THE INVERSE VOLATILITY PROBLEM FOR AMERICAN OPTIONS

IAN KNOWLES* AND AJAY MAHATO
Department of Mathematics
University of Alabama at Birmingham
Birmingham, AL 35294, USA

Abstract. The problem of determining equity volatility from a knowledge of American option prices for a range of exercise (strike) prices and expirations is solved by minimization of a convex functional.

1. Introduction. It is common to model a financial asset, such as a stock or a commodity, via the stochastic differential equation

\[
\frac{dS_t}{S_t} = m(S_t, t) dt + \sigma(S_t, t) dB_t, \tag{1}
\]

where, for each time \(t\), \(S_t(\omega)\) is a random variable representing the price of the financial asset for the trial \(\omega\), \(m\) is the drift, which relates to the “trend” of the asset, \(\sigma\) is the volatility (the “wobble” in the asset graph), and \(B_t(\omega)\) is a Brownian motion stochastic process used to model the randomness. Financial derivatives are contracts that derive their value from such an underlying asset. In particular, a European put (call) option on a stock is a contractual right to sell (buy) one share of the stock at a specified price \(K\) (the strike, or exercise, price) at a specified future time \(T\) (the expiration, or expiry, date), and an American option is defined similarly with the additional provision of allowing early exercise of the contract.

It is known that the daily financial transactions associated with these forward-looking instruments give rise to a huge amount of raw data, buried inside of which is the market’s best guess as to what the future holds. We are concerned here with the possibility of extracting future-oriented information from this type of data with the aid of certain computational inverse algorithms.

In their Nobel Prize winning paper [3] the economists Fischer Black and Myron Scholes determined that, under certain rather severe restrictions, the arbitrage-free price \(v = v(S, t; K, T)\) of a European put contract satisfies a homogeneous partial differential equation (PDE) of diffusion type in time \(t\) and the value \(S\) of the underlying asset:

\[
\mathcal{L}_B v = \frac{\partial v}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 v}{\partial S^2} + \mu S \frac{\partial v}{\partial S} - rv = 0, \quad t \in [0, T], \quad S \in [0, \infty), \tag{2}
\]

with boundary conditions

\[
v(0, t) = K e^{-r(T-t)}, \quad v(S, t) \to 0 \text{ as } S \to \infty, \tag{3}
\]

2010 Mathematics Subject Classification. Primary: 35K10; Secondary: 65N21.

Key words and phrases. Inverse volatility, American options, convex functional.

* Corresponding author.
and a terminal condition
\[ v(S, T) = (K - S)^+ = \max(K - S, 0), \tag{4} \]
where \( \sigma \) is the (assumed constant) volatility, \( \mu \) is the risk-neutral drift, and \( r \) is the short-term interest rate, which for practical purposes may be taken to be the interest rate on a 13-week US government treasury bill.

The work of many authors over the years [9, 19, 22, 23, 25, 26, 27] makes it clear that several of the assumptions laid down by Black and Scholes are basically incompatible with modern market data. Notable among these are objections to the constancy of \( \sigma \). As recently as the early nineteen eighties, assertions such as “the Black-Scholes volatility is constant” seemed to hold true, at least while the market believed it so; but then, after the crash of 1987, volatility was anything but constant, and in fact it has now even become fashionable to speak of, and invest in, the volatility of the volatility. So, at the very least one may regard the volatility as a function of \( S \) and \( t \), \( \sigma = \sigma(S, t) \), and we assume such a structure in the sequel.

Now, financial data specifying the market price \( v \) of options is readily available in quantity at various strike values \( K \) around the current price of the underlying asset (the “spot” price), and for values of the expiration \( T \) up to six months or so into the future. Given that the computer models currently used by stock analysts do not project very far into the future, one is led quite naturally to the so-called inverse volatility problem: determine a market-inspired estimate of the future volatility function \( \sigma(S, t) \) from a knowledge of current market prices \( v \) of options with different strikes and future expirations. In financial circles this is usually known as the calibration problem.

As we intend to concentrate here on the calibration problem using dividend-free American option price data as our main example, and it is known that under these conditions early exercise of American call options is not advantageous, in the sequel we exclusively consider American put option data. It is not difficult to extend this material to American call and put options on dividend paying stocks. Before proceeding we note that, as has been elegantly observed by Kholodnyi [12], an American put option satisfies the following semilinear Black-Scholes terminal value problem:

\[
\mathcal{L}_{BS} v = -r K h((K - S)^+ - v), \quad S \geq 0, \quad t \in [0, T],
\]
\[
v(0, t) = K e^{-r(T-t)}, \quad v(S, t) \to 0 \text{ as } S \to \infty, \tag{6}
\]
\[
v(S, T) = (K - S)^+, \tag{7}
\]
where \( h \) is the Heaviside function.

A number of approaches have been proposed to compute volatility surfaces from American option prices. A Tikhonov regularization technique requiring an initial estimate of volatility has been implemented in [7], and constrained optimization methods based upon solving the complementarity problem have been used in [11, 29]; a related least squares minimization approach was used in [1]. In [20] a generalized Newton minimization is applied via a quadratic approximation to obtain the implied volatility. Related work on perpetual American options with time homogeneous volatility appears in [2] and alternatives involving so-called jump-diffusion models of the underlying assets are discussed in [5, 6].

We approach this task as follows. The value \( v \) of an option contract also depends on the exercise (strike) price, \( K \), and the expiration date, \( T \), of the contract, so \( v = v(S, t; K, T) \). In 1994 Bruno Dupire [8] noticed that for European options the
function \( v(S,t;K,T) \) satisfies the “dual” Black-Scholes equation (also known also as the Dupire equation) in the variables \( K \) and \( T \):

\[
L^*_{BS} v = \frac{\partial v}{\partial T} - \frac{1}{2} K^2 \sigma^2(K,T) \frac{\partial^2 v}{\partial K^2} + \mu K \frac{\partial v}{\partial K} - (\mu - r)v = 0. \tag{8}
\]

In what follows, it is convenient to denote the right hand side term in equation (5) by the function \( g^*(S,t;K,T) \), and we can thus rewrite equation (5) as the equation

\[
L_{BS} v = g^*(S,t;K,T), \tag{9}
\]

which we call the inhomogeneous Black-Scholes equation. We now show that for American options the price \( v(S,t;K,T) \) satisfies a dual version of the inhomogeneous Black-Scholes equation (which we call the inhomogeneous Dupire equation) in the variables \( K \) and \( T \)

\[
L^*_{BS} v = \frac{\partial v}{\partial T} - \frac{1}{2} K^2 \sigma^2(K,T) \frac{\partial^2 v}{\partial K^2} + \mu K \frac{\partial v}{\partial K} - (\mu - r)v = g(S,t;K,T), \tag{10}
\]

where the function \( g(S,t;K,T) \) is defined below.

**Theorem 1.1.** Set \( \Omega = [0, \infty) \) and let \( v_0 \) satisfy the Cauchy problem

\[
L_{BS} v_0 = g^*(S,t;K,T), \\
v_0|_{\partial \Omega} = 0, \quad v_0|_{t=T} = 0. \tag{11}
\]

If \( v(S,t;K,T) \) satisfies the inhomogeneous Black-Scholes equation (9) with boundary condition (6) and terminal condition (7) in the variables \( S \) and \( t \), then the function \( v(S,t;K,T) \) also satisfies the dual equation (10) in the variables \( K \) and \( T \), where we define

\[
g(S,t;K,T) = L^*_{BS} v_0. \tag{12}
\]

**Proof.** As \( g^*(S,t;K,T) \) is bounded, by [21, Theorem 4.1] there exists a solution \( v_0 \) for the Cauchy problem (11). The function \( v - v_0 \) satisfies

\[
L_{BS}(v - v_0) = 0, \tag{13}
\]

in the variables \( S \) and \( t \), as well as boundary condition (6) and terminal condition (7). Since \( v - v_0 \) satisfies the Black-Scholes equation and associated conditions in \( S \) and \( t \), by standard theory [4, §2] it must also satisfy the Dupire equation (8) in \( K \) and \( T \), so that \( L^*_{BS}(v - v_0) = 0 \). Thus

\[
L^*_{BS} v = L^*_{BS} v_0 = g(S,t;K,T),
\]

as required.

**Remark 1.** It is worth noting here that if one defines \( w = v - v_0 \) then we have that the American option price \( v \) may be written as

\[
v = w + v_0.
\]

Here, as \( w = w(S,t;K,T) \) satisfies the classical Black-Scholes conditions by (13), it may be considered to be the the price of the European-style put option associated with the same underlying asset, were such an entity to exist. The function \( v_0 \), which must be non-negative (by the non-negativity of the boundary data and the maximum principle for these parabolic equations), then takes on the rôde of the *premium*, the necessarily non-negative price difference between the less flexible European and the more flexible American option.
In this paper we present a new variational algorithm for computing, via our inhomogeneous Dupire equation (10), the volatility \( \sigma(K,T) \) from a knowledge of American option prices at various strikes and expirations. The method used is an adaption of the variational approach involving the minimization of convex functionals used in [18] for numerical differentiation (formulated as an inverse problem), and in [16, 17] for solving the inverse groundwater modeling problem. This approach has distinct advantages over traditional least squares techniques in that, as the associated functionals are convex, one typically has unique global minima and stationary points, so that one is not concerned about the steepest descent minimization processes “getting stuck” in spurious local minima. This is an enormous advantage numerically in difficult situations. In addition and for similar reasons, the method is quite adept at the simultaneous recovery of multiple coefficient functions in the differential equation, as we shall see presently.

2. Reconstruction of volatility. For simplicity of exposition we assume that there is no dividend for the underlying asset. Thus the risk-neutral drift \( \mu \) in (9) and (10) is equal to the interest rate \( r \), and the inhomogeneous Dupire equation (10) can be written as

\[
\frac{\partial v}{\partial T} - \frac{1}{2} K^2 \sigma^2(K,T) \frac{\partial^2 v}{\partial K^2} + rK \frac{\partial v}{\partial K} = g(S_0,t_0,K,T),
\]

where \( t_0 \) is the fixed time at which the option data is gathered, and \( S_0 = S(t_0) \) is the associated “spot” price. Let \( T_0 < T_1 < \ldots < T_n \) be expiration times, and for each expiration \( T_i \), \( 0 \leq i \leq n \), let \( K_{i1}, \ldots, K_{im_i} \) be the associated strike prices. We assume that the volatility \( \sigma = \sigma(K,T) \) and \( g = g(S_0,t_0,K,T) \) are piecewise constant in \( T \), i.e.

\[
\sigma(K,T) = \sum_{i=1}^{n} \chi_{[T_{i-1},T_i]}(T) \sigma_i(K) \quad \text{and} \quad g(S_0,t_0,K,T) = \sum_{i=1}^{n} \chi_{[T_{i-1},T_i]}(T) g_i(K),
\]

where \( \chi_{[T_{i-1},T_i]}(T) \) denotes the characteristic (indicator) function of the time interval \([T_{i-1},T_i] \). Fixing \( i \), set

\[
w_\lambda(K) = \int_{T_{i-1}}^{T_i} e^{-\lambda T} v(K,T) dT,
\]

where \( \lambda > 0 \) is a parameter. For each fixed \( i, 1 \leq i \leq n \), we now Laplace transform the inhomogeneous Dupire equation (14) over \([T_{i-1},T_i] \) to obtain

\[
\int_{T_{i-1}}^{T_i} e^{-\lambda T} v_K dT - \frac{1}{2} K^2 \sigma_i^2 \int_{T_{i-1}}^{T_i} e^{-\lambda T} v_{KK} dT \underbrace{w''_\lambda}_{w'} + rK \int_{T_{i-1}}^{T_i} e^{-\lambda T} v_K dT = \int_{T_{i-1}}^{T_i} e^{-\lambda T} g_i dT,
\]

where the primes indicate differentiation with respect to \( K \). On integrating the first term by parts we get
\[ [e^{-\lambda T}v]_{T_i}^{T_{i-1}} + \lambda \int_{T_{i-1}}^{T_i} e^{-\lambda T} v \, dT - \frac{1}{2} K^2 \sigma_i^2 w''_\lambda + rK w'_\lambda = g_i \int_{T_{i-1}}^{T_i} e^{-\lambda T} \, dT, \]

and rearranging terms gives,

\[- \frac{1}{2} K^2 \sigma_i^2 w''_\lambda + rK w'_\lambda + \lambda w_\lambda = g_i \int_{T_{i-1}}^{T_i} e^{-\lambda T} \, dT - v(K, T_i) e^{-\lambda T_i} + v(K, T_{i-1}) e^{-\lambda T_{i-1}}. \]

Next, dividing by \( \frac{1}{2} K^2 \sigma_i^2 \) throughout, we obtain

\[- (w''_\lambda - \frac{2r}{K^2 \sigma_i^2} w'_\lambda) + \frac{\lambda}{\frac{1}{2} K^2 \sigma_i^2} w_\lambda = \frac{1}{\frac{1}{2} K^2 \sigma_i^2} g_i \int_{T_{i-1}}^{T_i} e^{-\lambda T} \, dT \]

\[+ \frac{1}{\frac{1}{2} K^2 \sigma_i^2} [ -v(K, T_i) e^{-\lambda T_i} + v(K, T_{i-1}) e^{-\lambda T_{i-1}}]. \]

Finally, on multiplying by the integrating factor

\[ P(K) = e^{-\frac{2r}{K^2 \sigma_i^2} \int_{K_{i-1}}^{K_i} \frac{dk}{K^2 (K^2 + 1)}}, \]

we now have an equation in Sturm-Liouville form:

\[- (P(K) w'_\lambda)' + \lambda Q(K) w_\lambda = \alpha_\lambda F(K) + \beta_\lambda(K) Q(K), \]

where

\[ Q(K) = \left( \frac{2}{K^2 \sigma_i^2(K)} \right) P(K), \]

\[ F(K) = Q(K) g_i(K), \]

\[ \alpha_\lambda = \int_{T_{i-1}}^{T_i} e^{-\lambda T} \, dT, \]

\[ \beta_\lambda(K) = -v(K, T_i) e^{-\lambda T_i} + v(K, T_{i-1}) e^{-\lambda T_{i-1}}. \]

If we can recover the functions \( P(K), Q(K) \) and \( F(K) \) for each \( i, 1 \leq i \leq n \), we can find the volatility \( \sigma_i(K) \), and the \( g_i(K) \), from the following formulae:

\[ \sigma_i(K) = \sqrt{\frac{2P(K)}{K^2 Q(K)}}. \]

\[ g_i(K) = \frac{F(K)}{Q(K)}. \]

We now focus attention on a variational approach to the recovery of one such trio of coefficient functions \( P, Q, F \) defined on an interval \( a \leq K \leq b \). It is assumed that we are given the functions \( w_\lambda(K) \) for \( K \in [a, b] \) and all \( \lambda > 0 \). For positive functions \( p \) and \( q \) and a function \( f \) all defined on \( [a, b] \), let \( c = (p, q, f) \). Define \( w = w_{\lambda,c}(K) \) to be the solution to the boundary value problem

\[ L_{p,\lambda q}(w) = -(p(K) w')' + \lambda q(K) w = \alpha_\lambda f(K) + \beta_\lambda(K) q(K), \]

\[ w(a) = w_\lambda(a), \quad w(b) = w_\lambda(b). \]

Let

\[ D = \{(p, q, f) : \text{the equation (25) is disconjugate on } [a, b]\}, \]

so that for each \( (p, q, f) \) in \( D \) every non-trivial solution of \( (25) \) has at most one zero on \([a,b]\). It is known [10, Theorem 6.1, p. 351] that \( (25) \) is disconjugate if and
only if the boundary value problem (25, 26) can always be solved uniquely. It is also known (c.f. [18, Proposition 2.1]) that this set is open and convex in \((L[a, b])^3\) and \((L^2[a, b])^3\). For each \(\lambda > 0\) define the functional \(G_\lambda\) on the convex set \(D\) by

\[
G_\lambda(c) = \int_a^b [p(K)(w'_{\lambda,c}^2 - w_{\lambda,c}^2) + \lambda q(K)(w_{\lambda}^2 - w_{\lambda,c}^2) - 2(\alpha f(K) + \beta q(K))(w_{\lambda} - w_{\lambda,c})] dK. \tag{28}
\]

3. Properties of the functional \(G_\lambda\). The main properties of the functionals \(G_\lambda\) are summarized in the following

**Theorem 3.1.**  (a) For any \(c = (p, q, f)\) in \(D\),

\[
G_\lambda(c) = \int_a^b p(w'_{\lambda} - w'_{\lambda,c})^2 + \lambda q(w_{\lambda} - w_{\lambda,c})^2. \tag{29}
\]

(b) \(G_\lambda(c) \geq 0\) for all \(c = (p, q, f)\) in \(D\), and \(G_\lambda(c) = 0\) if and only if \(w_{\lambda} = w_{\lambda,c}\).

(c) For \(h = (h_1, h_2, h_3) \in (L^\infty[a, b])^3\) the first Gâteaux derivative of \(G_\lambda\) is given by

\[
G_\lambda'(p, q, f)[h] = \int_a^b (w'_{\lambda} - w'_{\lambda,c}) h_1 + \left[ \lambda (w_{\lambda}^2 - w_{\lambda,c}^2) - 2\beta (w_{\lambda} - w_{\lambda,c}) \right] h_2 - 2\alpha (w_{\lambda} - w_{\lambda,c}) h_3. \tag{30}
\]

(d) The second Gâteaux derivative of \(G_\lambda\) is given by

\[
G_\lambda''(c)[h, k] = 2(\lambda_{p,\lambda}^{-1} (e_\lambda(h), e_\lambda(k))), \tag{31}
\]

where \(h = (h_1, h_2, h_3), k = (k_1, k_2, k_3)\),

\[
e_\lambda(h) = -(h_1 w'_{\lambda,c})' + \lambda h_2 w_{\lambda,c} - \beta h_2 - \alpha h_3,
\]

and \((\cdot, \cdot)\) denotes the usual inner product in \(L^2[a, b]\).

**Proof.**  (a) If \(v \in W^{1,2}[a, b]\) and \(\phi \in W^{1,2}_0[a, b]\) then from integration by parts we have

\[
\int_a^b p(x)v'\phi' dx = \left. \frac{p(x)v\phi}{0} \right|_a^b - \int_a^b \phi(p(x)v')' dx = - \int_a^b \phi(p(x)v')' dx. \tag{32}
\]

Consequently, from (32) using \(\phi = w_{\lambda} - w_{\lambda,c} \in W^{1,2}_0[a, b]\),

\[
G_\lambda(c) = \int_a^b p(w'_{\lambda} - w'_{\lambda,c})^2 + \lambda q(w_{\lambda}^2 - w_{\lambda,c}^2) - 2(\alpha f + \beta q)(w_{\lambda} - w_{\lambda,c})
\]

\[
= \int_a^b p(w'_{\lambda} - w'_{\lambda,c})^2 + 2p(w_{\lambda} - w_{\lambda,c})w'_{\lambda,c}
\]

\[
+ \lambda q(w_{\lambda}^2 - w_{\lambda,c}^2) - 2(\alpha f + \beta q)(w_{\lambda} - w_{\lambda,c})
\]
As p and q are chosen to be positive and \( \lambda > 0 \), from (a) we get (b).

(c) The first Gâteaux derivative of the functional \( G_\lambda \) is given by

\[
G'_\lambda(p,q,f)[h] = \lim_{\varepsilon \to 0} \frac{G_\lambda(c + \varepsilon h) - G_\lambda(c)}{\varepsilon}
\]

\[
= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_a^b \left( p + \varepsilon h_1 \right)(w_\lambda^2 - w_{\lambda,c+\varepsilon h}^2) + \lambda(q + \varepsilon h_2)(w_\lambda^2 - w_{\lambda,c+\varepsilon h}^2)
\]

\[
+ 2(\alpha f + \beta q)(w_\lambda - w_{\lambda,c}) + 2\varepsilon \beta(w_\lambda - w_{\lambda,c+\varepsilon h})
\]

\[
+ \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_a^b p(w_\lambda^2 - w_{\lambda,c+\varepsilon h}^2) + \lambda(q + \varepsilon h_2)(w_\lambda^2 - w_{\lambda,c+\varepsilon h}^2)
\]

\[
+ 2\varepsilon \beta(w_\lambda - w_{\lambda,c+\varepsilon h})
\]

If we can show that the second limit above is zero then we have proved 3.1(c).

Let the integral in the second term be denoted by \( I \) and note that

\[
-(pw'_\lambda c)^2 + \lambda w_{\lambda,c} = \alpha f + \beta q,
\]

(33)

\[-((p + \varepsilon h_1)w'_\lambda c + \varepsilon h_2)w_{\lambda,c+\varepsilon h} = \alpha(f + \varepsilon h_3) + \beta(q + \varepsilon h_2).
\]

(34)

The first term in the integral \( I \) may now be expanded as

\[
\varepsilon^{-1} \int_a^b p(w_\lambda^2 - w_{\lambda,c+\varepsilon h}^2)
\]

\[
= \varepsilon^{-1} \int_a^b p(w_\lambda^2 + w_\lambda w_{\lambda,c+\varepsilon h})(w'_\lambda c - w_{\lambda,c+\varepsilon h})
\]

from (32) using \( \phi = w_\lambda - w_{\lambda,c+\varepsilon h} \),
\[ L_{p, λ}(w_{λ,c+εh} - w_{λ,c}) = -(p(w_{λ,c+εh} - w_{λ,c}))' + λq(w_{λ,c+εh} - w_{λ,c}) \]
\[ = -(pw_{λ,c+εh})' + λqw_{λ,c+εh} - \left( -(pw_{λ,c})' + λqw_{λ,c} \right) \]
\[ = ε[(h_1 w_{λ,c+εh}′) - λh_2 w_{λ,c+εh} + βh_2 + αh_3]. \] (35)

By the definition of the second Gâteaux derivative, we have
\[ G''_λ(c)[h, k] = \lim_{ε \to 0} \frac{G'_λ(c + εh)[k] - G'_λ(c)[k]}{ε} \]
\[ = \lim_{ε \to 0} \frac{1}{ε} \int_a^b \left( (w_{λ,c+εh}'' - w_{λ,c}''')k_1 + [λ(w_{λ,c+ε}'' - w_{λ,c}''')] - 2β(w_{λ,c} - w_{λ,c+εh}) \right)k_2 \]
\[ - 2α(w_{λ,c} - w_{λ,c+εh})k_3 - (w_{λ,c}'' - w_{λ,c}''')k_1 \]
\[ - [λ(w_{λ,c}'' - w_{λ,c}''')] - 2β(w_{λ,c} - w_{λ,c+εh})]k_2 + 2α(w_{λ,c} - w_{λ,c})k_3 \]
\[ = \lim_{ε \to 0} \frac{1}{ε} \int_a^b \left( (w_{λ,c+εh}'' - w_{λ,c}''')k_1 + [λ(w_{λ,c+ε}'' - w_{λ,c}''')] - 2β(w_{λ,c} - w_{λ,c+εh}) \right)k_2 \]
\[ - 2α(w_{λ,c} - w_{λ,c+εh})k_3 \]
\[ = \lim_{ε \to 0} \frac{1}{ε} \int_a^b \left( (w_{λ,c} - w_{λ,c+εh})((k_1(w_{λ,c}′ + w_{λ,c+εh}′))' \right) \]
\[ + [λ(w_{λ,c}'' - w_{λ,c}''')] - 2β(w_{λ,c} - w_{λ,c+εh})]k_2 \]
\[ - 2α(w_{λ,c} - w_{λ,c+εh})k_3 \]
\[ = \lim_{ε \to 0} \frac{1}{ε} \int_a^b (w_{λ,c} - w_{λ,c+εh}) \cdot \]
\[ [(k_1(w_{λ,c}′ + w_{λ,c+εh}′))' + (λ(w_{λ,c} + w_{λ,c+εh}) - 2β)]k_2 - 2αk_3], \]
using (35).
Define a convex functional $G$ with derivatives of $w$ only if from Theorem 3.1(b) that $G$ is therefore convex on $D$. We know from Theorem 3.1(b) that $G$ has a global minimum (zero) at $c = (p, q, f)$ if and only if $w_\lambda = w_{\lambda,c}$. Choose $N \geq 3$ positive distinct real numbers $\lambda_j$, $1 \leq j \leq N$, so that

$$0 < \lambda_j T < 2, \quad T \in [T_{i-1}, T_i].$$

Define a convex functional $G$ on the domain $D$ defined by (27) with

$$G(c) = \sum_{j=1}^{N} G_{\lambda_j}(c).$$

We have $G(c) \geq 0$, and $G(c) = 0$ when $c = (P, Q, F)$. Moreover, when $G(c) = 0$ from Theorem 3.1(b) we have that $w_\lambda = w_{\lambda,c}$ for $1 \leq i \leq N$. From the uniqueness theorem [13, Theorem 3.5] we know that, under certain (computer-verifiable) conditions on the nature of the flows of certain associated vector fields (which amount here to an admissibility restriction on the data $v(K,T)$), the condition $w_\lambda = w_{\lambda,c}$ for at least three distinct values of $\lambda$ implies that

$$c = (p, q, f) = (P, Q, F),$$

as required.

$\square$
so \( G \) has a unique global minimum at \( c = (P, Q, F) \), which is also a stationary point for \( G \). By \cite[Proposition 42.6(1)]{31} we know that if the convex functional \( G \) has a stationary point at a point \((p, q, f)\) then it must have a global minimum there, and from the foregoing (assuming admissible data) that stationary point must uniquely occur at \((P, Q, F)\). So, the desired function trio \((P, Q, F)\) now appears as the unique global minimum of a convex functional with a unique stationary point. In practical numerics this is an important consideration, as many (if not most) least-square type minimization methods suffer greatly from the minimization process getting stuck in spurious local minima. That this cannot happen here is one of the significant advantages of our approach.

4. The algorithm. \( G(c) \) is a nonnegative convex functional as it is the sum of nonnegative convex functionals, and it also has a unique stationary point at \( c = (P, Q, F) \). The idea here is that by using \( G \) rather than just one of the \( G_{\lambda} \), in addition to gaining favorable uniqueness properties, we are blending additional time-based data into the inverse problem, and this is intended to improve the well-posedness of the problem. We note in passing from \cite{14} that this inverse recovery is conditionally well-posed in the weak-\( L^2 \) sense, so from a theoretical standpoint, the recoveries are expected to be quite stable, which indeed is the case.

We minimize this functional for \( N = 20 \) using the steepest descent method to recover the coefficients \( P(K), \ Q(K) \) and \( F(K) \). The gradients for \( G \) at the point \( c = (p_0, q_0, f_0) \) with respect to \( p, q \) and \( f \) are given by

\[
\nabla_{L^1,p} G(c) = \sum_{j=1}^{N} \left( w_{\lambda_j}^2 - w_{\lambda_j,c}^2 \right),
\]

\[
\nabla_{L^1,q} G(c) = \sum_{j=1}^{N} \left[ \lambda_j (w_{\lambda_j}^2 - w_{\lambda_j,c}^2) - 2\beta (w_{\lambda_j} - w_{\lambda_j,c}) \right],
\]

\[
\nabla_{L^2,f} G(c) = \sum_{j=1}^{N} -2\alpha (w_{\lambda_j} - w_{\lambda_j,c}),
\]

respectively.

Instead of using these gradients directly, we use the corresponding Neuberger-gradients (see \cite{24}) as the regular gradient has numerical problems that are extensively discussed in \cite{18}. In particular, the \( L^1 \)-gradient with respect to \( q \) is zero on the boundary of \([a, b]\), given that \( w_{\lambda} \) and \( w_{\lambda,c} \) are equal there, and consequently, under this gradient the algorithm is unable to properly recover \( Q \). The Neuberger-gradient process smooths an \( L^r \)-gradient and preserves boundary data during the descent, an important property not shared by other descent techniques. Our Neuberger-gradient \( g = \nabla_{H^1} G \) can be found from an \( L^r \)-gradient \( \nabla_{L^r} G \) by solving the boundary value problem

\[
-g'' + g = \nabla_{L^r} G, \quad g(a) = g(b) = 0.
\]  

Below is the steepest descent algorithm used to get one descent step in \( p \):

1. Initialize \( p(K), q(K) \) and \( f(K) \) with \( c = (p_0, q_0, f_0) \).
2. Find \( w_{\lambda,c} \) and \( w_{\lambda,c}' \) by solving \((25),(26)\).
3. Find the \( L^1 \)-gradient of \( G \) in \( p \), \( \nabla_{L^1,p} G(c) \).
4. Find the Neuberger gradient in \( p \), \( \nabla_{H^1,p} G(c) \).
5. Evaluate \( p_{new}(K) = p_0(K) - \alpha \nabla_{H^1,p} G(c) \).
Get option prices \((v_1 \text{ and } v_2)\) for two consecutive expirations.

Find \(w_\lambda(K)\), \(w'_\lambda(K)\), \(\alpha(K)\), and \(\beta_\lambda(K)\).

Initialize \(p(K)\), \(q(K)\) and \(f(K)\) with \(c = (p_0, q_0, f_0)\).

Find \(w_\lambda,c\) and \(w'_\lambda,c\).

Find the \(L^1\) gradient in \(p\), \(\nabla_{L^1,p} G(c)\).

Find the Neuberger gradient in \(p\), \(\nabla_{H^1,p} G(c)\).

Find \(\alpha\) that gives lowest value of \(G\) by changing \(p(K)\) to \(p(K) - \alpha \nabla_{H^1,p} G(c)\).

\(\alpha = 0\) yes no

Update \(p(K)\) by \(p(K) - \alpha \nabla_{H^1,p} G(c)\).

Find \(w_\lambda,c\) and \(w'_\lambda,c\).

Find the \(L^1\) gradient in \(q\), \(\nabla_{L^1,q} G(c)\).

Find the Neuberger gradient in \(q\), \(\nabla_{H^1,q} G(c)\).

Find \(\alpha\) that gives lowest value of \(G\) by changing \(q(K)\) to \(q(K) - \alpha \nabla_{H^1,q} G(c)\).

\(\alpha = 0\) yes no

Update \(q(K)\) by \(q(K) - \alpha \nabla_{H^1,q} G(c)\).

Find \(w_\lambda,c\) and \(w'_\lambda,c\).

Find the \(L^2\) gradient in \(f\), \(\nabla_{L^2,f} G(c)\).

Find the Neuberger gradient in \(f\), \(\nabla_{H^2,f} G(c)\).

Find \(\alpha\) that gives lowest value of \(G\) by changing \(f(K)\) to \(f(K) - \alpha \nabla_{H^2,f} G(c)\).

\(\alpha = 0\) yes no

Update \(f(K)\) by \(f(K) - \alpha \nabla_{H^2,f} G(c)\).

Is there a descent in \(p,q\) or \(f\)?

\(\alpha = 0\) yes no

stop

yes no

no yes

Is there a descent in \(p,q\) or \(f\)?

\(\alpha = 0\) yes no

stop

yes no

Figure 1. Flowchart
6. Find \( G(p, q_0, f_0) \) using \( p_{\text{new}}(K) \) for \( p(K) \).
7. Find \( \alpha > 0 \) that gives the lowest value of \( G(p, q_0, f_0) \), \( \alpha_{\text{min}} \).
8. Set \( p(K) = p_{\text{new}}(K) \) where \( p_{\text{new}} \) is evaluated using \( \alpha_{\text{min}} \).

The descent in \( q \) and \( f \) is similar to that for descent in \( p \).

| Spot Price \((S_0)\) | $ 799.87 |
|----------------------|----------|
| Expiry \((T)\) in days | 7 | 14 | 21 | 27 | 35 | 62 | 153 |
| Strike \((K)\) in $ | 760 | 770 | 780 | 790 | 800 | 810 | 820 |
| \( v(K,T) \) | 1.45 | 2.10 | 3.00 | 5.15 | 8.70 | 14.50 | 21.70 |
| 760 | 2.95 | 4.20 | 5.85 | 8.55 | 12.60 | 18.00 | 24.25 |
| 770 | 4.20 | 5.90 | 8.05 | 11.00 | 15.30 | 20.25 | 26.70 |
| 780 | 5.35 | 7.15 | 9.70 | 12.80 | 17.05 | 22.35 | 29.10 |
| 790 | 6.45 | 8.50 | 11.30 | 14.55 | 18.75 | 24.50 | 30.00 |
| 800 | 10.70 | 13.25 | 16.35 | 20.20 | 24.15 | 29.55 | 35.30 |
| 810 | 26.15 | 29.60 | 33.55 | 37.65 | 42.60 | 47.55 | 53.00 |
| 820 | 7 | 14 | 21 | 27 | 35 | 62 | 153 |
| 830 | 21.70 | 24.25 | 26.70 | 29.10 | 33.90 | 41.00 | 50.35 |
| 840 | 30.90 | 32.20 | 35.90 | 39.05 | 41.00 | 48.60 | 51.25 |
| 850 | 40.10 | 41.00 | 43.30 | 44.45 | 48.60 | 55.00 | 52.05 |

**Figure 2.** GOOGLE put option data, 19th April, 2013

**Figure 3.** Recovered Google Volatility
After computing the corresponding gradients in \( q \) and \( f \), the \( q_{\text{new}}(K) \) and \( f_{\text{new}}(K) \) are given by

\[
q_{\text{new}}(K) = q(K) - \alpha \nabla_{H^1,q} G(c),
\]
\[
f_{\text{new}}(K) = f(K) - \alpha \nabla_{H^1,f} G(c).
\]

The specific order of descent is somewhat problem dependent, and different combinations of descents in \( p, q \) and \( f \) were tried to get the best minimization. Typically one needs more \( p \)-descent steps relative to \( q \) and \( f \) steps as the descent progresses.

5. **Results.** One of the most popular American options on US exchanges is the option on Google Inc (GOOG). Put option prices on GOOG were taken from the official website of the Chicago Board Options Exchange (CBOE), for several expirations on April 19, 2013; see Figure 2. The data includes only the near-the-money options as they are the most heavily traded. Furthermore, such data is invariably in the “do not exercise” region, likely due to the following in-built buyer/seller asymmetry: for \( S_0, t_0 \) in the “exercisable” region, if the stock goes down the put buyer can exercise immediately and make a profit, and if it goes up or sideways the buyer can wait, or even exercise, taking advantage of the trader’s “time value of money” (for a particular strike \( K \)) argument that becomes a consideration only when one is the put owner. As there appear to be no compensating advantages for the put seller, such sellers are likely to be in short supply. So for any given \( S_0, t_0 \), when \( K, T \) is near the boundary of the do-not-exercise region the volume drops to zero, and effectively there is no data available for these values of \( K, T \).

One can observe from the data in Figure 2 that, for fixed \( T \), at the bottom end of the strike range the prices tend to zero, and at the top end they tend to the linear function \( K - S_0 \). So, in our data sets we always, for fixed \( T \), linearly extrapolate to zero on the left end of the \( K \) interval, and linearly extrapolate to \( K - S_0 \) on the right end. This modification is important as it brings the “American” nature of the data into full focus. To recover the coefficient functions \( P(K), Q(K) \) and \( F(K) \) in (18) a computer code was written in the programming language C.

We have option prices for discrete strikes and discrete expirations \( T_i, 0 \leq i \leq n \). We generated the function \( v(K, T) \) by using cubic spline interpolation to form the functions \( v(K, T_i) \) for each \( i \), and then for each \( K \), and \( T \) in a fixed expiration interval \([T_{i-1}, T_i]\), we set

\[
v(K, T) = v(K, T_{i-1}) + \frac{T - T_{i-1}}{T_i - T_{i-1}} (v(K, T_i) - v(K, T_{i-1})).
\]

The work in [15] indicates that linear interpolation is quite acceptable here, as solution surfaces generated by such parabolic equations are similar to those generated by analytic functions in that the general shape is very much determined from a knowledge of relatively few points on the surface. Due to this “rigidity” other methods such as bicubic splines also work well. The derivative \( v_K(K, T) \) was found using central differences.

For 20 fixed values of \( \lambda \) the functions \( v(K, T) \) and \( v_K(K, T) \) were Laplace transformed using (16) to \( w_1(K) \) and \( w_1'(K) \) respectively. The functions \( p(K) \) and \( q(K) \) were initialized using (17) and (19), and the initial \( \sigma_i \) was chosen to be the implied volatility. The function \( f(K) \) was initialized to be the zero function. Observe that \( g(S_0, t_0; K, T) = 0 \) for all \( K, T \) in the original data range \( 760 \leq K \leq 850 \) and \( 7 \leq T \leq 153 \) in (10), which is consistent with an American option and a spot price \( S_0 \) in the “do not exercise” region above the free boundary [30, p. 108]. In later
iterations we enforced, after each iteration step, the condition $f(S_0, t_0; K, T) = 0$ for all $K, T$ in the data range $760 \leq K \leq 850$ and $7 \leq T \leq 153$, as we know beforehand that the option should not be exercised with some time value left in the option. We performed a series of descents in $p, q, f$, using the aforementioned steepest descent algorithm with Neuberger gradients, until the functional could not be minimized any further. Specifically, descents were performed in the order

$$q f f q f p p p p f$$

at which point the functional could not be lowered any further by descent in any of $p, q$ and $f$; the descent is illustrated graphically in Figure 4. The line minimization of $G(c)$ in $\alpha$ was done using the one-variable Brent minimization technique, by
adapting the `brent()` function code in Numerical Recipes in C [28]. To avoid possible catastrophic cancellation in the Simpson rule formula used in the calculation of the integrals in the formula (28) for the functional $G_{\lambda}$, we used the formula (29) instead.

As can be seen in the recovery of the function $g(S_0, t_0, K, T)$ for the maturity interval $14 \leq T \leq 21$ illustrated in Figures 4 and 5, the recovery settled down after about the fourth iteration.

After running the code we recovered the functions $P(K)$, $Q(K)$ and $F(K)$. From (23) we calculated the volatility $\sigma_i$ for each value of $i$, and created volatility surface $\sigma(K, T)$ from (15) using the curve fitting toolbox in MATLAB. From the volatility surface in Figure 3 one can observe the volatility smile and term structure. To check the accuracy of the method, we priced this option by substituting our recovered volatility $\sigma(K, T)$ into the homogeneous (so as to be consistent with pricing data from an American option in the “do not exercise” region as mentioned above) Dupire equation (8) and solving with MATLAB’s PDE Toolbox, using the known prices for boundary data. We compared these prices with the actual option prices at six different expirations as shown in Figure 6. As can be seen, the fit is excellent. Similar results were also observed using American put option price data from Facebook, Yahoo, and Baidu at various times.

6. **Conclusion.** We have shown that volatility can be recovered from published American option prices using steepest descent minimization of a convex functional. This provides a “market view” of future volatility which in principle can be used to price and trade options more effectively.

**REFERENCES**

[1] Y. Achdou, G. Indragoby and O. Pironneau, Volatility calibration with American options, *Methods Appl. Anal.*, 11 (2004), 533–556, [https://projecteuclid.org/euclid.maa/1144939946](https://projecteuclid.org/euclid.maa/1144939946).

[2] A. Alfonsi and B. Jourdain, Exact volatility calibration based on a Dupire-type Call-Put duality for perpetual American options, *Nonlinear Differ. Equ. Appl.*, 16 (2009), 523–554.

[3] F. Black and M. Scholes, The pricing of options and corporate liabilities, *Journal of Political Economy*, 81 (1973), 637–654.

[4] I. Bouchouev and V. Isakov, Uniqueness, stability and numerical methods for the inverse problem that arises in financial markets, *Inverse Problems*, 15 (1999), R95–R116.
Figure 6. Comparing option and recovered prices from the Dupire Equation.

[5] P. Carr and A. Hirs, Why be backward?, Risk, 16 (2003), 103–107.
[6] P. Carr and A. Hirs, Forward evolution equations for knock-out options, in Advances in Mathematical Finance, Appl. Numer. Harmon. Anal., Birkhäuser Boston, Boston, MA, 2007, 195–217.
[7] S. Crépey, Calibration of the local volatility in a trinomial tree using Tikhonov regularization, Inverse Problems, 19 (2003), 91–127.
[8] B. Dupire, Pricing with a smile, Risk, 7 (1994), 18–20.
[9] E. Fama, The behavior of stock-market prices, Journal of Business, 38 (1965), 34–105.
[10] P. Hartman, Ordinary Differential Equations, S. M. Hartman, Baltimore, Md., 1973.
[11] J. Huang and J.-S. Pang, A mathematical programming with equilibrium constraints approach to the implied volatility surface of American options, Journal of Computational Finance, 4 (2000), 21–56.
[12] V. A. Kholodnyi, A nonlinear partial differential equation for American options in the entire domain of the state variable, in Proceedings of the Second World Congress of Nonlinear Analysts, Part 8 (Athens, 1996), 30 (1997), 5059–5070.
[13] I. Knowles, Uniqueness for an elliptic inverse problem, SIAM J. Appl. Math., 59 (1999), 1356–1370.
[14] I. Knowles and M. A. LaRussa, Conditional well-posedness for an elliptic inverse problem, SIAM J. Appl. Math., 71 (2011), 952–971.
[15] I. Knowles and M. A. LaRussa, Lavrentiev’s theorem and error estimation in elliptic inverse problems, in Spectral Theory, Function Spaces and Inequalities, vol. 219 of Oper. Theory Adv. Appl., Birkhäuser/Springer Basel AG, Basel, 2012, 91–103.
[16] I. Knowles, T. Le and A. Yan, On the recovery of multiple flow parameters from transient head data, J. Comp. Appl. Math., 169 (2004), 1–15.
[17] I. Knowles, M. Teubner, A. Yan, P. Rasser and J. Lee, Inverse groundwater modelling in the Willunga Basin, South Australia, Hydrogeology Journal, 15 (2007), 1107–1118.
[18] I. Knowles and R. Wallace, A variational method for numerical differentiation, Numerische Mathematik, 70 (1995), 91–110.
[19] S. Kon, Models of stock returns - a comparison, Journal of Finance, 39 (1984), 147–165.
[20] G. W. Kutner, Determining the implied volatility for American options using the QAM, The Financial Review, 33 (1998), 119–130.
[21] O. A. Ladyzhenskaya and N. N. Ural’tzeva, Linear and Quasilinear Equations of Parabolic Type, Second edition, revised. Izdat. “Nauka”, Moscow, 1973.
[22] D. Madan and E. Seneta, The variance gamma model for share market returns, Journal of Business, 63 (1990), 511–524.
[23] B. Mandelbrot, The variation of certain speculative prices, Journal of Business, 36 (1963), 394–411.
[24] J. W. Neuberger, Sobolev Gradients and Differential Equations, vol. 1670 of Lecture Notes in Mathematics, 2nd edition, Springer-Verlag, Berlin, 2010.
[25] R. Officer, The distribution of stock returns, Journal of the American Statistical Association, 67 (1972), 807–812.
[26] P. Praetz, The distribution of share price changes, Journal of Business, 45 (1972), 49–55.
[27] S. Press, A compound events model for security prices, Journal of Business, 40 (1967), 317–335.
[28] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, Numerical Recipes in C, 2nd edition, Cambridge University Press, Cambridge, 1992, The art of scientific computing.
[29] A. Sen, An optimization approach to computing the implied volatility of American options, Optimization Online Digest.
[30] P. Wilmott, S. Howison and J. Dewynne, Option Pricing, Oxford Financial Press, Oxford, 1993.
[31] E. Zeidler, Nonlinear Functional Analysis and its Applications. III, Springer-Verlag, New York, 1985, Variational Methods and Optimization, Translated from the German by Leo F. Boron.

Received for publication September 2018.
E-mail address: iknowles@uab.edu
E-mail address: amahato@uab.edu