fixed $T > 0$ throughout. Note that any kind of explicit expression for I with variable T would imply an explicit expression for any fixed T. The function I clearly exists and is unique, since the Black-Scholes call price increases from the lower to the upper no-arbitrage bound as σ increases. If a call price surface C satisfies the slightly strengthened bounds

$$(5) \qquad (S - K)^+ < C(S, K) < S,$$

which usually hold for $T > 0$, then the corresponding implied volatility σ_{imp} satisfies

$$\sigma_{\mathrm{imp}}(S, K) = I(S, K, C(S, K)), \quad S, K > 0.$$

An example where the lower bound in (5) fails would be an out-of-the-money call that cannot move into-the-money, as may happen in a binomial model, say. We impose (5) instead of (1) to make the domain of I open, so that I is differentiable on its whole domain $\mathcal{D}_I$.

Since the Black-Scholes call price is real analytic for $S, K, \sigma > 0$, and the Black-Scholes vega $\partial C_{\mathrm{BS}}/\partial\sigma$ is positive, the implicit function theorem [15] shows that I is real analytic (hence C^∞) on $\mathcal{D}_I$. The question we now ask is whether the function I admits a closed form. To give a partial answer, we recall in Section 2 the definition of the class of D-finite functions. The main result of this note (Theorem 1 in Section 3 below) is that I is *not* D-finite.

We briefly mention some other results of the kind "a certain function does not belong to a certain class". There is a wealth of theorems about transcendental functions, i.e., non-algebraic ones; for examples, see, e.g., Schmidt [20] and Bell et al. [3] and the references given there. Rubel [19] discusses "transcendentally transcendental" functions, which means that they do not satisfy an algebraic differential equation. The class of D-finite functions, which we are interested in, lies in between the algebraic and differentially algebraic classes. Finally, Bronstein et al. [5] have shown that the Lambert W function is not Liouvillian, which roughly means that it cannot be expressed by iterated integration and exponentiation starting with an algebraic function.

2. D-finite functions

Suppose that a C^∞-smooth function f is defined on an open set $\mathcal{D}_f \subseteq \mathbb{R}^n$. It is called D-finite if it satisfies PDEs

$$P_{1,d_1}(\mathbf{x})\frac{\partial^{d_1} f(\mathbf{x})}{\partial x_1^{d_1}} + P_{1,d_1-1}(\mathbf{x})\frac{\partial^{d_1-1} f(\mathbf{x})}{\partial x_1^{d_1-1}} + \cdots + P_{1,1}(\mathbf{x})\frac{\partial f(\mathbf{x})}{\partial x_1} + P_{1,0}(\mathbf{x})f(\mathbf{x}) = 0,$$

$$(6) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots$$

$$P_{n,d_n}(\mathbf{x})\frac{\partial^{d_n} f(\mathbf{x})}{\partial x_n^{d_n}} + P_{n,d_n-1}(\mathbf{x})\frac{\partial^{d_n-1} f(\mathbf{x})}{\partial x_n^{d_n-1}} + \cdots + P_{n,1}(\mathbf{x})\frac{\partial f(\mathbf{x})}{\partial x_n} + P_{n,0}(\mathbf{x})f(\mathbf{x}) = 0,$$

valid for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{D}_f$, where $d_i \geq 1$ for $i = 1, \ldots, n$, and the P_{ij} are polynomials such that P_{i,d_i} is not identically zero on $\mathcal{D}_f$ for $i = 1, \ldots, n$. If f is real analytic, let us fix an arbitrary point $\mathbf{x}_0 \in \mathcal{D}_f$ and consider the Taylor expansion of f at $\mathbf{x}_0$. When viewed as a formal power series, the PDEs (6) show that its partial derivatives generate a finite dimensional vector space over the field

of rational functions. This algebraic definition of D-finiteness is the usual one in the literature [17, 22]. It cannot be used directly for *functions*, though: Unlike the field of formal Laurent series, $C^\infty(\mathcal{D}_f)$ is not a vector space over the field of rational functions. Recall that in one dimension, the field of formal Laurent series $R((X))$ in the indeterminate X consists of the formal series

$$\left\{ \sum_{n \geq n_0} a_n X^n : n_0 \in \mathbb{Z}, \ a_n \in \mathbb{R}, n \geq n_0 \right\}.$$

The multivariate case is more involved; see Aparicio Monforte and Kauers [2].

Rational, and, more generally, algebraic functions are D-finite, as are the elementary functions exp, log, sin, and cos (but not tan and cot). About 60% of the special functions covered in the classical handbook by Abramowitz and Stegun [1] are D-finite. Examples include Bessel functions, Airy functions, exponential integral, dilogarithm, and hypergeometric series. Furthermore, the class of D-finite functions is closed under addition, multiplication, (in-)definite integration, and Laplace transform. Division, on the other hand, does *not* preserve D-finiteness in general, nor does composition, unless the inner function is algebraic. Thus, $x \to e^{\sqrt{x}}$ is D-finite, whereas $x \to e^{e^x}$ and $x \to 1/\sin x$ are not. The closure under algebraic substitution will be applied in the proof of our Theorem 1 below.

D-finite functions resp. power series have been studied extensively in the discrete mathematics and symbolic computation literature. Indeed, provided the coefficients of the polynomials in (6) and the initial conditions can be represented with a finite amount of information, one has a finite data structure that can represent reasonably general functions (or, in symbolic computation, rather formal power series). The closure properties mentioned above are effective, in the sense that there is an algorithm that computes the D-finite specification for a sum of two given D-finite functions, etc. There are also algorithms for proving identities involving D-finite functions, which are about to render tables of such formulas (certain definite integral evaluations, e.g.) obsolete to some extent. For further information, see, e.g., Chyzak and Salvy [6] or Koutschan [14].

Summarizing, the class of D-finite functions is extensive, but still there are plenty of "explicit" functions that are not D-finite. The following theorem thus gives a partial answer to the question asked in the title of the present note.

3. Main result

Theorem 1. *The function $I : \mathcal{D}_I \subset \mathbb{R}^3 \to (0, \infty)$ defined by (4) is not D-finite.*

Proof. It is often hard to show directly that a multivariate function is not D-finite, whereas there are several methods available that deal with univariate functions [4, 7, 12]. Thus, our strategy is to find a useful specialization of I to a univariate function. We let the underlying's initial price be proportional to the strike: $S = eK$. For $0 < K < 1/e$, the inequality

$$(e-1)K < \hat{c}(K) := (e-1)K + eK^2 < eK$$

holds, and so the function

$$f(K) := I(eK, K, \hat{c}(K))$$

is well-defined for $0 < K < 1/e$. We assume that I is D-finite, and want to show that then f would be D-finite, too, by studying its Taylor series at $K = 1/(2e)$. First, we expand the analytic function I in a neighborhood of the point $(\frac{1}{2}, \frac{1}{2e}, \frac{1}{2} - \frac{1}{4e}) \in \mathcal{D}_I$:

$$(7) \qquad I(S, K, c) = \sum_{i,j,k \geq 0} \gamma_{ijk} (S - \tfrac{1}{2})^i (K - \tfrac{1}{2e})^j (c - (\tfrac{1}{2} - \tfrac{1}{4e}))^k.$$

Since I is D-finite, so is the formal power series

$$\sum_{i,j,k\geq 0} \gamma_{ijk} X^i Y^j Z^k$$

in the indeterminates X, Y, Z. (Recall that this means that its partial derivatives generate a finite dimensional subspace of the space of formal Laurent series, which is a vector space over the field of rational functions $\mathbb{R}(X, Y, Z)$.) The algebraic substitution $X \to eX$, $Y \to X$, $Z \to (e-1)X + eX^2$ preserves D-finiteness [14, 17, 22], and so the univariate formal power series

$$(8) \qquad \sum_{i,j,k\geq 0} (eX)^i X^k (eX + eX^2)^k$$

is D-finite. This series represents the analytic function $x \mapsto f(x + \frac{1}{2e})$ in a neighborhood of zero. Indeed, by (7),

$$f(K) = \sum_{i,j,k\geq 0} \gamma_{ijk}(eK - \tfrac{1}{2})^i (K - \tfrac{1}{2e})^j (\hat{c}(K) - (\tfrac{1}{2} - \tfrac{1}{4e}))^k,$$

and this is (8) with X replaced by $K - 1/(2e)$. This shows that D-finiteness of I implies D-finiteness of f.

We will employ the following variant of the Black-Scholes formula (see Roper and Rutkowski [18]; note that the maturity $T > 0$ is fixed throughout):

$$(9) \qquad C_{\mathrm{BS}}(S, K, \sigma) = (S - K)^+ + S \int_0^{\sigma\sqrt{T}} N'\Big(\frac{\log(S/K)}{v} + \frac{v}{2}\Big) dv.$$

Define the function $F : (0, \infty) \to (0, 1/e)$ by

$$(10) \qquad F(x) := \int_0^x N'\Big(\frac{1}{v} + \frac{v}{2}\Big) dv.$$

Then, for $S = eK$, the log-moneyness is $\log(S/K) = 1$, and equation (9) becomes

$$C_{\mathrm{BS}}(eK, K, \sigma) = (e - 1)K + eKF(\sigma\sqrt{T}).$$

By the definition of f, we infer that the equation

$$(e - 1)K + eKF(\sqrt{T}f(K)) = \hat{c}(K), \quad 0 < K < 1/e,$$

must hold, and hence

$$F(\sqrt{T}f(K)) = \frac{\hat{c}(K) - (e-1)K}{eK} = K,$$

or

$$f(K) = F^{-1}(K)/\sqrt{T}, \quad 0 < K < 1/e.$$

The following lemma shows that F^{-1} is not D-finite, hence f and I are not D-finite, either. $\qquad\square$

Lemma 2. *Let the function $F : (0, \infty) \to (0, 1/e)$ be defined by (10). Then the inverse function $F^{-1} : (0, 1/e) \to (0, \infty)$ is not D-finite.*

Proof. We will show that F^{-1} has a singularity at zero whose type is incompatible with D-finiteness. First, we determine the asymptotics of F itself as $x \to 0^+$. For $0 \leq v \leq x$, we have

$$1 \geq e^{-v^2/8} \geq e^{-x^2/8} = 1 + \mathrm{O}(x^2),$$

and so

$$(11) \qquad N'\Big(\frac{1}{v} + \frac{v}{2}\Big) = \frac{1}{\sqrt{2e\pi}} e^{-1/(2v^2)}(1 + \mathrm{O}(x^2)), \quad x \to 0^+.$$

Substitution, integration by parts, and the expansion of N at infinity yield

$$\frac{1}{\sqrt{2\pi}}\int_0^x e^{-1/(2v^2)}dv = \frac{1}{\sqrt{2\pi}}\int_{1/x}^\infty \frac{e^{-y^2/2}}{y^2}dy$$

$$= \frac{x}{\sqrt{2\pi}}e^{-1/(2x^2)} - \frac{1}{\sqrt{2\pi}}\int_{1/x}^\infty e^{-y^2/2}dy$$

$$= \frac{x}{\sqrt{2\pi}}e^{-1/(2x^2)} - 1 + N(1/x)$$

$$(12) \qquad\qquad = \frac{x^3}{\sqrt{2\pi}}e^{-1/(2x^2)}(1 + \mathrm{O}(x^2)).$$

From (11) and (12) we conclude

$$(13) \qquad F(x) = \frac{x^3}{\sqrt{2e\pi}}e^{-1/(2x^2)}(1 + \mathrm{O}(x^2)), \quad x \to 0^+,$$

and thus, taking logarithms,

$$(14) \qquad -\log F(x) = \frac{1}{2x^2} + \mathrm{O}\Big(\log\frac{1}{x}\Big), \quad x \to 0^+.$$

Now we study the asymptotics of F^{-1}. For small $x > 0$, formula (13) yields $F(x) \le e^{-1/(2x^2)}$. Replacing x by $F^{-1}(x)$ in this inequality and rearranging, we obtain

$$(15) \qquad F^{-1}(x) \le \frac{1}{\sqrt{2\log(1/x)}}.$$

Now we replace x by $F^{-1}(x)$ in (14) and use (15) in the O-estimate:

$$(16) \qquad \log\frac{1}{x} = \frac{1}{2F^{-1}(x)^2} + \mathrm{O}\Big(\log\log\frac{1}{x}\Big).$$

This implies that (15) is sharp:

$$(17) \qquad F^{-1}(x) \sim \frac{1}{\sqrt{2\log(1/x)}}, \quad x \to 0^+.$$

Observe that a univariate D-finite function g solves an ODE

$$P_d(x)g^{(d)}(x) + \cdots + P_1(x)g'(x) + P_0(x)g(x) = 0$$

with polynomial coefficients, where P_d is not identically zero. Such a function has an analytic continuation throughout the complex plane, up to a finite number of poles and branch points. At each such singularity x_0, an asymptotic expansion holds, which may feature powers of $x - x_0$, exponential terms, and *integral* powers of $\log(x - x_0)$. This 19th century result [13] is a powerful method for proving non-D-finiteness; see Flajolet et al. [7, 8] for details. In particular, a *fractional* power of a logarithm, as in (17), is a "forbidden" asymptotic element, and shows that F^{-1} is not D-finite. □

References

[1] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55 of National Bureau of Standards Applied Mathematics Series, For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.

[2] A. Aparicio Monforte and M. Kauers, *Formal Laurent series in several variables*. To appear in Expositiones Mathematica, 2012.

[3] J. P. Bell, N. Bruin, and M. Coons, *Transcendence of generating functions whose coefficients are multiplicative*, Trans. Amer. Math. Soc., 364 (2012), pp. 933–959.

[4] J. P. Bell, S. Gerhold, M. Klazar, and F. Luca, *Non-holonomicity of sequences defined via elementary functions*, Annals of Combinatorics, 12 (2008), pp. 1–16.

[5] M. Bronstein, R. M. Corless, J. H. Davenport, and D. J. Jeffrey, *Algebraic properties of the Lambert W function from a result of Rosenlicht and of Liouville*, Integral Transforms Spec. Funct., 19 (2008), pp. 709–712.

[6] F. Chyzak and B. Salvy, *Non-commutative elimination in Ore algebras proves multivariate identities*, J. Symbolic Comput., 26 (1998), pp. 187–227.

[7] P. Flajolet, S. Gerhold, and B. Salvy, *On the non-holonomic character of logarithms, powers and the nth prime function*, Electronic Journal of Combinatorics, 11 (2005), pp. 1–16.

[8] P. Flajolet, S. Gerhold, and B. Salvy, *Lindelöf representations and (non-)holonomic sequences*, Electronic Journal of Combinatorics, 17 (2010).

[9] M. Forde and A. Jacquier, *Small-time asymptotics for implied volatility under the Heston model*, Int. J. Theor. Appl. Finance, 12 (2009), pp. 861–876.

[10] P. Friz, S. Gerhold, A. Gulisashvili, and S. Sturm, *On refined volatility smile expansion in the Heston model*, Quantitative Finance, 11 (2011), pp. 1151–1164.

[11] J. Gatheral, *The Volatility Surface, A Practitioner's Guide*, Wiley, 2006.

[12] S. Gerhold, *On Some Non-Holonomic Sequences*, Electronic Journal of Combinatorics, 11 (2004), pp. 1–8.

[13] E. L. Ince, *Ordinary differential equation*, Dover, 1926.

[14] C. Koutschan, *Advanced Applications of the Holonomic Systems Approach*, PhD thesis, J. Kepler University Linz, 2009.

[15] S. G. Krantz and H. R. Parks, *The implicit function theorem*, Birkhäuser Boston Inc., Boston, MA, 2002. History, theory, and applications.

[16] R. W. Lee, *Implied volatility: statics, dynamics, and probabilistic interpretation*, in Recent advances in applied probability, Springer, New York, 2005, pp. 241–268.

[17] L. Lipshitz, *D-finite power series*, J. Algebra, 122 (1989), pp. 353–373.

[18] M. Roper and M. Rutkowski, *On the relationship between the call price surface and the implied volatility surface close to expiry*, Int. J. Theor. Appl. Finance, 12 (2009), pp. 427–441.

[19] L. A. Rubel, *A survey of transcendentally transcendental functions.*, Am. Math. Mon., 96 (1989), pp. 777–788.

[20] W. M. Schmidt, *Eisenstein's theorem on power series expansions of algebraic functions*, Acta Arith., 56 (1990), pp. 161–179.

[21] M. Schweizer and J. Wissel, *Term structures of implied volatilities: absence of arbitrage and existence results*, Math. Finance, 18 (2008), pp. 77–114.

[22] R. P. Stanley, *Differentiably finite power series*, European J. Combin., 1 (1980), pp. 175–188.

Vienna University of Technology, Institute of Mathematical Methods in Economics, Wiedner Hauptstr. 8/105-1, A-1040 Vienna, Austria

E-mail address: `sgerhold at fam.tuwien.ac.at`