Determining Volatility via (Inverse) Optimization*

Thomas F. Coleman
Department of Computer Science
and
Cornell Theory Center
Cornell University
Ithaca, NY 14853

Optimization principles underpin much of computational finance. Historically this has been particularly true in the area known as portfolio analysis - the selection and modification of a collection of financial instruments such as stocks, bonds, and options. Indeed, Harry Markowitz was a co-recipient of the 1990 Nobel Prize in economics in recognition of his work involving novel application of optimization concepts to portfolio analysis.

The 1997 Nobel Prize in economics was won by Robert Merton and Myron Scholes for their role in the discovery and development of the most famous equation in finance, the Black-Scholes (B-S) equation [2, 14]. This work has, at first blush, little to do with optimization. Instead, it rests on notions from stochastic calculus, differential equations, and finance 'arbitrage' principles. However, it should be noted that application of the Black-Scholes equation (and, more interestingly, a generalized version) requires knowledge of a key parameter, the volatility. This is where numerical optimization can help.

*This expository article is targetted to a reader familiar with numerical methods; little mathematical finance background is required. A more complete and mathematical treatment is given in [3].

In this article we discuss the (generalized) Black-Scholes equation and sketch how inverse optimization problems can be formulated to yield smooth volatility surfaces. The latter is useful both in the accurate pricing of exotic options as well as computing sensitivities.

1 The Black-Scholes (B-S) Equation.

The Black-Scholes equation is the cornerstone of options pricing. Its' derivation is thoroughly covered in many introductory books on mathematical finance, e.g., [10, 11, 16]. One of the basic assumptions behind the Black-Scholes model is that the behaviour of the underlying S follows geometric Brownian motion,

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \qquad (1)$$

where the left-hand-side is an instantaneous relative change in the underlying (e.g., the stock value), μ is known as the 'drift' term (average rate of growth of the stock value) and μdt is the predictable component of the relative change in the stock price. The second term on the right-hand-side, σdW_t, corresponds to the random behaviour of the stock: σ is known as the *volatility* of the stock (volatility is the standard deviation, or square-root of the variance, of the returns of the underlying) and dW_t is random variable drawn from a normal distribution (the mean of dW_t is zero, the variance is dt).

Equation (1) has proven to be a very useful model of stock behaviour. Nevertheless, even with 'optimal' choices of σ and μ, it does not always capture reality. We discuss a generalization to overcome this gap in §2.

The question we are now faced with is how to value on option defined on an underlying S, e.g., stock, whose growth follows (1). The answer, under a number of assumptions including a no-arbitrage assumption, satisfies the Black-Scholes partial differential equation:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 V}{\partial s^2} + rs\frac{\partial V}{\partial s} - rV = 0, \quad (2)$$

where V is the value of the option (i.e., that which we are trying to determine), r is the risk-free interest rate. Detailed derivation of (2) is given in most texts on mathematical finance, e.g., [10, 11, 16].

Interestingly, the drift term μ does not appear in equation (2). Consequently, knowledge of μ is not required to evaluate V. (This is known as risk neutral valuation.) However, additional information beyond boundary conditions is required before (2) can be of practical value. In particular, a value for σ, the volatility, is required; σ is not directly observable. The methods for determining the volatility σ fall into two categories. First, σ can be estimated from first principles, i.e., based on the definition of standard deviation of the returns and using historical data. Typically this leads to 1-dimensional regression problem, e.g., [10, 16].

This first principles approach is easy to implement but has several unpleasant aspects. For example, there is the question of how far back in time to go and how the data should best be weighted. In addition, it is disconcerting that the regression solution for σ does not usually yield, through (2), the actual price for any known traded option.

A second, more common approach, is calculation of *implied volatility* (implied vol) to yield a value for σ . Implied volatility is determined by solving a simple 1-dimensional inverse problem involving a similar traded option (with known price) on the same underlying. Implied volatility is thus that value of σ that, when substituted into equation (2) along with appropriate boundary conditions, yields the known price of the corresponding option. Thus for each traded option there is a corresponding implied vol, and one such value can be used to determine the "fair" price of a new option on the same underlying.

Use of implied vol is very common in the trading world. Indeed, implied vol is often 'quoted' instead of option prices. Every financial engineering software package includes an implied vol computation. For example, in the MATLAB Financial Toolbox [13], a Black-Scholes implied volatility computation for a European call option is invoked by

$\sigma =$ **blsimpv(sc,K,r,T, call)**

where **sc** is the current price of the underlying, **K** is the strike price, **r** is the risk free interest rate, **T** is the time to maturity, and **call** is the value (or price) of the call option under consideration. Function **blsimpv** uses Newton's method.

Despite the popularity of the implied volatility concept, there are problems with its use. For example, given several differing implied vol computations on the same underlying, how should a value for σ be chosen to price an exotic option on the same underlying? a more pernicious problem has to do with hedging. We refer the reader to basic books on finance, e.g., [10, 11, 16], for discussions of hedging strategies. Here it suffices to say that hedging involves computing the sensitivity of V with respect to different parameters. The choice of σ can greatly affect the sensitivity calculation (and thus the hedging strategy) and so an arbitrary choice from a set of 'implied vols' can be misleading. Moreover, there is much evidence to indicate that σ varies with time and/or strike price (e.g., [6, 8, 9]): this suggests σ is better thought of as a function of (s, t), i.e., $\sigma = \sigma(s, t)$. A framework for this approach is discussed in §2.

2 Generalized Black-Scholes.

A reasonable and realistic alternative is to think of volatility as a surface, $\sigma = \sigma(s, t)$, rather than a constant. In particular, a more general model of the evolution of the stock price, replacing (2), is the 1-factor continuous diffusion approach:

$$\frac{dS_t}{S_t} = \mu(S_t, t)dt + \sigma(S_t, t)dW_t, \qquad (3)$$

where both $\sigma = \sigma(s, t)$ and $\mu = \mu(s, t)$ are continuous differentiable functions of the underlying s and t. Note that S_t is a stochastic variable and W_t is standard Brownian motion. The value of a European option where the underlying is defined by (3) satisfies the generalized Black-Scholes equation [14]:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma(s, t)^2 s^2 \frac{\partial^2 V}{\partial s^2} + rs\frac{\partial V}{\partial s} - rV = 0. \quad (4)$$

Unlike standard Black-Scholes, equation (4) does not enjoy an explicit solution; however, a discretized PDE approach can be used provided the surface $\sigma(s, t)$ is available for evaluation at all the grid points. Equation (4) obviously represents a potentially more powerful approach than the standard B-S equation which requires volatility to be a single number. But we are left with the question: how can the volatility surface be obtained?

Similar to the scalar case, an inverse (implied) point of view can be invoked. This approach uses current (or very recent) and trusted data, i.e., reliable prices on recently traded options on the underlying of question. A straightforward implementation to this inverse problem, assuming model (4), yields a large-scale optimization problem.

To see this suppose we have m data triplets, $(\bar{v}_j, T_j, K_j)$, corresponding to recently traded options on the same underlying: the option price is $\bar{v}_j$, T_j is the time to maturity, and K_j is the strike price. Discretize s and t consistent with the numerical procedure to solve (4), to yield an M-vector s and N-vector t. Surface $\sigma(s, t)$ is thus represented as an M-by-N matrix Σ (of unknowns). We are now faced with a vastly underdetermined problem, in general, since the number of data points, m, satisfies $m \ll M * N$. (A typical value is $m \approx 20$ whereas the product $M * N$ could easily be $100,000$ or more). To take up the slack, and introduce smoothness into σ, Osher and Lagnado [12] propose minimization over $\sigma(s, t)$, Σ after discretization, of the function

$$\sum_{j=1}^{m} (v_j(\sigma(s, t))) - \bar{v}_j)^2 + \lambda \|\nabla \sigma(s, t)\|_2, \quad (5)$$

where λ is a positive constant and $\|\cdot\|_2$ denotes the L^2 norm.

A difficulty with this approach, in addition to the delicate choice of the parameter λ, is the computational challenge. Problem (5) is a very large optimization problem ($M * N$ variables Σ). Moreover, the first term in (5) is very nonlinear and dense - density is due to the evaluation of the discretized aproximation of the PDE in (4) which depends on the entire surface $\sigma(s,t)$, i.e., each point of the matrix Σ is involved in the evaluation of (4) for each j, $j = 1, \cdots, m$.) Indeed, in light of the extreme computational expense, Osher and Lagnado compute only very approximate solutions to (5) using a steepest descent procedure. Unfortunately, approximate solutions can yield rough surfaces σ. Rough volatility surfaces, in turn, can cause severe pricing and, especially, hedging problems.

An alternative approach [3] yields a smaller more tractable optimization problem. The solution σ is smooth. The essential idea is to build in smoothness from the start: assume $\sigma(s,t)$ is a bi-cubic spline, e.g., [1, 7], defined on p knots. The knots are located in a regular way comensurate with the known data. In more detail, let the number of spline knots be $p \leq m$. Choose a set of fixed spline knots $\{(\bar{s}_j, \bar{t}_j)\}_{j=1}^{p}$. Given the spline knots with corresponding local volatility values $\bar{\sigma}_i \stackrel{\text{def}}{=} \sigma(\bar{s}_i, \bar{t}_i)$, an interpolating cubic spline $c(s,t)$ with a fixed end condition (e.g., the natural spline end condition) is uniquely defined by setting $c(\bar{s}_i, \bar{t}_i) = \bar{\sigma}_i, i = 1, \cdots, p$. The freedom in this problem is represented by the volatility values $\{\bar{\sigma}_i\}$ at the given knots $\{(\bar{s}_i, \bar{t}_i)\}$. If $\bar{\sigma}$ is a p-vector, $\bar{\sigma} = (\bar{\sigma}_1, \cdots, \bar{\sigma}_p)^T$, then we denote the corresponding interpolating spline with the specified end condition as $c(s,t;\bar{\sigma})$.

For $j = 1, \cdots, m$, let

$$v_j(c(s,t;\bar{\sigma})) \stackrel{\text{def}}{=} v(c(s,t;\bar{\sigma}), K_j, T_j).$$

To allow the possibility of incorporating additional a priori information, l and u are lower and upper bounds that can be imposed on the local volatilities at the knots. Thus, we define the *inverse spline local volatility approximation problem*: Given p spline knots, $(\bar{s}_1, \bar{t}_1) \cdots, (\bar{s}_p, \bar{t}_p)$, solve for the p-vector $\bar{\sigma}$

$$\min_{\bar{\sigma} \in \Re^p} \quad f(\bar{\sigma}) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{j=1}^{m} [v_j(c(s,t;\bar{\sigma})) - \bar{v}_j]^2 \quad (6)$$

$$\text{subject to} \quad l \leq \bar{\sigma} \leq u. \quad (7)$$

Note that (6) is a small optimization problem, typically with $m < p \approx 20$ variables, the solution has certain guaranteed smoothness properties (due to the use of the bi-cubic spline model), and, the given data will be usually be satisfied provided it is consistent.

3 Concluding Remarks.

The volatility surface produced by the bi-cubic spline optimization approach discussed above is visually smooth in the area of interest [3]. Indeed, given the location of the knots it can be argued that the computed surface σ is the smoothest surface consistent with the (discretized) model (4) and the given data. However, the real test of any volatility surface computation, in addition to its computational attractiveness, is its useability vis-à-vis pricing and, especially, hedging with the generalized Black-Scholes model (4). Hedging involves computing the sensitivity of the option price with respect to different parameters. Initial encouraging experiments are reported in [4]. (Efficient implementation of some sensitivity calculations involves applying either automatic differentiation or finite-differencing in a structured way [3, 5]).

We conclude with three remarks. First, while the bi-cubic spline optimization approach appears to produce a smooth, attractive, and useful volatility surface in the usual area of interest - in an $(s,t)-$region around known strike and maturities for current option data - the volatility surface becomes less reliable outside of this region. This is usually not a problem but can be troublesome when pricing (or hedging with) long-dated options. Pricing of long-dated options is an active area of investigation. Second, further work needs to be done on how to choose the number and location of the knots. Certainly fewer knots facilitate smoothness - our experience supports using fewer knots than the number of data points, but enough to force f in (6) to be close to zero. Finally, we expect that the bicubic optimization approach to this volatility surface problem can be applied to other inverse problems involving nonlinear, underdetermined systems both within and outside of finance.

4 Acknowledgement.

Thanks to my colleagues Yohan Kim, Yuying Li, and Arun Verma for their help and suggestions.

References

[1] G. Birkhoff and C. De Boor, *Piecewise polynomial interpolation and approximation*, in Approximation of Functions, H. L. Garabedian, ed., Elsevier (New York), 1965, pp. 164–190.

[2] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, Journal of Political Economy, 81 (1973), pp. 637–659.

[3] T. F. Coleman, Y. Li, and A. Verma, *Reconstructing the unknown local volatility function*, The Journal of Computational Finance, 2(3):77-102, 1999.

[4] T. F. Coleman, Y. Kim, Y. Li, and A. Verma, *Dynamic Hedging in a Volatile Market* Cornell Theory Center Technical Report, May, 1999.

[5] T. F. Coleman and A. Verma, *Structure and efficient Jacobian calculation*, in Computational Differentiation: Techniques, Applications, and Tools, M. Berz, C. Bischof, G. Corliss, and A. Griewank, eds., SIAM, Philadelphia, Penn., 1996, pp. 149–159.

[6] S. R. Das and R. K. Sundaram, *Of smiles and smirks: A term-structure perspective*, Tech. Report NBER working paper No. 5976, Department of Finace, Harvard Business School, Harvard University, Boston, MA 02163, 1997.

[7] C. de Boor, *Spline Toolbox for use with MATLAB*, The Math Works Inc., 1997.

[8] E. Derman and I. Kani, *Riding on a smile*, Risk, 7 (1994), pp. 32–39.

[9] B. Dupire, *Pricing with a smile*, Risk, 7 (1994), pp. 18–20.

[10] J. Hull, Options, Futures, and Other Derivatives, 3rd Edition. Prentice Hall, 1997.

[11] R. Jarrow and S. Turnbull, Derivative Securities, South-Western Pub., 1996

[12] R. Lagnado and S. Osher, *Reconciling differences*, Risk, 10 (1997), pp. 79–83.

[13] Users Guide, *Financial Toolbox*, The Mathworks, Inc., Version 2, 1999.

[14] R. Merton, *The theory of rational option pricing*, Bell Journal of Economics and Management Science, 4 (1973), pp. 141–183.

[15] R. Zvan, K. Vetzal, and P. Forsyth, *Swing high Swing low*, Risk Magazine, (1998), pp. 71–74.

[16] P. Wilmott, and S. Howison, and J Dewynne, The Mathematics of Financial Derivatives. Cambridge University Press, 1997.