Smooth and Bid-Offer Compliant Volatility Surfaces under General Dividend Streams

Olivier Bachem  Gabriel Drimus  Walter Farkas

First version: October 2012
Current version: July 2013

This research has been carried out within the NCCR FINRISK project on “Mathematical Methods in Financial Risk Management”
Smooth and Bid-Offer Compliant Volatility Surfaces under General Dividend Streams

Olivier Bachem\textsuperscript{2} Gabriel Drimus\textsuperscript{1}\textdagger  Walter Farkas\textsuperscript{1,2}

First version: October, 2012
This version: July, 2013

Abstract

Given bid-offer quotes for a set of listed vanilla options, a fundamental need of option market makers is to interpolate and extrapolate the available quotes to a full arbitrage-free surface. We propose a methodology which directly controls the trade-off between smoothness and bid-offer compliance of the resulting volatility surface. Unlike previous literature, the method applies simultaneously to all listed maturities and aims to smooth the implied risk neutral densities. Additionally, we consider asset dynamics which allow for general dividend streams – continuous, discrete yield and discrete cash – a modeling aspect of key importance in option markets.

KEYWORDS: implied volatility surface, risk neutral density, discrete dividends.
JEL: C63, G13.

1 Introduction

One of the most fundamental — and surprisingly non-trivial — problems in quantitative finance is the construction of an arbitrage free implied volatility surface consistent with the observed market quotes of listed vanilla options. All aspects of a derivatives business from pricing, trading to risk-management depend on how this fundamental task is implemented.

Two different approaches can be followed in constructing an implied volatility surface. In the parametric approach, one seeks to prescribe an

\textsuperscript{1}Department of Banking and Finance, University of Zürich, Plattenstrasse 14, CH-8032 Zürich, Switzerland. Email: gabriel.drimus@bf.uZH.ch and walter.farkas@bf.uZH.ch.
\textsuperscript{2}Department of Mathematics, ETH Zürich, Rämistrasse 101, CH-8092 Zürich, Switzerland. Email: bachemo@student.ethz.ch.
\textdagger National Center for Competence in Research (NCCR) FINRISK funding, through Project D1 ”Mathematical Methods in Financial Risk Management” is gratefully acknowledged.
analytic formula for the implied volatility curve at each maturity. These parametric curves are usually inspired by the limiting behavior of implied volatility in certain stochastic volatility models. Two of the most popular are the SABR-parametrization, based on the short-term asymptotics of the SABR model in Hagan et al. (2002) and the SVI-parametrization of Gatheral (2006), based on the long-term asymptotics of the Heston (1993) model, as recently shown in Gatheral, Jacquier (2011). The non-parametric (or semi-parametric) approach is usually applied in terms of absolute call prices and seeks to impose no (or very little) a priori structure, except for the no-arbitrage conditions. We primarily note here the spline-based method in Fengler (2009) and the implicit finite difference method in Andreasen, Huge (2010). The approach followed in this paper belongs to this latter class of methods.

There exists an inherent trade-off between the smoothness of the implied volatility surface – or of the risk-neutral densities – and the quality of fit to observed market quotes. While a parametric approach may yield very smooth implied volatilities, it may be difficult to control how far outside the bid-offer quotes the interpolation departs. On the other hand, a non-parametric approach may lead to very irregular implied risk-neutral densities. Neither of these issues is desirable. Additionally, current methods usually apply sequentially, either starting from short maturities and advancing towards distant maturities or in reverse. This implies that one gives priority to the fit at one maturity over the fit at the next maturity, in order to prevent calendar spread arbitrage.

The methodology proposed in this paper differs from existing methods in several ways. Firstly, we directly control the trade-off between smoothness and the bid-offer constraints of the original market quotes. In particular, our method allows us to impose that the resulting volatility surface lie within the bid-offer quotes of all, or of a selected subset e.g. close to the money, listed vanilla options. Secondly, the method applies globally to all listed maturities and hence does not require to choose a (possibly arbitrary) order in which maturities are fitted. More recently, a full surface estimation approach is taken in Fengler, Lin (2011) and Glaser, Heider (2012). Thirdly, we smooth directly the risk neutral densities rather than the implied volatilities. This is a key feature of our approach since a seemingly smooth implied volatility curve can hide an unreasonable risk-neutral density; for a convincing illustration we refer to Rebonato, Cardoso (2004). Finally, we work throughout under asset dynamics which assume general dividend streams: continuous, discrete yield and discrete cash. In practice, the modeling of discrete cash dividends is particularly important as they have a direct impact on the implied volatility surface, see e.g. Bos et al. (2004).

The rest of the paper is organized as follows. The next section introduces the asset dynamics and presents the no-arbitrage conditions for a surface of call option prices under general dividend streams. Section three is the core
section which develops the volatility surface construction algorithm. Section four applies the method to option data on dividend paying stocks. The final section summarizes the main results.

2 Asset dynamics and no-arbitrage conditions

2.1 General dividend streams

We begin by prescribing the underlying price dynamics. Historically, several approaches have been proposed to incorporate discrete cash dividends. Comprehensive overviews can be found in Bos et al. (2004), Vellekoop, Nieuwenhuis (2006) and Klassen (2009). As discussed in these references, all approaches have certain advantages and disadvantages but the approach considered most desirable is the so called 'piecewise lognormal-model' in which the stock price evolves according to a Geometric Brownian Motion between dividend dates and exhibits downward jumps equal to the dividend amounts at the ex-dividend dates.

As we intend to work in a non-lognormal world, we shall allow the volatility to be an arbitrary process and assume the following dynamics for the asset price $S_t$ under the risk-neutral measure $Q$:

$$dS_t = (r(t) - q(t)) S_t dt + \sigma_t S_t dW_t - \sum_{i \in D} (d_i + \delta_i S_{t_i}) \cdot 1_{\{t = t_i\}}$$

where $W_t$ is a $Q$-Brownian Motion, the interest rate $r(t)$ and the continuous dividend yield $q(t)$ are both deterministic and time-dependent; the volatility process $\sigma_t$ is only assumed to be measurable and integrable. We let the set $D$ denote the index set of all discrete dividend payments $(d_i)_{i \in D}$. Each discrete dividend payment $d_i$ is given by the triplet $\{t_d, d_i, \delta_i\}$ where $t_d$ denotes the ex-dividend date, $d_i$ the discrete cash payment and $\delta_i$ the discrete proportional payment.

In practical applications, near-term dividends (e.g. up to one year) are assumed to be all cash and longer-term dividends (e.g. beyond 3 years) to be all proportional, with a mix of cash and proportional dividends in between. We note that in the stock price has non-zero probability of becoming negative on an ex-dividend date. However, as noted in Bos, Vandermark (2002) and Klassen (2009) this usually does not pose significant issues in practice provided the cash dividend stream is properly truncated beyond a certain maturity.

The interest rate $r(t)$ can be assumed to be a step function and easily calibrated to match an interest rate term structure, for example a LIBOR curve. The continuous dividend yield $q(t)$ can also be used to account for any borrow fees associated with the stock. For later convenience, we further define the zero bond factor $B(t_1, t_2)$, the continuous dividend factor $D(t_1, t_2)$
and the discrete yield dividend factor \( p_d(t_1, t_2) \):

\[
B(t_1, t_2) = e^\int_{t_1}^{t_2} r(t) dt \\
D(t_1, t_2) = e^\int_{t_1}^{t_2} q(t) dt \\
p_d(t_1, t_2) = \prod_{t_1 < t_i \leq t_2} \frac{1}{1 - \delta_i}.
\]

### 2.2 Static no-arbitrage conditions

We next identify the static no-arbitrage conditions that must be satisfied by a surface of European call prices. The conditions along the strike dimension are identical to the case when the asset pays no discrete dividends and can be found in Breeden, Litzenberger (1978) or more recently in Fengler (2009). The conditions along the maturity dimension will require more care due to the impact of discrete dividends.

Let \( F(T) \) denote the forward price of maturity \( T \) and \( C(K, T) \) denote the surface of European call prices across strikes \( K \) and maturities \( T \). All call prices must satisfy the trivial bounds

\[
0 \leq \max \left[ 0, \frac{1}{B(0, T)} (F(T) - K) \right] \leq C(K, T) \leq \frac{F(T)}{B(0, T)}.
\]

Additionally, for all \( K \) and \( T \), we have

\[
C(K, T) = \frac{1}{B(0, T)} \mathbb{E}_Q \left[ \max(S_T - K, 0) \right] \\
= \frac{1}{B(0, T)} \int_0^\infty \max(S - K, 0) \phi(S, T) dS
\]

where \( \phi(\cdot, T) \) denotes the marginal risk-neutral density of \( S_T \). Upon differentiation with respect to \( K \) we obtain

\[
C_K(K, T) = -\frac{1}{B(0, T)} \int_K^\infty \phi(S, T) dS
\]

which allows us to conclude

\[
-\frac{1}{B(0, T)} \leq C_K(K, T) \leq 0.
\]

Differentiating again with respect to \( K \), we get

\[
C_{KK}(K, T) = \frac{1}{B(0, T)} \phi(K, T) \geq 0. \tag{2}
\]

As first obtained by Breeden, Litzenberger (1978), the second derivative of the undiscounted call price surface with respect to \( K \) is equal to the risk
neutral density. In practice, this condition is also known as the ‘no negative butterflies’ condition, see Gatheral (2006).

For the no-arbitrage conditions in time we recall the result from Fengler (2009) stating that the dividend-corrected call price surface needs to be non-decreasing on the forward moneyness grid

\[ C(K, T_2) \geq \frac{1}{D(T_1, T_2)} C \left( \frac{D(T_1, T_2)}{B(T_1, T_2)} K, T_1 \right) \]

for all \( K \) and \( T_1 \leq T_2 \).

By an identical argument it follows that the condition above also has to hold under general dividend streams for any \( T_1 \) and \( T_2 \) such that there is no ex-dividend date in \([T_1, T_2]\). If the maturity interval \([T_1, T_2]\) contains at least one ex-dividend date, the calendar no-arbitrage condition becomes

\[ C(K, T_2) \geq \frac{1}{D(T_1, T_2) \cdot p_d(T_1, T_2)} \cdot C \left( K \cdot \frac{D(T_1, T_2) \cdot p_d(T_1, T_2)}{B(T_1, T_2)} + \sum_{T_1 < t^d_i \leq T_2} d_i \cdot \frac{D(T_1, t^d_i) \cdot p_d(T_1, t^d_i)}{B(T_1, t^d_i)} \right) \]

which follows most conveniently from our construction in the next section; the derivation steps are included in Appendix 6.2. Additionally, at an ex-dividend date \( t^d_i \) we have the extra no-arbitrage condition, which holds for all \( K \geq 0 \):

\[ C(K, t^d_i) = \frac{1}{B(0, t^d_i)} \mathbb{E}_Q \left[ (S_{t_i} - K)^+ \right] \]

\[ = \frac{1}{B(0, t^d_i)} \mathbb{E}_Q \left[ (S_{t_i} - (1 - \delta_i) - d_i - K)^+ \right] \]

\[ = (1 - \delta_i) C \left( \frac{K + d_i}{1 - \delta_i}, t^d_i \right). \]  

(4)

In particular, we remark that, as a function of maturity \( T \), the call price surface must exhibit a jump across the ex-dividend date. Differentiating (4) twice with respect to \( K \) and using equation (2), we obtain that also the risk-neutral density will exhibit a jump across the ex-dividend date:

\[ \phi(K, t^d_i) = \frac{1}{1 - \delta_i} \cdot \phi \left( \frac{K + d_i}{1 - \delta_i}, t^d_i \right). \]  

(5)

### 2.3 A transformation of the option quotes

These jumps at the ex-dividend dates make the construction of the call price surface difficult under general dividend streams. Therefore, we propose to perform the construction on a suitable transformation of the call price surface, with the aim to simplify the no-arbitrage conditions. An intuitive
way to introduce this transformation is by considering a new asset \( A_t \) which
starts out at \( A_0 = S_0 \) and collects all dividends as follows: continuous
dividends are reinvested in the stock and discrete dividends are partially
invested in the stock and partially invested in the bank account.

We now determine the value of the asset \( A_t \) at some arbitrary time
\( t \geq 0 \). Paying attention to all the dividend flows, the result can be obtained
by induction. As a first step, note that for any \( t \in [0, t^d_1) \), we have
\( A_t = D(0, t^d_1) \cdot S_t \) since our initial holding of one share has grown to
\( D(0, t^d_1) \) shares by reinvesting the continuous dividends. At the first discrete dividend date
\( t = t^d_1 \), we have:
\[
A_{t^d_1} = D(0, t^d_1) \cdot S_{t^d_1} + D(0, t^d_1) \cdot \left( S_{t^d_1} \cdot \delta_1 + d_1 \right)
\]
and using the identity \( S_{t^d_1} = S_{t^d_1} \cdot (1 - \delta_1) - d_1 \), we obtain after rearranging
terms
\[
A_{t^d_1} = D(0, t^d_1) \cdot \frac{1}{1 - \delta_1} \cdot S_{t^d_1} + D(0, t^d_1) \cdot \frac{1}{1 - \delta_1} \cdot d_1. \tag{6}
\]
Recall that we initially hold \( D(0, t^d_1) \) shares at time \( t^d_1 \). However, the amount
in \( [0, t^d_1) \) shows that we can purchase \( \frac{\delta_1}{1 - \delta_1} \) extra shares for each share that we
own and set aside an amount \( D(0, t^d_1) \cdot \frac{1}{1 - \delta_1} \cdot d_1 \) in the bank account. After
this readjustment, for any \( t \in [t^d_1, t^d_2) \), we have
\[
A_t = D(0, t^d_1) \cdot \frac{1}{1 - \delta_1} \cdot S_t + D(0, t^d_1) \cdot \frac{1}{1 - \delta_1} \cdot B(t^d_1, t) \cdot d_1.
\]
By a similar reasoning, we obtain the general result which holds for any
\( t \geq 0 \):
\[
A_t = D(0, t) \cdot p_d(0, t) \cdot S_t + \sum_{t^d_i \leq t} D(0, t^d_i) \cdot p_d(0, t^d_i) \cdot B(t^d_i, t) \cdot d_i \tag{7}
\]
where \( p_d(0, t) \) is the proportional dividend factor defined in the previous
section. We remark that the asset \( A_t \) is a non-dividend paying asset and, by
no-arbitrage, it must grow at the risk-free interest rate \( r(t) \) under the risk-
neutral measure \( Q \). In particular \( \tilde{A}_t = A_t / B(0, t) \), the value of \( A_t \) expressed
in units of the bank account, is a \( Q \)-martingale. Therefore, if we let
\[
\tilde{C} \left( \tilde{K}, T \right) = \mathbb{E}_Q \left( \tilde{A}_T - \tilde{K} \right)_+
\]
the static no-arbitrage conditions for $\tilde{C}(\cdot, \cdot)$ simplify to

$$0 \leq (S_0 - \bar{K})_+ \leq \tilde{C}(\bar{K}, T) \leq S_0,$$

$$-1 \leq \frac{\partial \tilde{C}}{\partial \bar{K}} \leq 0,$$

$$\frac{\partial^2 \tilde{C}}{\partial \bar{K}^2} \geq 0$$

and

$$\tilde{C}(\bar{K}, T_2) \geq \tilde{C}(\bar{K}, T_1)$$

for any $T_1 \leq T_2$.

where we have used the fact that $\tilde{A}_0 = A_0 = S_0$.

We can obtain a relationship between options on $S_t$ and options on $\tilde{A}_t$, as follows:

$$D(0, T) \cdot p_d(T) \cdot C(K, T) = D(0, T) \cdot p_d(T) \cdot \frac{1}{B(0, T)} \mathbb{E}_Q [S_T - K]_+$$

$$= \frac{1}{B(0, T)} \cdot \mathbb{E}_Q [A_T - D(0, T) \cdot p_d(T) \cdot K$$

$$- \sum_{t^d_i \leq T} D(0, t^d_i) \cdot p_d(t^d_i) \cdot B(t^d_i, T) \cdot d_i]_+$$

$$= \tilde{C}(\bar{K}(K, T), T)$$

(11)
where make use of relation (7) for \( A_T \) and where we define the strike transformation
\[
\tilde{K}(K, T) := \frac{D(0, T) \cdot p_d(T) \cdot K}{B(0, T)} + \sum_{t^d_i < T} \frac{D(0, t^d_i) \cdot p_d(t^d_i)}{B(0, t^d_i)} \cdot d_i. \tag{12}
\]

Therefore, by equations (11) and (12), we can transform any set of European option quotes on \( S_t \) to European option quotes on \( \tilde{A}_t \). This will be the first step we take in our option surface construction algorithm. We can then interpolate and extrapolate the quotes to a wide surface \( \tilde{C}(\tilde{K}, \tilde{T}) \) by observing the simpler no arbitrage conditions in (8)-(10). Finally, we transform back to obtain the desired option surface on \( S_t \). We show in Figure 1 an example of how the adjusted strikes \( \tilde{K} \) compare to the original strikes \( K \) across the listed maturities on July 4th 2012 for the Novartis stock (ticker: NOVN).

3 Construction of the Volatility Surface

Our volatility surface construction methodology is based on modifying – and subsequently combining – two different and seemingly unrelated methods from the literature, namely, the spline based method in Fengler (2009) and the implicit finite difference method in Andreasen, Huge (2010). Therefore, the method will consist of two major steps: an initial spline based step which builds on Fengler (2009) and a subsequent application of the Andreasen, Huge (2010) idea in a non-standard way, by an explicit solution which skips its optimization step.

Let the set \( \mathcal{O} \) denote the index set of all option quotes available. Then an option quote \( o_i \) is given by the 5-tuple \((T_i, K_i, \tilde{K}_i, P_{\text{bid}}^i, P_{\text{ask}}^i)\) for \( i \in \mathcal{O} \) where \( \tilde{K}_i \) is calculated using equation (12). We work on a rectangular grid \( \{u_0, u_1, \ldots, u_n\} \times \{T_1, T_2, \ldots, T_m\} \), with the following properties:

(i) The space grid points \( \{u_i\} \) will usually be chosen uniformly spaced in practical implementations. However, this is not a requirement in the algorithm described below and hence we treat the general case. Additionally, the set \( \{u_i\} \) does not depend on the set of strikes \( \tilde{K}_i \). As will be shown next, in order to compute the call price at some arbitrary strike \( \tilde{K} \), we simply use the cubic spline parametrization. The only constraint is that \( u_0 \) should be smaller than the smallest strike \( \tilde{K}_i \) and \( u_n \) should be larger than the largest strike \( \tilde{K}_i \).

(ii) The set \( \{T_j\} \) should include the listed maturities, but it may also include other special maturities e.g. ex-dividend dates. In our numerical examples, in addition to the listed maturities, the set \( \{T_j\} \) also includes the ex-dividend dates.
On the rectangular grid \( \{u_i\} \times \{T_j\} \), we apply – as in Fengler (2009) – a "pre-smoothing" step which seeks to fill initial estimates for the call price surface. These initial estimates need not satisfy any smoothness or arbitrage conditions, as these will be imposed later. In our implementation, to obtain the initial estimates we use one of Matlab’s bi-dimensional interpolation functions\(^1\) applied to the mid-implied volatilities of the market quotes. For the grid points \( \{u_i\} \) which lie outside the listed strike range for a given maturity, we apply a total variance extrapolation linear in log-forward moneyness, consistent with the slope coefficients in Lee (2004).

3.1 Step I: The spline based step

Our approach builds on the key idea in Fengler (2009), but is different in a number of ways: (i) we solve a global quadratic problem which applies simultaneously to all listed maturities, (ii) we seek to smooth the risk neutral densities as opposed to the call price surface and (iii) we apply bid-offer constraints to a subset (or possibly, all) of the resulting option prices. Additionally, the quadratic problem is applied to the transformed call prices \( \tilde{C}(\tilde{K}, T) \) not to the original prices \( C(K, T) \) thus allowing to apply simpler no-arbitrage constraints. In particular, we are able to incorporate discrete cash dividends, an important aspect in industry applications.

In formulating the quadratic optimization problem, we use the value - second derivative representation of a cubic spline as in Green, Silverman (1994); we remark that, unlike the treatment in Fengler (2009), we shall not restrict attention to natural cubic splines.

For each maturity \( t_j > 0 \), we approximate the transformed call price function \( \tilde{C}(\tilde{K}, T) \) by a cubic spline \( g_j(K) \) which has knots at the space grid points \( u_0 < u_1 < \ldots < u_n \). The cubic spline is represented by the value vector \( g^j = (g^j_0, g^j_1, \ldots, g^j_n) \) and the second-derivative vector \( \gamma^j = (\gamma^j_0, \gamma^j_1, \ldots, \gamma^j_n) \).

\(^1\)Specifically the Matlab function ”griddata” which is based on Delaunay triangulation and cubic splines.
following \((n-1) \times (n+1)\) matrices \(Q\) and \(R\):

\[
Q = \begin{pmatrix}
\frac{1}{h_0} & -\left(\frac{1}{h_0} + \frac{1}{h_1}\right) & 0 & \cdots & 0 \\
\frac{1}{h_1} & -\left(\frac{1}{h_1} + \frac{1}{h_2}\right) & \frac{1}{h_2} & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots \\
\frac{1}{h_{n-2}} & \cdots & \frac{1}{h_{n-2} + \frac{1}{h_{n-1}}} & 0 & 0 \\
\frac{1}{h_{n-1}} & 0 & \cdots & \cdots & \cdots
\end{pmatrix}
\]

\[
R = \begin{pmatrix}
\frac{h_0}{6} & \frac{h_0 + h_1}{6} & \frac{h_1}{6} & \frac{h_1 + h_2}{6} & \frac{h_2}{6} & 0 & \cdots & 0 \\
\frac{h_0}{6} & \frac{h_0 + h_1}{6} & \frac{h_1}{6} & \frac{h_1 + h_2}{6} & \frac{h_2}{6} & 0 & \cdots & 0 \\
\frac{h_0}{6} & \frac{h_0 + h_1}{6} & \frac{h_1}{6} & \frac{h_1 + h_2}{6} & \frac{h_2}{6} & 0 & \cdots & 0 \\
\frac{h_0}{6} & \frac{h_0 + h_1}{6} & \frac{h_1}{6} & \frac{h_1 + h_2}{6} & \frac{h_2}{6} & 0 & \cdots & 0 \\
\frac{h_0}{6} & \frac{h_0 + h_1}{6} & \frac{h_1}{6} & \frac{h_1 + h_2}{6} & \frac{h_2}{6} & 0 & \cdots & 0 \\
\frac{h_{n-2}}{6} & \frac{h_{n-2} + h_{n-1}}{6} & \frac{h_{n-1}}{6} & 0 & 0 & \cdots & 0 \\
\frac{h_{n-1}}{6} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{pmatrix}
\]

As shown in Appendix 6.3, the pair of vectors \((g^j, \gamma^j)\) represents a cubic spline if and only if

\[
Qg^j = R\gamma^j \quad \text{for } j = 1, 2, \ldots, m.
\]

(13)

One of the key features of our method is that the algorithm seeks to smooth the risk neutral density instead of the call price function. Note that the vector \(\gamma\) contains the values of the risk-neutral density evaluated at the grid points. We shall seek to minimize the "roughness of the vector \(\gamma\)" which we propose to measure by applying a second order finite difference operator on the vector \(\gamma\) itself, as the penalty. Intuitively, this is equivalent to minimizing an approximate derivative of the risk neutral density. The key aspect which allows us to formulate the penalty in this form is that the resulting optimization problem remains convex in the combined vector \((g, \gamma)\) and hence possesses a unique solution.

We define the second order finite difference operator \(S\) which uses the forward finite difference operator for \(u_0\), the central finite difference operator for \(u_i\) with \(i = 1, \ldots, n-1\) and the backward finite difference operator for \(u_n\):

\[
S = \begin{pmatrix}
\frac{2}{h_0^2} & \frac{2}{h_0 h_1} & \frac{2}{h_0^2} & \cdots & \frac{2}{h_0^2} & \frac{2}{h_0 h_1} & \frac{2}{h_0 h_2} & \frac{2}{h_0^2} & \cdots & \frac{2}{h_0^2} \\
\frac{2}{h_0 h_1} & \frac{2}{h_0 h_1} & \frac{2}{h_1^2} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{2}{h_0 h_1} & \frac{2}{h_0 h_1} & \frac{2}{h_1^2} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{2}{h_0 h_2} & \frac{2}{h_0 h_2} & \frac{2}{h_1 h_2} & \frac{2}{h_1^2} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{2}{h_0 h_1} & \frac{2}{h_0 h_1} & \frac{2}{h_1^2} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{2}{h_0 h_2} & \frac{2}{h_0 h_2} & \frac{2}{h_1 h_2} & \frac{2}{h_1^2} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{2}{h_1 h_2} & \frac{2}{h_1 h_2} & \frac{2}{h_1^2} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{2}{h_{n-2} h_{n-1}} & \frac{2}{h_{n-2} h_{n-1}} & \frac{2}{h_{n-2} h_{n-1}} & \frac{2}{h_{n-2} h_{n-1}} & \frac{2}{h_{n-2} h_{n-1}} & \frac{2}{h_{n-2} h_{n-1}} & \frac{2}{h_{n-2} h_{n-1}} & \frac{2}{h_{n-2} h_{n-1}} & \frac{2}{h_{n-2} h_{n-1}} & \frac{2}{h_{n-2} h_{n-1}}
\end{pmatrix}
\]

Denoting the pre-smoothing estimates by \(\tilde{g}^j = (\tilde{g}_0^j, \tilde{g}_1^j, \ldots, \tilde{g}_n^j)\) for all maturities \(j = 1, 2, \ldots, m\), we set up the target function to be minimized as
follows:

\[
f(g, \gamma) = \frac{1}{\bar{w}} \sum_{j=1}^{m} [(g^j - \tilde{g}^j)'W^j(g^j - \tilde{g}^j)] + C\lambda \sum_{j=1}^{m} [(\gamma^j)'S'S\gamma^j]
\]

where \(\lambda\) is the smoothing parameter, \(C\) is a normalization constant, \(W^j\) is a diagonal matrix with non-negative weights and \(\bar{w}\) is the average weight across all maturities. We remark that our objective function spans all maturities simultaneously, where for each fixed maturity, the term in the left sum is a measure of the fitting error and the term in the right sum is a measure of the roughness of the marginal risk-neutral density. An interesting observation is that our algorithm could accommodate different values of \(\lambda\) for different maturities. Therefore one could, at least in theory, impose different degrees of smoothness at different maturities. Of course, picking maturity-dependent smoothing parameters may prove to be rather challenging in practice and we have not experimented it in our numerical examples.

As the strike grid in practice will usually be much wider than the range of available market quotes, we indicated in the previous section that the pre-smoothing estimates \(\tilde{g}^j\) at distant strikes will be obtained by an extrapolation procedure consistent with the asymptotic slope bounds derived in Lee (2004). These serve as the "target" in the optimization problem above and the user can choose the weight vector \(W^j\) to emphasize certain segments of the smile curve, e.g. the put wing, if necessary. We remark that, since we always work on a finite grid, the construction method will guarantee that the volatility surface remains arbitrage free independent of the chosen extrapolation at distant strikes.

We next consider the no-arbitrage conditions required for the call price surface to be arbitrage-free. Imposing bid and offer constraints requires for all \(i \in \mathcal{O}\) (or, alternatively, a selected subset of liquid options in \(\mathcal{O}\)):

\[
p^\text{bid}_i \leq g_j(\tilde{K}_i) \leq p^\text{ask}_i \quad \text{where } i \text{ is such that } T_i = t_j
\]

The value of the spline at an option quote \(i \in \mathcal{O}\) is given by

\[
g_j(\tilde{K}_i) = \left[ \left( 1 - \frac{x}{h_k} \right) g_k^j + \frac{x}{h_k} g_{k+1}^j + \left( \frac{x^2}{2} - \frac{h_k x}{3} - \frac{x^3}{6 h_k} \right) \gamma_k^j + \left( \frac{x^3}{6 h_k} - \frac{h_k x}{6} \right) \gamma_{k+1}^j \right] \bigg|_{x=\tilde{K}_i-u_k} \tag{14}
\]

where \(k\) is such that \(u_k \leq \tilde{K}_i < u_{k+1}\). Denoting the vector of bid quotes at maturity \(t_j\) by \(p^j_{\text{bid}}\) and the vector of ask quotes by \(p^j_{\text{ask}}\), we notice that we can identify matrices \(L^j_{g}\) and \(L^j_{\gamma}\) such that the bid-offer conditions may be expressed linearly as

\[
p^j_{\text{bid}} \leq L^j_{g} g^j + L^j_{\gamma} \gamma^j \leq p^j_{\text{ask}} \quad \text{for } j = 1, 2, \ldots, m.
\]
For completeness, the construction of the matrices \( L^j_g \) and \( L^j_\gamma \) can be found in the Appendix.

The call price surface has to be convex in \( \tilde{K} \), which translates to the simple constraint

\[
\gamma^j \geq 0 \quad \text{for } j = 1, 2, \ldots, m.
\]

Due to convexity, it suffices to impose the no-arbitrage condition \( [9] \) on the first derivative with respect to the adjusted strike only on the first and last knot. Therefore, for \( j = 1, 2, \ldots, m \) we require

\[
g'_j(u_0) = \frac{g^j_1 - g^j_0}{h_0} - \frac{h_0}{6} (2\gamma^j_0 + \gamma^j_1) \geq -1
\]

and

\[
g'_j(u_n) = \frac{g^j_n - g^j_{n-1}}{h_{n-1}} + \frac{h_{n-1}}{6} (\gamma^j_{n-1} + 2\gamma^j_n) \leq 0.
\]

Additionally, appealing also to monotonicity, we notice that the basic bounds in \( [8] \) need only be imposed at the boundaries, for \( j = 1, 2, \ldots, m \)

\[
g^j_n \geq 0
\]

\[
g^j_0 \geq (S_0 - u_0)_+
\]

\[
g^j_0 \leq S_0.
\]

Finally, setting \( g^0_i = \max(S_0 - u_i, 0) \) for \( i = 0, 1, \ldots, n \), the no-arbitrage conditions in time are simply given by

\[
g^j \geq g^{j-1} \quad \text{for } j = 1, 2, \ldots, m.
\]

Assembling all pieces together, we arrive at the full quadratic minimiza-
The minimization problem presented above can easily be cast in the following standardized form for quadratic programs so that it can be solved by standard algorithms

\[
\min_{(g^j), \gamma^j} \left( \frac{1}{w} \sum_{j=1}^{m} [(g^j - \tilde{g}^j)'W(g^j - \tilde{g}^j)] + C\lambda \sum_{j=1}^{m} [(\gamma^j)'S'S\gamma^j] \right)
\]

subject to

\[
\begin{align*}
Qg^j &= R\gamma^j \\
\gamma &\geq 0 \\
g^j_1 - g^j_0 &= \frac{h_0}{6} (2\gamma^j_0 + \gamma^j_1) \geq -1 \\
g^j_n - g^j_{n-1} &= \frac{h_{n-1}}{6} (\gamma^j_{n-1} + 2\gamma^j_n) \leq 0 \\
g^j_n &\geq 0 \\
g^j_0 &\leq S_0 \\
g^j_0 &\geq (S_0 - u_0)_+ \\
g^j &\geq g^{j-1} \\
L_gg^j + L_\gamma^j &\leq p^j_{ask} \\
L_gg^j + L_\gamma^j &\geq p^j_{bid}
\end{align*}
\]

for \( j = 1, 2, \ldots, m \)

The minimization problem presented above can easily be cast in the following standardized form for quadratic programs so that it can be solved by standard algorithms

\[
\min_{x} \frac{1}{2} x' H x + f' x
\]

subject to

\[
Ax \leq b \\
Ex = d.
\]

Since both \( W \) and \( S'S \), and thus also \( H \), are positive semi definite matrices, the function \( \frac{1}{2} x' H x + f' x \) is convex and we may hence use the interior point method to find a solution relatively fast.

As an implementation detail, in order to make the smoothing parameter \( \lambda \) comparable across data sets, we further normalize all option values and the adjusted strike grid by dividing with the current stock price \( S_0 \).

We also found that setting the constant \( C \) to \( 10^{-11} \) normalizes the range of acceptable \( \lambda \) to approximately \( 10^{-3} \) to 10. The weights \( W \) allow to assign different credibilities to different initial estimates. For example, in our implementation we weight each maturity by the squared inverse of the average bid-ask spread at the corresponding maturity in order to account for the relatively higher spreads at later maturities.
3.2 Step II: An explicit solution of Andreasen, Huge (2010)

The global algorithm of Step I yields a set of arbitrage free option prices across strikes and for all listed maturities. However, for practical applications, we usually require to have a much finer set of maturities with an inter-maturity time as small as one day (or a couple of days). This task would lead to a prohibitively large global optimization problem. In this section, we show a creative application of the Andreasen, Huge (2010) algorithm to fill in the times between listed maturities very quickly. Given that we already have arbitrage free option prices at all listed maturities, it turns out that one can solve the Andreasen, Huge (2010) problem explicitly, thus skipping its iterative optimization step.

Throughout we continue to work in the adjusted strike space $\tilde{K}$ introduced in the previous section but will drop the tilde for brevity. We briefly recall the core part of the Andreasen, Huge (2010) algorithm: starting from $C(K_j, 0) = (S_0 - K_j)_+$, one solves, sequentially, for each pair of (listed) maturities $[T_i, T_{i+1}]$, $i = 0, \ldots, m - 1$ with $T_0 = 0$, a problem of the form:

$$\min_{\theta \in \mathbb{R}^{n-1}} \sum_{j=0}^{n} \left( C(K_j, T_{i+1}) - C^M(K_j, T_{i+1}) \right)^2$$

with

$$\left[ 1 - \frac{1}{2} \Delta T_i \theta_{i+1,j}^2 \delta_{KK} \right] C(K_j, T_{i+1}) = C(K_j, T_i)$$

for $j = 1, \ldots, n - 1$

where $\delta_{KK}$ denotes the second order central finite difference operator, defined for $j = 1, \ldots, n - 1$ as:

$$\delta_{KK} C(K_j, T_i) = 2 \frac{C(K_{j-1}, T_i) - C(K_j, T_i)}{K_j - K_{j-1}} + \frac{C(K_{j+1}, T_i) - C(K_j, T_i)}{K_{j+1} - K_j}$$

and $\delta_{KK} C(K_j, T_i) = 0$ for $j = 0$ and $j = n$. In Andreasen, Huge (2010) $C^M(K_j, T_{i+1})$ denote the option market quotes we try to fit for maturity $T_{i+1}$.

However, we notice that using instead of $C^M(K_j, T_{i+1})$ our interpolated values from Step I, we can solve explicitly for the vector $\theta_{i+1} \in \mathbb{R}^{n-1}$. This is possible because our values $\tilde{C}(K, T_i)$ already satisfy the no-arbitrage conditions [10] and therefore we can compute, for $i = 0, \ldots, m - 1$ and $j = 1, \ldots, n - 1$

$$\theta_{i+1,j} = \sqrt{\frac{2 \tilde{C}(K_j, T_{i+1}) - \tilde{C}(K_j, T_i)}{\Delta T_i \delta_{KK} \tilde{C}(K_j, T_i)}}$$

thus bypassing the optimization problem. Once the vector $\theta_{i+1}$ has been computed, it is shown in Andreasen, Huge (2010) that we can fill in an
arbitrage-free way any inter-maturity time $T \in (T_i, T_{i+1})$ by solving the system of equations

\[
\left[ 1 - \frac{1}{2} (T - T_i) \theta_{i+1,j}^2 \delta_{KK} \right] C(K_j, T) = C(K_j, T_i) \quad \text{for } j = 0, \ldots, n.
\]

The system of equations may be written as a tridiagonal system and can thus be solved in $O(n)$ steps. Defining the vector $c(T_i)$ as the call option prices for maturity $T_i$ at grid points $(K_j)$ we have

\[
A_i c(T_i) = c(T_i) \quad \text{for } i \text{ such that } T \in (T_i, T_{i+1})
\]

where

\[
A_i = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
-\frac{z_{i,1}}{h_0} & 1 + \frac{z_{i,1}}{h_0} + \frac{z_{i,1}}{h_1} & \cdots & 0 \\
& \ddots & \ddots & \ddots \\
& & -\frac{z_{i,n-1}}{h_{n-2}} & 1 + \frac{z_{i,n-1}}{h_{n-2}} + \frac{z_{i,n-1}}{h_{n-1}} \\
& & & 0 & 1
\end{pmatrix}
\]

with $z_{i,j} = \frac{1}{2}(T - T_i)\theta_{i+1,j}^2$. The entire Step II, which fills-in the inter-maturity layers, is very fast and does not require any iterative solutions to optimization or minimization problems. Any desired inter-maturity layer is filled by solving a tri-diagonal linear system whose coefficients depend on the bracketing listed maturities; the latter have been constructed in Step I presented in the previous section.

To summarize, we recall that the first step of our algorithm considers all listed maturities jointly, while the second step fills in the inter-maturity layers sequentially. Therefore, our approach could best be interpreted as a hybrid between the fully sequential traditional approach and the surface approach considered in Fengler, Lin (2011) and Glaser, Heider (2012).

### 4 Numerical Applications

We now apply the algorithm described in the previous section to construct call price surfaces for several stocks listed on the Swiss Stock Exchange SIX. In practice, for single stocks, the listed options are usually American and hence it is necessary to convert these quotes to corresponding European quotes. In our implementation, we calculate the implied volatility of the American bid and ask quotes using binomial trees that account for general dividends similar to Vellekoop and Nieuwenhuis (2006) and then calculate the corresponding European call option quotes. The cash dividend amounts have been inferred from dividend futures listed on the EUREX.
Figure 2: Call price surface for NOVN with $\lambda = 0.1$ (July 4th, 2012)

Figure 3: Risk neutral densities for NOVN with $\lambda = 0.1$ (July 4th, 2012)
Figure 4 shows the constructed call price surface for Novartis on July 4th, 2012 with $\lambda = 0.1$. As expected by equation (4), discontinuities in the call price surface are visible at times $T = 238$ and $T = 603$ and correspond to the expected discrete dividend payments of the stock. Similarly, by equation (5), the same discontinuities are observed in the implied risk neutral densities (Figure 3) which are displaced horizontally at the ex-dividend dates. We further note that the implied risk neutral densities have an intuitive bell shape with a variance that increases with time to maturity.

The choice of the smoothing parameter $\lambda$ is central as it determines the direct trade-off between smoothness and bid-offer compliance. The role of the smoothing parameter is clearly visible in Figure 4 which shows the risk-neutral densities for a very low smoothing parameter ($\lambda = 10^{-6}$). The mid prices are only adjusted to impose the no-arbitrage conditions and, as a result, the risk-neutral densities are very irregular. If we increase the smoothing parameter $\lambda$ to $10^{-3}$ (see Figure 5) the risk neutral densities become smoother, but may still exhibit local minima and maxima. Increasing $\lambda$ to 0.1, we can obtain very smooth risk neutral densities while still observing the market bid-offer quotes (Figure 3). We note, however, that risk-neutral densities may exhibit legitimate bi-modal shapes under certain market conditions, e.g. in anticipation of important earnings announcements, and applying excessive smoothing may distort these shapes. Of course, the appropriate degree of smoothness will depend on the application at hand: for a vanilla market making desk, for example, a lower degree of smoothness may be appropriate in order to avoid distorting the market too much while, for an exotics desk, a higher degree of smoothness may be
In order to measure the bid-offer compliance, we introduce the relative position of the call price surface within the bid ask spread of an option quote $i \in O$

$$\tau_i = -1 + 2 \frac{\mu_j(K'_i) - P_{bid}^i}{P_{ask}^i - P_{bid}^i}$$

for $j$ such that $t_j = T_i$

As we impose the bid-offer conditions in the quadratic program, $\tau$ necessarily lies in the interval $[-1, 1]$ where $\tau = 0$ implies that the call price surface at the option quote is given by the mid price. We denote the squared relative deviation from the mid price by $\tau^2$. The role of $\tau$ and $\tau^2$ is to show the degree to which our resulting call price surface is constrained by the bid-offer conditions: low values of $\tau^2$ mean that the output surface is relatively close to the mid-market quotes, whereas large values of $\tau^2$ (i.e. closer to 1) mean that the output surface tends to lie closer to the boundaries of the bid-offer ranges. The more smoothness we demand, the more constraining the bid-offer conditions become. Figure 8 shows the average $\tau^2$ for different smoothing parameters $\lambda$ for the Novartis dataset. It is evident that there is a direct trade-off between smoothness and bid-offer compliance, which we can control using the smoothing parameter $\lambda$.

Figure 6 presents a different view of the NOVN options surface for fixed maturity slices. For each fixed-term implied volatility plot, we show the implied volatility curves corresponding to two different values of $\lambda$ and overlay the bid-offer spreads as vertical bars. It can now be easily seen how the...
Figure 6: Implied volatilities for NOVN on July 4th, 2012 (solid black line: $\lambda = 0.001$, dashed gray line: $\lambda = 0.1$)
resulting implied volatility surface is positioned inside the bid-offer spreads. Figure 7 shows the results for a different stock, ABB. We have carried out similar calculations for two additional stocks, ROG and ZURN, and found that the approach works reliably across datasets and for different choices of required smoothness. The full set of numerical results can be found on the
We finally note that our MATLAB implementation, which was run on an Intel i7 3.5 GHz machine, produced the following execution times for different sizes of the \( u \)-grid and the NOVN dataset:

<table>
<thead>
<tr>
<th>Grid size</th>
<th>200 x 11</th>
<th>400 x 11</th>
<th>800 x 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>11.7s</td>
<td>50.7s</td>
<td>340.4s</td>
</tr>
</tbody>
</table>

The 200-point grid is closer to the actual sizes that would be used in practice; additionally, we expect a C/C++ implementation to produce further improvements in execution times.

5 Summary and Conclusions

In this paper, we address the problem of constructing a smooth arbitrage-free implied volatility surface under realistic market assumptions. Our approach starts from the basic and raw ingredients found in the market: bid-ask quotes for a set of listed vanilla options, an underlying asset which may pay discrete dividends (either proportional or cash), a funding curve (for example, a LIBOR curve) and possibly a borrow-fee curve for the underlying asset (i.e. the fee short-sellers must pay to long-holders for borrowing the stock).

We have created a website under [http://n.ethz.ch/~bachemo](http://n.ethz.ch/~bachemo) where a full set of graphs is available. For four different stocks (ABB, NOVN, ROG, ZURN) we show (i) the call price surfaces, (ii) the risk-neutral density surfaces, (iii) the fixed-term IV plots and (iv) the goodness of fit tau-squared metric. We note that (i)-(iii) are all shown for two different values of lambda corresponding to lower and higher smoothness.
We have developed a two-step procedure to construct the implied volatility surface with smooth underlying risk-neutral densities while simultaneously observing any required bid-offer constraints. In particular—and unlike earlier literature—we seek to smooth the risk-neutral densities rather than call prices or implied volatilities. As previously discussed, a seemingly smooth call price surface or implied volatility surface, may hide unreasonable risk-neutral densities which would be unsuitable for pricing non-vanilla products.

The first step of our procedure applies a global quadratic problem to all listed maturities and balances the trade-off between the quality of fit and the smoothness of the risk-neutral densities. Due to the global nature of the fitting problem, it is no longer required to choose a particular sequence in which maturities are fitted. The second step is designed to quickly fill-in the time slices between listed maturities. We show a direct method which requires only the solution of a tri-diagonal system for each intermediary maturity. Finally, we perform several numerical tests using market data on four Swiss stocks and find that the method works robustly under different fitting scenarios.

6 Appendix

6.1 The matrices $L^j_g$ and $L^j_\gamma$

To calculate the matrices $L^j_g$ and $L^j_\gamma$ for a given maturity $j$, we define the set $\mathcal{O}^j$ which contains the indices of the option quotes at the maturity $j$. We recall that the value of the spline at an option quote $i \in O^j$ is given by

$$g_j(\tilde{K}_i) = \left[ \left(1 - \frac{x}{h_k}\right) g^j_k + \frac{x}{h_k} g^j_{k+1} + \left(\frac{x^2}{2} - \frac{h_k x}{3} - \frac{x^3}{6 h_k}\right) \gamma^j_k + \left(\frac{x^3}{6 h_k} - \frac{h_k x}{6}\right) \gamma^j_{k+1}\right]_{x=\tilde{K}_i - u_k}$$

where $k$ is such that $u_k \leq \tilde{K}_i < u_{k+1}$.

We define $k_i$ such that $u_{k_i} \leq \tilde{K}_i < u_{k_i+1}$ for $i \in \mathcal{O}^j$, introduce the vector $l$ that enumerates the elements in $\mathcal{O}^j$ and set $x_i = \tilde{K}_i - u_{k_i}$ for $i \in \mathcal{O}^j$. $L^j_g$ and $L^j_\gamma$ are then $|\mathcal{O}^j| \times |u|$ matrices that can be constructed row-wise: for
We have

\[ L^i_j(l_i, k_i) = \left(1 - \frac{x_i}{h_{k_i}}\right) \]

\[ L^i_j(l_i, k_i + 1) = \frac{x_i}{h_{k_i}} \]

\[ L^i_1(l_i, k_i) = \left(\frac{x_i^2}{2} - \frac{h_{k_i}x_i}{3} - \frac{x_i^3}{6h_{k_i}}\right) \]

\[ L^i_1(l_i, k_i + 1) = \left(\frac{x_i^3}{6h_{k_i}} - \frac{h_{k_i}x_i}{6}\right) \]

where all other entries are zero.

### 6.2 Proof of equation (3)

To obtain the calendar no-arbitrage condition for the dividend paying asset \( S_t \), it is convenient to start from the simpler no-arbitrage condition satisfied by the undiscounted call prices on the dividend-adjusted asset \( \tilde{A}_t \), namely

\[ \tilde{C}(\tilde{K}, T_2) \geq \tilde{C}(\tilde{K}, T_1) \]

where we recall the strike transformation given in equation [12]

\[ \tilde{K}(K, T) := \frac{D(0, T) \cdot p_d(0, T) \cdot K}{B(0, T)} + \sum_{t_i^d \leq T} \frac{D(0, t_i^d) \cdot p_d(0, t_i^d)}{B(0, t_i^d)} \cdot d_i \]

which, by solving for \( K \), we can write equivalently as

\[ K(\tilde{K}, T) := \frac{B(0, T) \cdot \tilde{K}}{D(0, T) \cdot p_d(0, T)} - \sum_{t_i^d \leq T} \frac{B(t_i^d, T)}{D(t_i^d, T) \cdot p_d(t_i^d, T)} \cdot d_i. \]

In particular, for two maturities \( T_1 \leq T_2 \), we can check by simple algebra the relationship

\[ K(\tilde{K}, T_1) = K(\tilde{K}, T_2) \cdot \frac{D(T_1, T_2) \cdot p_d(T_1, T_2)}{B(T_1, T_2)} + \sum_{T_1 \leq t_i^d < T_2} d_i \cdot \frac{D(T_1, t_i^d) \cdot p_d(T_1, t_i^d)}{B(T_1, t_i^d)}. \]
Using equation \((11)\), we have
\[
\bar{C}(\bar{T}, T_2) \geq \bar{C}(\bar{T}, T_1) \iff \\
D(0, T_2) \cdot p_d(0, T_2) \cdot C(K(\bar{T}, T_2), T_2) \geq D(0, T_1) \cdot p_d(0, T_1) \cdot C(K(\bar{T}, T_1), T_1) \iff \\
D(0, T_2) \cdot p_d(0, T_2) \cdot C(K(\bar{T}, T_2), T_2) \geq D(0, T_1) \cdot p_d(0, T_1) \cdot C(K(\bar{T}, T_1), T_1) + \\
\frac{D(T_1, T_2) \cdot p_d(T_1, T_2)}{B(T_1, T_2)} + \\
\sum_{T_1 < t^*_i \leq T_2} \frac{d_d(T_1, t^*_i) \cdot p_d(T_1, t^*_i)}{B(T_1, t^*_i)}
\]
which, upon accounting for the scaling terms, leads to the calendar no-arbitrage condition stated in equation \((3)\).

### 6.3 Proof of equation \((13)\)

The vectors \(g\) and \(\gamma\) are by definition a valid \textit{value-second derivative} representation of a general cubic spline with knots \(u_0, u_1, \ldots, u_n\) if and only if the resulting cubic spline is in \(C^2\). We recall that the value of the spline for \(u \in [u_i, u_{i+1})\) is given by
\[
g_i(u) = \left[ \frac{1 - x}{h_i} g_i + \frac{x}{h_i} g_{i+1} + \left( \frac{x^2}{2} - \frac{h_i x}{3} - \frac{x^3}{6 h_i} \right) \gamma_i + \\
\left( \frac{x^3}{6 h_i} - \frac{h_i x}{6} \right) \gamma_{i+1} \right]_{x = u - u_i}
\]
where \(h_i = u_{i+1} - u_i\). We differentiate twice with respect to \(u\):
\[
g'_i(u) = \left[ - \frac{g_i}{h_i} + \frac{g_{i+1}}{h_i} + \left( x - \frac{h_i}{3} - \frac{x^2}{2 h_i} \right) \gamma_i + \left( \frac{x^2}{2 h_i} - \frac{h_i}{6} \right) \gamma_{i+1} \right]_{x = u - u_i}
\]
\[
g''_i(u) = \left[ \left( 1 - \frac{x}{h_i} \right) \gamma_i + \frac{x}{h_i} \gamma_{i+1} \right]_{x = u - u_i}
\]
Continuity requires that \(\lim_{u \to u_{i+1}} g_i(u) = g_{i+1}(u_{i+1})\) for \(i = 0, 1, \ldots, n - 2\). This is always satisfied by construction as can be easily seen in equation \((15)\). First order differentiability requires \(\lim_{u \to u_{i+1}} g'_i(u) = g'_{i+1}(u_{i+1})\) for \(i = 0, 1, \ldots, n - 2\). This is equivalent to
\[
- \frac{g_i}{h_i} + \frac{g_{i+1}}{h_i} + \frac{h_i}{6} \gamma_i + \frac{h_i}{3} \gamma_{i+1} = - \frac{g_{i+1}}{h_{i+1}} + \frac{g_{i+2}}{h_{i+1}} + \frac{h_{i+1}}{3} \gamma_{i+1} + \frac{h_{i+1}}{6} \gamma_{i+2}
\]
\[
\iff \frac{1}{h_i} g_i - \frac{1}{h_i} g_{i+1} = \frac{1}{h_{i+1}} g_{i+1} + \frac{1}{h_{i+1}} g_{i+2} = \frac{h_i}{6} \gamma_i + \frac{h_i + h_{i+1}}{3} \gamma_{i+1} + \frac{h_{i+1}}{6} \gamma_{i+2}
\]
This condition can be rewritten as the matrix equation defined in equation (13). Second order differentiability, i.e. \( \lim_{u \to u_{i+1}} g''(u) = g''(u_{i+1}) \) for \( i = 0, 1, \ldots, n - 2 \), holds by construction as is easily seen in equation (16) and hence the matrix equation (13) is a necessary and sufficient condition for the vectors \( g \) and \( \gamma \) to be a valid value-second derivative representation of a cubic spline.

References


