

B-spline techniques for volatility modeling

Sylvain Corlay

▶ **To cite this version:**

Sylvain Corlay. B-spline techniques for volatility modeling. 2013. hal-00830378v4

HAL Id: hal-00830378
https://hal.science/hal-00830378v4

Preprint submitted on 11 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

B-spline techniques for volatility modeling

Sylvain Corlay*

May 29, 2014

Abstract

This paper is devoted to the application of B-splines to volatility modeling, specifically the calibration of the leverage function in stochastic local volatility models and the parameterization of an arbitrage-free implied volatility surface calibrated to sparse option data. We use an extension of classical B-splines obtained by including basis functions with infinite support.

We first come back to the application of shape-constrained B-splines to the estimation of conditional expectations, not merely from a scatter plot but also from the given marginal distributions. An application is the Monte Carlo calibration of stochastic local volatility models by Markov projection.

Then we present a new technique for the calibration of an implied volatility surface to sparse option data. We use a B-spline parameterization of the Radon-Nikodym derivative of the underlying's risk-neutral probability density with respect to a roughly calibrated base model. We show that this method provides smooth arbitrage-free implied volatility surfaces.

Finally, we sketch a Galerkin method with B-spline finite elements to the solution of the partial differential equation satisfied by the Radon-Nikodym derivative.

Keywords: B-splines, Tikhonov regularization, Radon-Nikodym, local volatility, stochastic volatility, finite elements, second-order cone programming, calibration

Introduction

This article is concerned with the calibration of volatility models to market option prices. We address the particle method for the calibration of the leverage function in stochastic local volatility models, the problem of fitting a smooth and arbitrage-free implied volatility surface to sparse option data, and the numerical treatment of the Kolmogorov forward equation.

We dedicate the first section to giving background on B-splines. The classical B-spline basis functions have compact support and practitioners usually handle extrapolation by adding external "ghost knots" with a certain multiplicity. We favor an alternative extrapolation scheme which consists of supplementing the basis with functions of infinite support as proposed in [26].

Section 2 is devoted to the problem of estimating a conditional expectation from given bivariate data and the knowledge of the marginal distributions. We review the Bayesian interpretation of Tikhonov regularization and we address the problem of compatibility between the marginal distributions and the conditional expectation. We show that shape-constrained B-splines are well suited to the problem of accounting for these compatibility conditions. Section 2.3 gives some general background on second-order cone programming and its application to shape-constrained B-splines.

In Section 3, we take on the recently devised particle method for calibrating the leverage function in stochastic local volatility models, where each time step requires the estimation of a conditional expectation, with a complete knowledge of the marginal distributions. We show that using the constrained regression method presented in Section 2 to account for the compatibility conditions significantly improves the accuracy of the method.

Section 4 is devoted to the problem of calibrating of a smooth arbitrage-free implied volatility surface from sparse option data. Our method is based on a B-spline parameterization of the Radon-Nikodym derivative with respect to a prior density. It allows for beliefs on the asymptotics of the volatility surface

*Bloomberg Quant Research, 731 Lexington Avenue, New York, NY 10022, USA. scorlay@bloomberg.net

to be accounted for through the choice of a base model. While most approaches proposed in the literature rely on some kind of general-purpose nonlinear optimizer such as the Levenberg-Marquardt algorithm. Finally, the problem of calibrating an arbitrage-free surface to market option prices can be formulated as a second-order cone program, which is solved efficiently using off-the-shelf software such as CVXOPT [5], Mosek [4] or CPLEX [1].

Finally, in Section 5 we look at discretization schemes of the Kolmogorov forward partial differential equation. We propose using a B-spline-based finite element space discretization of the forward PDE satisfied by the Radon-Nikodym derivative of the underlying's risk-neutral distribution with respect to a base model. This leads to the same kind of surface parameterization as we have considered earlier.

Notation: We use the following conventions: $\inf(\emptyset) = +\infty$ and $\sup(\emptyset) = -\infty$. $\mathcal{P}_n(\mathbb{R})$ is the set of real polynomials of order n. If X is a random variable on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $\mathbb{P}_X$ denotes its pushforward measure. Indexing of knots and B-spline basis functions start at 0.

1 Univariate B-splines and extrapolation

1.1 B-splines of infinite support

In this section, we present the extension of classical B-splines devised in [26] by Schumaker to include basis functions with infinite support.

Definition 1.1 (B-splines of infinite support, [26])**.** *Let k be a nonnegative integer and $\Gamma := \{\gamma_0 \leq \gamma_1 \leq \cdots \leq \gamma_{k-1}\}$ be a sorted collection of k knots. (If $k = 0$, $\Gamma := \emptyset$.) Let C_0 and C_1 be two positive constants. For a nonnegative integer $n \leq k$, a B-spline of order n associated with the knots Γ is a function of the form $\sum_{j=0}^{k+n} w_j b^\Gamma_{j,n}$, where the weights $(w_j)_{0\leq j\leq k+n}$ are real numbers and the functions $\left(b^\Gamma_{j,n}\right)_{n\geq 0,\ 0\leq j\leq k+n}$ are defined by*

$$\begin{cases} b^\Gamma_{0,0}(x) := \mathbf{1}_{(-\infty,\inf(\Gamma))}(x), \qquad b^\Gamma_{k,0} := \mathbf{1}_{[\sup(\Gamma),+\infty)}(x), \\ \text{and} \quad b^\Gamma_{j,0}(x) = \mathbf{1}_{[\gamma_{j-1},\gamma_j)}(x), \quad 1 \leq j \leq k-1, \end{cases} \tag{1}$$

and for $1 \leq n \leq k$, with the induction formula

$$\begin{cases} b^\Gamma_{j,n}(x) := \frac{(\gamma_j - x)}{C_0} b^\Gamma_{j,n-1}(x), & j = 0, \\ b^\Gamma_{j,n}(x) := b^\Gamma_{j-1,n-1}(x) + \frac{(\gamma_j - x)}{C_0} b^\Gamma_{j,n-1}(x), & 1 \leq j < \min(n,k), \\ b^\Gamma_{j,n}(x) := \begin{cases} b^\Gamma_{j-1,n-1}(x) + \frac{\gamma_j - x}{\gamma_j - \gamma_{j-n}} b^\Gamma_{j,n-1}(x), & \text{if } k > n, \\ b^\Gamma_{j-1,n-1}(x) + b^\Gamma_{j,n-1}(x), & \text{if } k = n, \end{cases} & j = \min(n,k), \\ b^\Gamma_{j,n}(x) := \frac{x - \gamma_{j-n-1}}{\gamma_{j-1} - \gamma_{j-n-1}} b^\Gamma_{j-1,n-1}(x) + \frac{\gamma_j - x}{\gamma_j - \gamma_{j-n}} b^\Gamma_{j,n-1}(x), & \min(n,k)+1 \leq j < \max(n,k), \\ b^\Gamma_{j,n}(x) := \begin{cases} \frac{x - \gamma_{j-n-1}}{\gamma_{j-1} - \gamma_{j-n-1}} b^\Gamma_{j-1,n-1}(x) + b^\Gamma_{j,n-1}(x), & \text{if } k > n, \\ b^\Gamma_{j-1,n-1}(x) + b^\Gamma_{j,n-1}(x), & \text{if } k = n, \end{cases} & j = \max(n,k), \\ b^\Gamma_{j,n}(x) := \frac{(x - \gamma_{j-n-1})}{C_1} b^\Gamma_{j-1,n-1}(x) + b^\Gamma_{j,n-1}(x), & \max(n,k)+1 \leq j < k+n, \\ b^\Gamma_{j,n}(x) := \frac{(x - \gamma_{j-n-1})}{C_1} b^\Gamma_{j-1,n-1}(x), & j = k+n. \end{cases} \tag{2}$$

Regarding the terms $\frac{x - \gamma_{j-n-1}}{\gamma_{j-1} - \gamma_{j-n-1}}$ and $\frac{\gamma_j - x}{\gamma_j - \gamma_{j-n}}$, the convention when the denominator is equal to zero is to replace it by 1 and 0 respectively.
The so-defined collection of functions $\left(b^\Gamma_{i,n}\right)_{0\leq i<k+n+1}$ for $0 \leq n \leq k$ are called the B-spline basis functions of order n. The B-splines of order n form a vector space of dimension $k+n+1$.

Definition 1.2 (B-splines of higher order)**.** *With the same notation as in Definition 1.1, for a nonnegative integer $n > k$, a B-spline of order n associated with the knots Γ is a function of the form $\sum_{j=0}^{k+n} w_j b^\Gamma_{j,n}$, where the weights $(w_j)_{0\leq j\leq k+n}$ are real numbers and where the functions $\left(b^\Gamma_{j,n}\right)_{n>k,0\leq j\leq k+n}$ are defined*

by the induction formula

$$\begin{cases} b^{\Gamma}_{j,n}(x) := \frac{(\gamma_j - x)}{C_0} b^{\Gamma}_{j,n-1}(x), & j = 0, \\ b^{\Gamma}_{j,n}(x) := b^{\Gamma}_{j-1,n-1}(x) + \frac{(\gamma_j - x)}{C_0} b^{\Gamma}_{j,n-1}(x), & 1 \leq j < \min(n,k), \\ \left(b^{\Gamma}_{j,n}(x)\right)_{\min(n,k) \leq j < \max(n,k)+1} & \text{any basis of } \mathcal{P}_{\max(n,k)-\min(n,k)}(\mathbb{R}) = \mathcal{P}_{n-k}(\mathbb{R}) \\ b^{\Gamma}_{j,n}(x) := \frac{(x - \gamma_{j-n-1})}{C_1} b^{\Gamma}_{j-1,n-1}(x) + b^{\Gamma}_{j,n-1}(x), & \max(n,k)+1 \leq j < k+n, \\ b^{\Gamma}_{j,n}(x) := \frac{(x - \gamma_{j-n-1})}{C_1} b^{\Gamma}_{j-1,n-1}(x), & j = k+n. \end{cases} \tag{3}$$

The positive constants C_0 and C_1 are the same as in Definition 1.1

Remark (On the choice of the constants C_0 and C_1)**.** *A desirable property for the B-spline basis functions is that if the collection of knots Γ is affinely transformed, the corresponding B-splines are affinely transformed as well. In other words, C_0 and C_1 should scale with Γ. In our implementation, we used $C_0 = C_1 = \frac{\gamma_{k-1} - \gamma_0}{k-1}$ if $\gamma_{k-1} > \gamma_0$ and $C_0 = C_1 = 1$ otherwise.*

We have defined B-spline basis functions of arbitrary order associated with an arbitrary finite collection of knots. In Figure 1, we display the B-spline basis functions of order 0, 1, 2 and 3, for the same collection of 8 knots, where we have taken $C_0 = C_1 = \frac{\gamma_{k-1} - \gamma_0}{k-1}$.

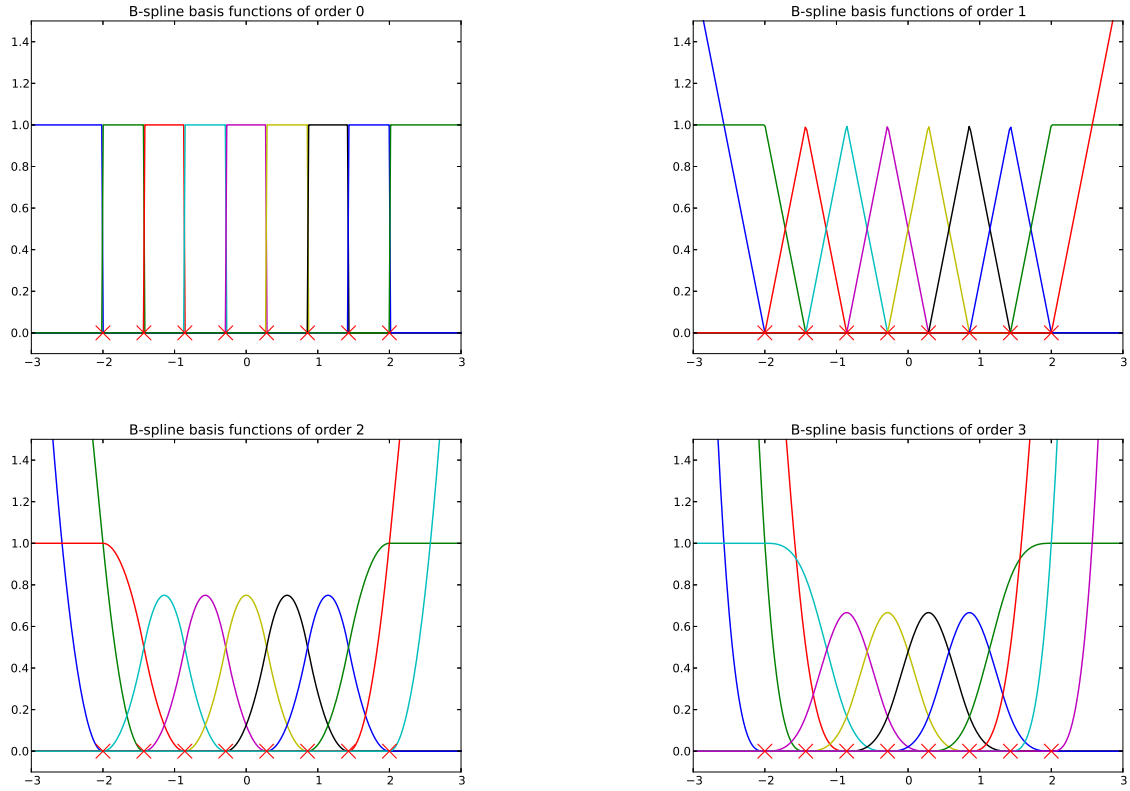


Figure 1: The B-spline basis functions of order 0, 1, 2 and 3, corresponding to the same collection of 8 knots.

Proposition 1.1 (Properties of B-splines)**.** *With the same notation as in Definition 1.1:*

- *If $\gamma_0 < \gamma_1 < \cdots < \gamma_{k-1}$ (with strict inequalities), then the vector space spanned by $\left(b^{\Gamma}_{j,n}\right)_{0 \leq j \leq k+n}$ is the set of C^{n-1} piecewise polynomial functions of order n over $\mathbb{R}$ with breakpoints Γ.*
- *If $\gamma_{l-1} < \gamma_l = \cdots = \gamma_{l+m-1} < \gamma_{l+m}$ for some $0 \leq l \leq k-1$ and $m \geq 1$, then the functions contained in $\operatorname{span}\left(b^{\Gamma}_{j,n}\right)_{0 \leq j \leq k+n}$ are piecewise polynomial functions of order n and are only C^{n-m} at γ_l.*

In other words, the multiplicity of a knot diminishes the regularity of the spanned set of piecewise polynomial functions at the corresponding breakpoint.

Remark (Basis truncation)**.** *When using B-splines for regression, a good way to avoid explosion of the extrapolation is to remove the basis functions of unbounded support that have a polynomial order strictly higher than t, $b^\Gamma_{0,n}, b^\Gamma_{1,n}, \cdots, b^\Gamma_{n-t-1,n}$ and $b^\Gamma_{k+t+1,n}, \cdots, b^\Gamma_{k+n,n}$. The resulting vector space has dimension $k + 2t - n + 1$. For $t = -1$, this reduces to the usual B-splines of compact support.*

This illustrates one of the main advantages of this basis over the classical approach using ghost points, as limiting the extrapolation order simply amounts to truncating the basis rather than imposing linear equality constraints.

The derivatives of these B-spline basis functions can be decomposed onto a B-spline basis of lower order.

Proposition 1.2 (Differentiation of B-splines)**.** *With the same notation, if $0 < n \leq k$,*

$$\left(b^\Gamma_{j,n}\right)' = \begin{cases} -\frac{n}{C_0} b^\Gamma_{j,n-1} & \text{for } \leq j < \min(n,k), \\ -\frac{n}{\gamma_n - \gamma_0} b^\Gamma_{n,n-1} & \text{for } j = n, \\ \frac{n}{\gamma_{j-1} - \gamma_{j-n-1}} b^\Gamma_{j-1,n-1} - \frac{n}{\gamma_j - \gamma_{j-n}} b^\Gamma_{j,n-1} & \text{for } n+1 \leq j < k, \\ \frac{n}{\gamma_{k-1} - \gamma_{k-n-1}} b^\Gamma_{k-1,n-1} & \text{for } j = k, \\ \frac{n}{C_1} b^\Gamma_{j-1,n-1} & \text{for } \max(n,k)+1 \leq j < k+n+1. \end{cases} \tag{4}$$

Proposition 1.3. *If we define the B-splines of order -1 by $b^\Gamma_{j,-1} := \delta_{x_j}$ for $j = 0, \cdots, k-1$, where δ_x denotes the Dirac mass centered at x. If $k > 0$ and $n = 0$, we have*

$$\left(b^\Gamma_{j,n}\right)' = \begin{cases} \text{for } 0 & 0 \leq j < \min(n,k), \\ \text{for } -b^\Gamma_{n,n-1} & \text{for } j = n, \\ \text{for } b^\Gamma_{j-1,n-1} - b^\Gamma_{j,n-1} & n+1 \leq j < k, \\ \text{for } b^\Gamma_{k-1,n-1} & \text{for } j = k, \\ \text{for } 0 & \max(n,k)+1 \leq j < k+n+1, \end{cases} \tag{5}$$

which can be seen as a limit case of Equation (4)*.*

For $0 \leq p \leq n$, the decomposition of the pth derivative of B-spline basis functions of order n onto the basis of order $n - p$ is obtained by iterating over this decomposition.

Remark (Integration and inner products of B-splines)**.** *Primitives and integrals of B-splines, as well as inner products of B-splines have closed-form expressions. An exact quadrature method is to use Gauss-Legendre points on each interval defined by the knots. A comprehensive study of methods to compute inner products of B-splines is carried out in [28].*

1.2 Evaluation and representation of B-splines

The forward evaluation scheme for basis functions

We can reformulate (2) in a simpler way. Starting from $b^\Gamma_{j,n}(x) = 0$, we write for $0 < n \leq k$

$$\begin{array}{ll} \text{For } 0 \leq j < \min(n,k) & b^\Gamma_{j,n}(x) \mathrel{+}= \frac{\gamma_j - x}{C_0} b^\Gamma_{j,n-1}(x), \\ & b^\Gamma_{j+1,n}(x) \mathrel{+}= b^\Gamma_{j,n-1}(x), \\ \text{For } \min(n,k) \leq j < \max(n,k) & b^\Gamma_{j,n}(x) \mathrel{+}= \frac{\gamma_j - x}{\gamma_j - \gamma_{j-n}} b^\Gamma_{j,n-1}(x), \\ & b^\Gamma_{j+1,n}(x) \mathrel{+}= \frac{x - \gamma_{j-n}}{\gamma_j - \gamma_{j-n}} b^\Gamma_{j,n-1}(x), \\ \text{For } \max(n,k) \leq j < k+n & b^\Gamma_{j,n}(x) \mathrel{+}= b^\Gamma_{j,n-1}(x), \\ & b^\Gamma_{j+1,n}(x) \mathrel{+}= \frac{x - \gamma_{j-n}}{C_1} b^\Gamma_{j,n-1}(x). \end{array} \tag{6}$$

This formulation is used to evaluate B-spline basis function in two ways:

1. The first and most natural approach is to use Formula (6) at each query point.

2. The second method is to implement Formula (6) in terms of operations in the *polynomial algebra*. With this pre-processing stage, we end up with a representation of the B-spline basis as the collection of their polynomial coefficients on each interval.

   The evaluation from the piecewise polynomial representation can be carried out using Horner's method, which is more efficient than recomputing the basis functions at new query points.

   Therefore the use of the second method, which involves a pre-processing stage, is beneficial if we evaluate the B-spline on a large number of points. The threshold for the number of evaluations is approximately equal to n evaluations by interval.

In every case, one can use knowledge of the support of B-spline basis functions for their representation in memory and their evaluation. For a fixed $x \in \mathbb{R}$, if $0 \leq n \leq k$ there are $n+1$ B-spline basis functions of order n that can be non-zero at x. More precisely, if $\gamma_i \leq x \leq \gamma_{i+1}$ (with the conventions that $\gamma_{-1} = -\infty$ and $\gamma_k = +\infty$) the only B-spline basis functions that are not equal to zero are $b^\Gamma_{j,n}$ for $i \leq j \leq i+n$.

The backward evaluation scheme

Regarding the evaluation of a B-spline function $f = \sum \alpha_j b^\Gamma_{j,n}$, the natural and naive approach would be to use the forward evaluation scheme (3) already presented for the basis functions and to compute the weighted sum. This is efficient if the B-spline basis functions have already been evaluated. However, if this is not the case, there is a more direct algorithm.

Indeed, using that the basis functions $b^\Gamma_{j,n}$ are decomposed onto the basis functions $b^\Gamma_{j-1,n-1}$ and $b^\Gamma_{j,n-1}$ (Equation (2)), we can show that $f = \sum \alpha^{(1)}_j b^\Gamma_{j,n-1}$ where the loadings $\alpha^{(1)}_j$ are piecewise polynomial of order 1 and carry on with the decomposition of $b^\Gamma_{j,n-1}$ onto a lower order basis. We find $f = \sum \alpha^{(i)}_j b^\Gamma_{j,n-i}$ where the loadings $\alpha^{(i)}_j$ are piecewise polynomial of order i. The algorithm stops when $i = n$ with the decomposition of f onto the trivial basis $\left(b^\Gamma_{j,0}\right)_{0 \leq j \leq k}$. To get the loadings $\alpha^{(i+1)}_j$ from $\alpha^{(i)}_j$, we start from $\alpha^{(i+1)}_j(x) = 0$ and write

$$\begin{array}{ll}
\text{For } 0 \leq j < \min(n-(i+1), k) & \alpha^{(i+1)}_j(x) \mathrel{+}= \alpha^{(i)}_j(x) \frac{\gamma_j - x}{C_0} + \alpha^{(i)}_{j+1}(x), \\
\text{For } \min(n-(i+1), k) \leq j < \max(n-(i+1), k) & \alpha^{(i+1)}_j(x) \mathrel{+}= \alpha^{(i)}_j(x) \frac{\gamma_j - x}{\gamma_j - \gamma_{j-n}} + \alpha^{(i)}_{j+1}(x) \frac{x - \gamma_{j-n}}{\gamma_j - \gamma_{j-n}}, \\
\text{For } \max(n-(i+1), k) \leq j < k+n-(i+1) & \alpha^{(i+1)}_j(x) \mathrel{+}= \alpha^{(i)}_j(x) + \alpha^{(i)}_{j+1}(x) \frac{x - \gamma_{j-n}}{C_1}.
\end{array} \tag{7}$$

This method is called backward evaluation. The scheme was proposed in [26, Chapter 5] for the case of B-splines of compact support. It can be carried out in the polynomial algebra as well, to obtain a piecewise polynomial representation of f.

Remark. *Backward and forward evaluation schemes can be used for the evaluation of derivatives of B-splines using Equation* (4).

The case of equally spaced knots

A critical stage of all evaluation schemes is the localization of the query points in the knot vector. In the general case, this is done by bisection with $O(\log(k))$ complexity. However, in the case where the knots are evenly spaced, this is reduced to an integer part computation. The case of equally spaced knots leads to further simplifications: all bounded spline basis functions have the same polynomial representation up to a parallel shift, and unbounded basis functions are symmetric. We can exploit these properties to save a significant amount of memory and computing.

2 Multiple regression and Bayesian considerations

In this section, we address the estimation of conditional expectations by multiple regression, with special attention paid to the case of the B-splines. We also recall the Bayesian theoretical foundation of Tikhonov regularization.

Then we tackle the problem of estimating a conditional expectation, not merely from a scatter plot but also given the marginal distributions $\mathbb{P}_X$ and $\mathbb{P}_Y$. The regression problem must be constrained to account for these compatibility conditions. The problem can be formulated as a second-order cone program.

Eventually, we show that this technique can be used as a time-stepping scheme in the particle method proposed in [17] for the calibration of stochastic local volatility models.

Multiple regression as an approximation of conditional expectation

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X, Y be two real random variables such as $Y \in L^2(\mathbb{P})$.

- $\mathbb{E}[Y|X]$ is the projection of Y onto the vector space $\left\{f(X), f \in L^2(\mathbb{P}_X)\right\}$, *i.e.*, it is the solution of
$$\min_{f \in L^2(\mathbb{P}_X)} \|Y - f(X)\|_2 ,$$
- while the multiple regression of Y with respect to a finite collection $(f_i)_{i\in I} \in \left(L^2(\mathbb{P}_X)\right)^I$ is the projection of Y onto the subspace $\operatorname{span}_{i\in I}\{f_i(X)\}$, *i.e.*, it is the solution of
$$\min_{w\in\mathbb{R}^I} \left\| Y - \sum_{i\in I} w_i f_i(X) \right\|_2 .$$

Hence, the larger the vector space $\operatorname{span}_{i\in I}(f_i)$, the better the approximation of $\mathbb{E}[Y|X]$ by the multiple regression of Y with respect to $(f_i(X))_{i\in I}$.

2.1 Regression of empirical distributions

In practice, we usually only have a finite sample $(x_j, y_j)_{1\le j\le N}$ of independent draws from (X, Y). A common approach is then to approximate the multiple regression of Y with respect to $(f_i(X))_{i\in I}$ by the regression of the corresponding empirical distributions. When doing so, enlarging the vector space onto which we project can be detrimental rather than beneficial. Indeed, performing a better regression of the empirical distribution does not mean that we get a better regression of the actual distribution of X with respect to Y. This phenomenon, also called "over-fitting", occurs for example when using a very fine grid for piecewise linear regression. Certain practitioners refrain from using a fine grid because of it. This means that they do not believe in wiggly results, that is, they have a prior belief on the smoothness of the conditional expectation.

Rather than refraining from refining the grid, another approach to the problem of over-fitting is the Bayesian approach, that is, to determine the most likely conditional expectation of Y with respect to X given the observed sample $(x_j, y_j)_{1\le j\le N}$ and the prior distribution for the conditional expectation.

Bayesian foundations of Tikhonov regularization

We now assume that X and $\mathcal{E}$ are L^2 real random variables and F is a random variable valued in $L^2(\mathbb{P}_X)$. We also assume that X, $\mathcal{E}$ and F are independent, and that $\mathcal{E} \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \sigma_{\mathcal{E}}^2)$. We define $Y := F(X) + \mathcal{E}$. If we assume that our prior distribution for F is proportional to $\exp\left(-\frac{\psi(F)}{2\sigma_F^2}\right)$ for some functional $\psi : L^2(\mathbb{P}_X) \to \mathbb{R}_+$, using Bayes' lemma and the independence of X, $\mathcal{E}$ and F, we obtain the following relation for the likelihood $L(F|(X,Y))$ of F knowing X and Y

$$L(F|(X,Y)) \propto L(F)L((X,Y)|F) = L(F)L((X,\mathcal{E})|F) \propto L(F)L(\mathcal{E}) \propto \exp\left(-\frac{1}{2\sigma_{\mathcal{E}}^2}\left(Y - F(X)\right)^2 - \frac{\psi(F)}{2\sigma_F^2}\right),$$

The functional ψ is usually a measure of irregularity such as $\psi(f) := C\int_{\mathbb{R}} \left(f^{(p)}(x)\right)^2 dx$ for a nonnegative integer p. If $(X_i, Y_i)_{1\le i\le N}$ are N independent copies of (X, Y), the likelihood of $F \in L^2(\mathbb{P}_X)$ given this sample satisfies

$$L(F|(X_i,Y_i)_{1\le i\le N}) \propto \exp\left(-\frac{1}{2\sigma_{\mathcal{E}}^2}\sum_{i=1}^{N}\left(Y_i - F(X_i)\right)^2 - \frac{\psi(F)}{2\sigma_F^2}\right).$$

Hence, maximizing the likelihood of f amounts to solving the minimization problem

$$\min_{f\in \text{span}_{i\in I}(f_i)} \underbrace{\frac{1}{N}\sum_{i=1}^{N}(y_i - f(x_i))^2}_{\text{estimator of } \mathbb{E}[(Y-f(X))^2]} + \underbrace{\frac{\sigma_{\mathcal{E}}^2}{\sigma_F^2}\frac{1}{N}}_{\text{Tikhonov factor}} \psi(f). \tag{8}$$

This shows us that the Tikhonov factor should be proportional to $\frac{1}{N}$ where N is the sample size, which is consistent with the intuition that the larger the sample is, the less we need to regularize to avoid over-fitting.

The quadratic case

In the case where the functional ψ is such that $\psi\left(\sum_{i\in I} w_i f_i\right)$ is a quadratic form in the loadings $(w_i)_{i\in I}$, the minimization problem (8) simply amounts to the minimization of a quadratic form. It is the case, for example when $\psi(f) = \int_{\mathbb{R}}(f^{(p)}(x) - g(x))^2 dx$ for some $g \in L^2(\mathbb{R})$. We solve the Tikhonov-regularized regression problem by solving the corresponding set of normal equations.

Let V be the matrix defined by $V_{ij} := \frac{1}{N}\sum_{l=0}^{N} f_i(x_l)f_j(x_l)$, $i,j \in I$ and c be the vector defined by $c_i := \frac{1}{N}\sum_{l=0}^{N} f_i(x_l)y_l$, $i \in I$. We assume that the quadratic form ψ is defined by $\psi(w) := \frac{1}{2}wPw + qw$, $w \in \mathbb{R}^I$. After some algebra, the minimization problem (8) amounts to

$$\min_{w\in\mathbb{R}^I} \frac{1}{2}wVw - cw + \lambda\left(\frac{1}{2}wPw + qw\right). \tag{9}$$

We obtain the following system of normal equations by differentiating (9)

$$(V + \lambda P)w + (\lambda q - c) = 0. \tag{10}$$

Quadratic forms of interest and measure of smoothness

In the case where the basis functions $(f_i)_{i\in I}$ are B-spline basis functions, measures of smoothness of the form $\psi(f) = \int_{\mathbb{R}}\left(f^{(p)}\right)^2(x)dx$ for some p can be explicitly derived in terms of the loadings w.

To begin with, if $f = \sum_{j=0}^{k+n} w_j b_{j,n}^{\Gamma}$ has non-zero weights on basis functions that have unbounded support and of extrapolating order higher or equal to p, we get $\psi(f) = +\infty$. Therefore, the basis truncation order t should always satisfy $t < p$. In other words, $w_i = 0$ for $0 \le j \le n-p$ and $k+p \le j \le k+n+1$. For example, with a penalization order $p = 2$, the maximum extrapolation order t should be strictly lower than 2. We obtain

$$\int_{\mathbb{R}}\left(\sum_{j=0}^{n+k} w_j \left(b_{j,n}^{\Gamma}\right)^{(p)}(x)\right)^2 dx = \sum_{i=0}^{n+k}\sum_{j=0}^{n+k} w_i w_j \int_{\mathbb{R}} \left(b_{i,n}^{\Gamma}\right)^{(p)}(x)\left(b_{j,n}^{\Gamma}\right)^{(p)}(x)dx.$$

Using the explicit decomposition of $\left(b_{j,n}^{\Gamma}\right)^{(p)}$ onto the B-spline basis of order $n-p$, $\left(b_{j,n-p}^{\Gamma}\right)_{0\le i<k+n-p+1}$, the coefficients of the quadratic form depend on inner products of basis functions of order $n-p$, $P_{ij} := C\int_{\mathbb{R}} b_{i,n-p}^{\Gamma}(x)b_{j,n-p}^{\Gamma}(x)dx$, which can be computed exactly using Gauss-Legendre quadrature or any of the other methods to compute inner products of B-splines presented in [28].

Remark (Penalization of order $n+1$)**.** *We are restricted to a penalization order satisfying $p \le n$. Using the Dirac comb $\left(b_{j,-1}^{\Gamma}\right)_{0\le j<k}$ introduced in Proposition 1.3, we can penalize the derivative of order $n+1$ in the same fashion. Regarding the inner product of B-splines of order -1, we use the convention*

- $\int_{\mathbb{R}}(b_{j,-1}^{\Gamma}(x))^2dx := \frac{1}{2}\left(\frac{1}{\gamma_j - \gamma_{j-1}} + \frac{1}{\gamma_{j+1}-\gamma_j}\right)$, $1 \le j < k-1$,

- $\int_{\mathbb{R}}(b^{\Gamma}_{0,-1}(x))^2dx := \frac{1}{2}\frac{1}{\gamma_1-\gamma_0}$ *and* $\int_{\mathbb{R}}(b^{\Gamma}_{k-1,-1}(x))^2dx := \frac{1}{2}\frac{1}{\gamma_{k-1}-\gamma_{k-2}}$,

which corresponds to the trapezoidal rule.

Remark (Invariance by re-scaling)**.** *A desirable property is that if the sample* $(x_j, y_j)_{1\leq j\leq N}$ *and the knots* Γ *are simultaneously affinely transformed, the result of the penalized regression remains the same.*

On the one hand, an affine transformation of the y-axis affects the regression error term and the penalization term in the same fashion and will not change the shape of the penalized regression. In the other hand, an affine transformation of the x-axis only affects the smoothness penalization term and thus its relative importance w.r.t. the regression error. In general, if $g(x/\lambda) = f(x)$ *for some* $\lambda > 0$*, then* $g^{(p)}(x/\lambda)/\lambda^p = f^{(p)}(x)$ *and* $\int_{\mathbb{R}}\left(f^{(p)}(x)\right)^2 dx = \frac{1}{\lambda^{2p-1}}\int_{\mathbb{R}}\left(g^{(p)}(u)\right)^2 du$*. Therefore, for a penalization order* $p > 0$*, we recommend a penalization factor proportional to* σ_X^{2p-1} *where the quantity* σ_X *scales proportionally with* X*, like the mean absolute deviation or the standard deviation. Finally the Tikhonov regularization factor should be of the form*

$$K\frac{\sigma_X^{2p-1}}{N}, \tag{11}$$

where K *is independent of* σ_X *and the sample size* N*.*

Numerical experiments with penalized regression

In Figure 2, we present the penalized regression of the same sample of (X, Y) with B-splines of various orders, various numbers of knots and penalization order $p = 2$. In every case, we used a Tikhonov regularization factor of $\frac{\sigma_X^{2p-1}}{N}$. We observe that the results are not very dependent on the spline order, or the number of knots once it is large enough. No additional tuning has been done. For these experiments, the random variables X and Y are defined by

$$\begin{aligned} &X \overset{\mathcal{L}}{\sim} \mathcal{N}\left(0, \sigma_X^2\right) \\ &Y := \tanh(2X/\sigma_X) + Z^2, \qquad \text{where} \quad Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \sigma_Z^2) \ \text{ is independent of } X, \end{aligned} \tag{12}$$

with $\sigma_X = \sigma_Z = 1$. This test case is nonlinear and presents changes of convexity.

In Figure 3, we illustrate the principal flaws of non-parametric regression of different orders. We display the shape of the smoothing kernel used for the regression. Order-0 local regression (local mean) will suffer from a general bias of order 1 in the case of a large bandwidth. Namely, when the bandwidth tends to infinity, the estimated value for $\mathbb{E}[Y|X = x]$ converges to $\mathbb{E}[X]$ regardless of x. In the case of too narrow bandwidth, order-0 local regression over-fit the data (See the first row in Figure 3). Similarly, order-1 local regression (local linear regression) suffers from an order-2 bias in the case of a large bandwidth and over-fits in the case of a narrow bandwidth (second row in Figure 3). Finally, local regression of order 2 will generally present exploding results on the wings.

Penalized B-spline regression has proven to be a rather robust regression method in comparison with classical non-parametric approaches. Non-parametric regression methods are very sensitive to parameters such as the regression order, the selection of the bandwidth, the shape of the smoothing kernel, and give poor control on extrapolation. Moreover, the piecewise polynomial form of B-splines allows for a natural representation of the regression result in memory and a natural way to evaluate it at new values. It has a solid theoretical foundation as a maximum likelihood estimator of the conditional expectation. More importantly, we will see that unlike non-parametric regression, B-splines allow to account for linear shape constraints such as non-negativity, monotonicity and convexity, and linear integral constraints. Such linear constraints come at practically no cost as the regression then amounts to a quadratic program.

2.2 Compatibility with the marginal distributions

A common application of multiple regression is the estimation of the conditional expectation of a random variable Y given another random variable X from a scatter plot of the joint distribution.

However, it often happens that additional information is available. For example, it is common that we completely know the marginal distributions of X and Y. If X and Y are two real L^1 random variables and $f : \mathbb{R} \to \mathbb{R}$ is a measurable function such that $\mathbb{E}[Y|X] = f(X)$ a.s. then the following conditions hold:

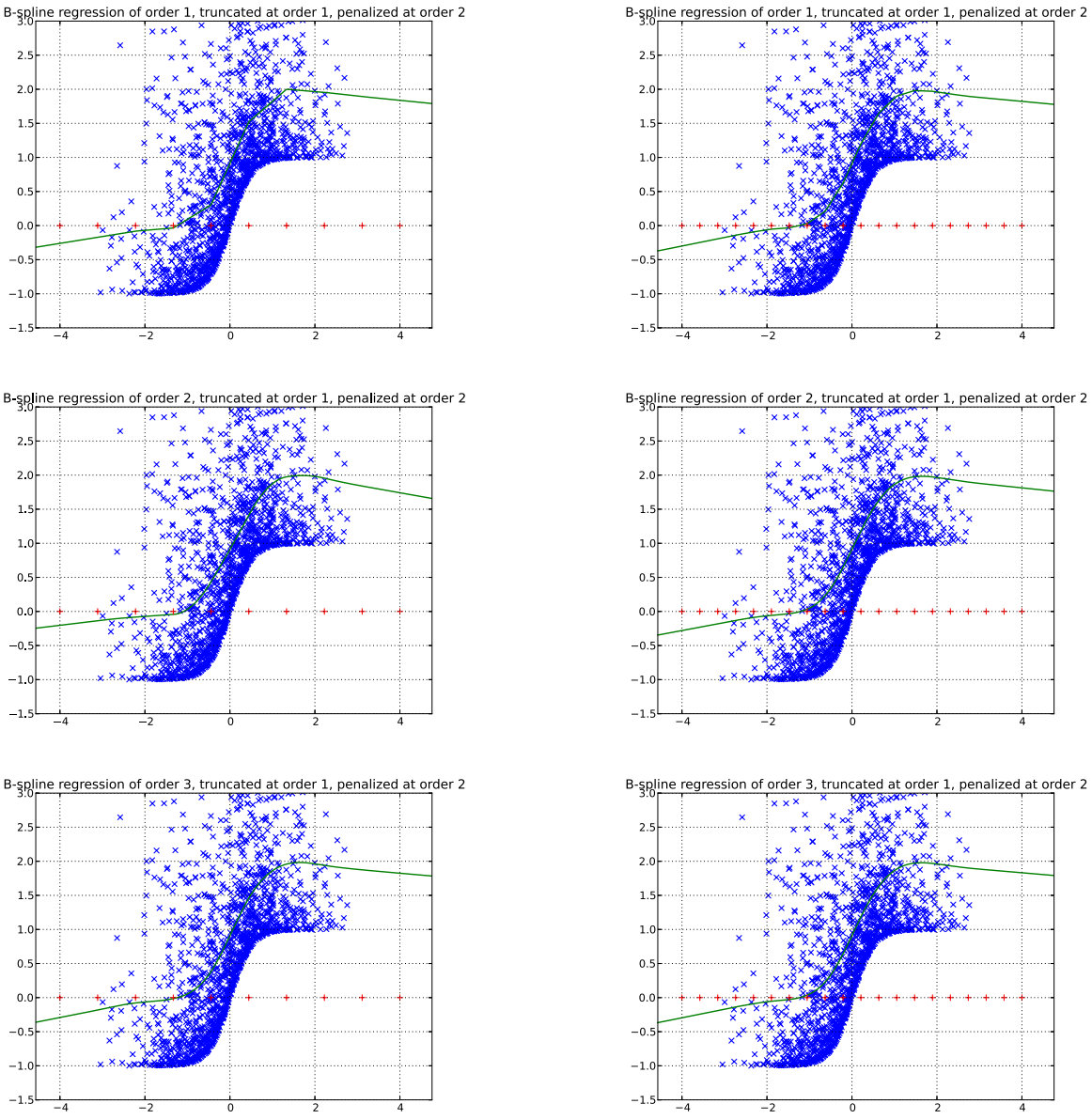


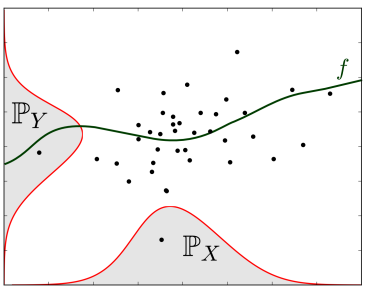
Figure 2: Penalized B-spline regression of a scatter plot sampled from Distribution (12). In every case, the penalization order is $p = 2$ and the truncation order is $t = 1$. The Tikhonov penalization factor is $\frac{\sigma_X^{2p-1}}{N}$ where $N = 1600$ is the sample size and σ_X the standard deviation of the X sample. We notice that the result does not depend significantly on the spline order. When using a finer discretization grid (right column), the results obtained with different spline orders get closer to each other.

$$\mathbb{E}[Y] = \int_{\mathbb{R}} f(x) d\mathbb{P}_X(x) \tag{13a}$$

$$f(x) \text{ is } \mathbb{P}_X\text{-a.s. in the convex hull of } \operatorname{supp}(\mathbb{P}_Y) \tag{13b}$$

For any nonnegative convex function ϕ,

$$\mathbb{E}[\phi(f(X))] \leq \mathbb{E}[\phi(Y)]. \tag{13c}$$



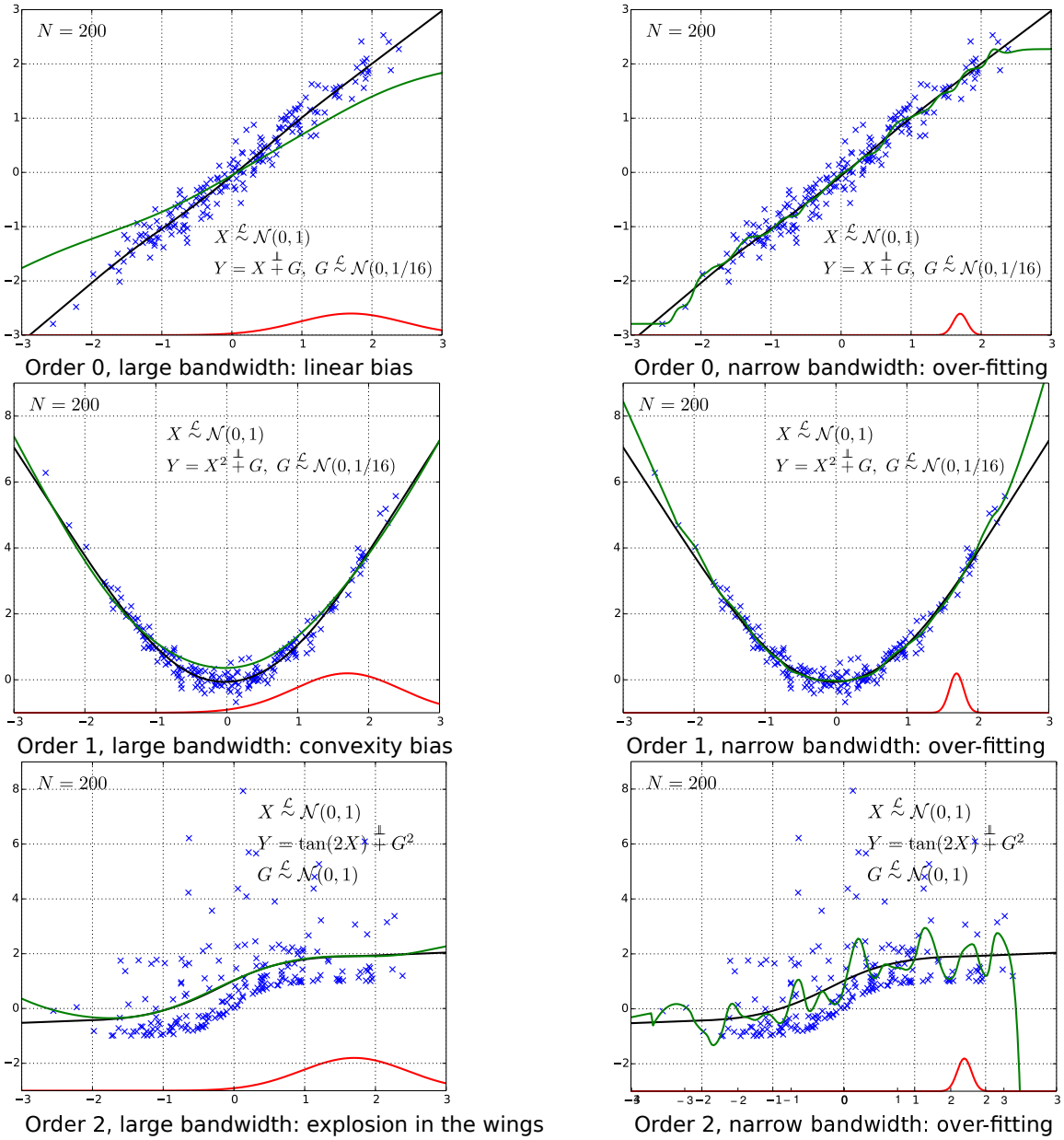


Figure 3: Illustration of the different flaws on non-parametric regression. In each figure the black line corresponds to a second-order B-spline regression with 10 knots evenly spaced between -2.5 and 2.5, with a Tikhonov-regularization of order $p = 2$ and a regularization factor of $\frac{\sigma_X^{2p-1}}{N}$. The green line is the result of the non-parametric regression. The first row corresponds to order-0 non-parametric regression (local mean); the second row corresponds to order-1 non-parametric regression (local linear regression); and the last row corresponds to order-2 non-parametric regression (local second-order polynomial fit).

Therefore, from a Bayesian point of view, it does not make sense to consider an estimate of the conditional expectation that does not satisfy these properties.

- The first condition (13a) amounts to a linear integral equality constraint.
- The second one (13b) consists of a set of linear inequality constraints. (In practice, it often amounts to a nonnegativity constraint in the regression.)
- Condition (13c) means that Y dominates $f(X)$ for the convex order. A consequence is that $\mathbb{E}\left[f(X)^2\right] \leq \mathbb{E}\left[Y^2\right]$, which is a quadratic inequality constraint.

The resulting constrained optimization problem can be formulated as a second-order cone program. Details on this special class of optimization problems is provided in Section 2.3.

Remark (Estimation of the conditional median rather than the conditional expectation)**.** *In [20], He and Ng proposed a constrained L^1 regression technique based on B-splines. More precisely, the quantity of interest that is parameterized with a spline is a conditional quantile distribution $\mathbb{P}[Y \leq g_\tau(x)|X = x]$, and $\tau = 1/2$ corresponds to the conditional median.*

2.3 Shape constraints and second-order cone programming

In this section, we first give some background on second-order cone programming and quadratic programming. Then, we review the shape constraints on B-splines that qualify as second-order cone constraints.

Second-order cone programming

A second-order cone program is a minimization problem of the form

$$\begin{array}{ll} \text{minimize} & f^T x \\ \text{subject to} & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \cdots, N, \end{array} \tag{14}$$

where $A_i \in M_{n_i-1,n}(\mathbb{R})$, $b_i \in \mathbb{R}^{n_i-1}$, $c_i \in \mathbb{R}^n$ and $d_i \in \mathbb{R}$. Second-order cone constraints, of the form $\|A_i x + b_i\|_2 \leq c_i^T x + d_i$ reduce to

- linear inequality constraints if $n_i = 1$, $(0 \leq c_i^T x + d_i)$,
- quadratic constraints if $c_i = 0$, $(\|A_i x + b_i\|_2^2 \leq d_i^2)$.

Moreover, in the case where the objective function itself is a positive definite quadratic form, we can recast it as a second-order cone program by appending an additional scalar t to the optimization variable. The optimization problem

$$\begin{array}{ll} \text{minimize} & x^T P^T P x + 2q_0^T x \\ \text{subject to} & \text{a collection of second-order cone constraints} \end{array}$$

where P is an invertible matrix, amounts to the minimization problem

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \text{the same collection of second-order cone constraints on } x \\ \text{and} & \left\|Px + P^{-1}q_0\right\|_2 \leq t \end{array}$$

where the new optimization variable is $(x_1, \cdots, x_n, t) \in \mathbb{R}^{n+1}$. Highly efficient software packages to solve second-order cone program are available, such as CVXOPT [5], Mosek [4] or CPLEX [1].

Shape constraints on B-splines

As we have seen in Section 1, if k is a nonnegative integer, $\Gamma := \{\gamma_0 \leq \gamma_1 \leq \cdots \leq \gamma_{k-1}\}$ is a sorted collection of k knots and if $n \leq k$, the B-spline basis functions of order n are nonnegative functions. Hence, the nonnegativity of each one of the the loadings is a sufficient condition for nonnegativity. It is also a finite set of linear constraints. This condition happens to be necessary for B-splines of order 0 and 1 as well as for the Dirac comb $\left(b_{i,-1}^\Gamma\right)_{0 \leq i < k}$ introduced in Proposition 1.3.

As we have seen in Proposition 1.2, derivatives of B-spline basis functions are explicitly decomposed onto the basis of lower order. Thus nonnegativity constraints on the first and second derivatives of B-splines translate into monotonicity and convexity constraints.

Remark. *There is no simple sufficient and necessary condition for spline nonnegativity of order $n \geq 2$. However, in [25] Papp and Alizadeh devised a method to handle the global nonnegativity constraints on B-splines without restraining to the case of nonnegative coefficients on a nonnegative basis, while remaining within the scope of second-order cone programs. In this article, we settle for the sufficient condition mentioned above.*

Other linear constraints

Equality and inequality constraints on the value of a B-spline or one of its derivatives at a certain point obviously qualify as linear constraints. It is also the case for inequality and equality constraints on limits of a B-spline or its derivatives at $-\infty$ or $+\infty$.

Regarding integral constraints, if $f_w = \sum_{j=0}^{k+n} w_j b_{j,n}^{\Gamma}$ is a B-spline of order n and μ a given locally finite measure, we have $I(w) := \int_{\mathbb{R}} f_w(x)d\mu(x) = \sum_{j=0}^{k+n} w_j \int_{\mathbb{R}} b_{j,n}^{\Gamma}(x)d\mu(x)$. Therefore, (if the quantities $\int_{\mathbb{R}} b_{j,n}^{\Gamma}(x)d\mu(x)$ are known for $0 \leq j < k+n+1$), equality and inequality constraints on $I(w)$ qualify as linear constraints.

Hierarchy of equality constraints: a modified Moore-Penrose pseudoinverse

Shape-constrained B-splines can be used as an interpolation method rather than a multiple regression method. In this case, there is temptation to consider the interpolation condition as firm equality constraints, and to use a measure of smoothness for the objective function in the resulting second-order cone program. However, we can encounter feasibility issues when using this approach. The input data could be unreachable with the given knots and spline order.

A more robust approach is to use all the degrees of freedom to achieve a least-square fit of the data points, and among the solutions of this problem, maximize smoothness. If interpolation is feasible, it will be achieved and the most regular interpolator will be returned.

The singular value decomposition of a real matrix $A \in M_{l,p}(\mathbb{R})$ is the decomposition $A = UDV^*$ where U and V are (complex) unit matrices and D is a $l \times p$ diagonal matrix. We denote by D^+ the $p \times l$ diagonal matrix obtained by inverting non-zero entries of D and transposing it. It satisfies the following properties:

- For $b \in \mathbb{R}^l$, $\widehat{x} := (VD^+U^*)b$ is a solution to $\min_{x \in \mathbb{R}^p} \|Ax - b\|_2$.
- If the minimization problem $\min_{x \in \mathbb{R}^p} \|Ax - b\|_2$ has multiple solutions, $\widehat{x}$ is the one which has the minimal Euclidean norm. $A^+ := VD^+U^*$ is called the Moore-Penrose pseudoinverse of A.

One could prefer to minimize another quadratic form $x \mapsto x^T Q x$, different from the Euclidean norm. If Q is positive definite and $Q = G^T G$ is its Cholesky decomposition, we define $\widehat{x} := G^{-1}\left(AG^{-1}\right)^+ b$. Using the properties of the Moore-Penrose pseudoinverse mentioned above, we find

- $\widehat{u} := G\widehat{x}$ minimizes $\left\|AG^{-1}u - b\right\|_2$, which implies that $\widehat{x}$ minimizes$\|Ax - b\|_2$,
- Gx has a minimal Euclidean norm, and thus $x^T Q x$ is minimal.

The matrix $A_Q^+ := G^{-1}\left(AG^{-1}\right)^+$ is the pseudoinverse of A that minimizes the quadratic form $Q = G^T G$. In general, we would always recommend to use this approach when all constraints are linear equality constraints, in order to give a best fit result in the case of infeasibility. However, the same analysis cannot be carried out in presence of inequality constraints. We refer to [12] for a thorough review of methods to handle hierarchies of constraints with more general quadratic programs.

3 Application to Guyon and Henry-Labordère's particle method

In this section, we take on the application of the shape-constrained regularized B-spline regression to the calibration of the leverage function in stochastic local volatility models.

Let S_t be the price of a risky asset at time t. We assume for the sake of simplicity that the asset does not pay any dividend and that interest and repo rates are zero. Then, arbitrage pricing theory tells us that under any risk-neutral probability, S_t is a martingale.

Knowledge of the call and put option prices of all strikes and maturities is equivalent to the knowledge of the risk-neutral densities of S_t for every maturity t. The celebrated Local Volatility Model [10] is the only Markov diffusion to match the corresponding continuum of marginal distributions. The

local volatility function is given by Dupire's stripping formula. (We use the Bachelier convention for instantaneous volatilities, that is, $dS = \sigma dW$, rather than the lognormal convention $dS = S\sigma dW$.)

$$\sigma^2_{\text{Dup}}(T,x) = \frac{\frac{\partial C}{\partial T}}{\frac{1}{2}\frac{\partial^2 C}{\partial K^2}}. \tag{15}$$

However, it is an arbitrary choice for the modeling of transition probabilities. It may not be a good model to price and hedge products that depend on these transition probabilities (see [11]). Pure stochastic volatility models, such as SABR [19], the Heston model [21] or Bergomi's model [7] are a first attempt of the modeling of these transition probabilities. A potential problem is that they do not have enough degrees of freedom to match all quoted vanilla option prices. A widespread approach [23] is the embedding of an additional level-dependent function $l(t,x)$, the leverage function, into the diffusion equation:

$$dS_t = a_t l(t,S_t) dW_t. \tag{16}$$

The local volatility term $l(t,S_t)$ is not only used as a means to achieve exact calibration to the market prices. It also allows one to recover the implied volatility skew using a combination of the contribution of the spot-vol correlation and the contribution of the local volatility term.

3.1 Calibration of the leverage function

Let us now consider the situation in which the Dupire local volatility is already calibrated. Equivalently, the entire implied volatility surface is known and therefore the risk-neutral marginal distributions $\mathbb{P}_{S_t}$ are known for any $t \geq 0$. The process a_t is a pure stochastic volatility process, which has also been determined. It only depends on S_t through the correlation of its driving process(es) with the Brownian motion W_t. We now tackle the calibration of the leverage function $l(t,S_t)$.

The calibration condition arises from Gyöngy's Markov projection theorem [18], which tells us that the so-defined process S_t has the same marginals as Dupire's Local Volatility Model $d\widetilde{S}_t = \sigma_{\text{Dup}}\left(t,\widetilde{S}_t\right) d\widetilde{W}_t$ if and only if

$$\sigma^2_{\text{Dup}}(t,x) = l(t,x)^2 \mathbb{E}\left[a_t^2 \middle| S_t = x\right]. \tag{17}$$

If the leverage function satisfies this condition, then (S_t, a_t) is the solution to a non-linear stochastic integro-differential equation

$$dS_t = \frac{\sigma_{\text{Dup}}(t,S_t)}{\sqrt{\mathbb{E}\left[a_t^2 | S_t\right]}} a_t dW_t. \tag{18}$$

A theoretical study on the existence of solutions to (18) was carried out in [2] by Abergel and Tachet from a PDE viewpoint in the case where a is a one-dimensional Itô process (see Appendix A). Regarding its numerical treatment, in Article [17], Guyon and Henry-Labordère devised a purely forward Monte Carlo method to integrate (18) and calibrate the leverage function $l(t,x)$. Let us also mention the related work of Van der Stoep, Grzelak and Oosterlee [9].

Let T be the horizon maturity for the calibration and $t_0 = 0 < \cdots < t_n = T$ a subdivision of $[0,T]$. We assume that the Dupire local volatility function σ_{Dup} is already calibrated and that the model parameters for a_t are already fixed. The calibration procedure proceeds as follows:

Monte Carlo calibration of the leverage function [17]
For every $0 \leq k < n$, simulate N independent draws of $S_{t_{k+1}}, a_{t_{k+1}}$ with a single Euler step (or another stepping scheme) from

- the N draws of the previous time step of S_{t_k}, a_{t_k}
- the calibrated leverage function $x \mapsto l(t_k,x)$.

We ensure the calibration condition (17) for the next time step by setting

$$l(t_{k+1},x) := \frac{\sigma_{\text{Dup}}(t_{k+1},x)}{\sqrt{\mathbb{E}\left[a_{t_{k+1}}^2 \middle| S_{t_{k+1}} = x\right]}}.$$

At each time step, the computation of the diffusion coefficient for a given path integrates contributions from all the other simulated paths via the estimation of the conditional expectation of a_t^2 knowing S_t. In this estimation, in addition to the scatter plot, the risk-neutral probability distribution of S_t is known (from the already interpolated implied volatility surface). Furthermore, the distribution of a_t^2 in most stochastic volatility models if also a given. For example, in the case of the Heston model, it is a noncentral chi-squared distribution. In the cases of the SABR model and Bergomi's model, a_t^2 is lognormal. Hence, we are in the situation studied in the previous section, where the marginal distributions are known.

(a) Condition (13a) results in an equality constraint for the instantaneous forward variance: $\mathbb{E}\left[a_t^2\right] = \int_{\mathbb{R}} f(x) d\mathbb{Q}_{S_t}(x) = \sum_{i=0}^{k+n} w_i \int_{\mathbb{R}} b_{i,n}^{\Gamma}(x) d\mathbb{Q}_{S_t}(x)$.

(b) The convex hull condition (13b) amounts to a nonnegativity constraint.

(c) A consequence of the third compatibility condition (13c) is the quadratic inequality constraint

$$\mathbb{E}\left[a_t^4\right] \geq \mathbb{E}\left[f(S_t)^2\right] = \sum_{i=0}^{k+n} \sum_{j=0}^{k+n} w_i w_j \int b_{i,n}^{\Gamma}(x) b_{j,n}^{\Gamma}(x) d\mathbb{Q}_{S_t}(x). \tag{19}$$

In Section 3.2, we present numerical experiments with the shape-constrained particle method for the calibration. We show that adding the shape constraints allows us to significantly reduce the number of Monte Carlo runs necessary to reach a desired level of accuracy.

Remark (Evaluation and representation of B-splines for the particle method)**.**

- *In the case of the particle method, the resulting B-splines are evaluated at the very same points where the basis functions where evaluated in the first place. Hence, we should keep the corresponding values and simply sum over the already evaluated basis functions rather than relying on backward evaluation schemes.*
- *The basis functions are evaluated on a Monte Carlo sample. In this context, it is typically beneficial to pre-compute the piecewise polynomial representation of the basis for the evaluation on this large number of points.*
- *Another advantage for the piecewise polynomial representation is that it is the best suited memory representation of the calibrated leverage function for any future use, which does not rely on the definition of the B-splines.*

Implementing the linear equality condition on the instantaneous forward variance $\mathbb{E}\left[a_t^2\right] = \int_{\mathbb{R}} f(x) d\mathbb{Q}_{S_t}(x)$ requires computing the quantities $\int b_{i,n}^{\Gamma}(x) d\mathbb{Q}_{S_t}(x)$. In the general case of an arbitrary risk-neutral distribution $\mathbb{Q}_{S_t}$, a numerical quadrature must be performed. However, in Section 4, we propose an interpolation method for the implied volatility surface which produces marginal distributions of the form $\mathbb{Q}_{S_t} = g\mathbb{Q}_t^0$ where the "prior" $\mathbb{Q}^0$ is the distribution of a simple base model like the Black-Scholes model and g is a B-spline parameterization of the Radon-Nikodym derivative of $\mathbb{Q}_{S_t}$ with respect to $\mathbb{Q}_t^0$.

In this case, computing $\int b_{i,n}^{\Gamma}(x) d\mathbb{Q}_{S_t}(x) = \int b_{i,n}^{\Gamma}(x) g(x) d\mathbb{Q}_t^0(x)$ amounts to integrating a piecewise polynomial with respect to the $\mathbb{Q}_t^0$ measure, which can be exactly evaluated in the cases of a Black-Scholes or Bachelier prior.

3.2 Numerical experiments with the shape-constrained particle method

Let us consider a model where the stochastic volatility process is the exponential of an Ornstein-Uhlenbeck process (like in Scott's model [27]), and the Black-Scholes implied volatility surface to be recovered is flat. More precisely for a volatility of volatility ν, a mean-reversion parameter θ, we consider the following model:

$$\begin{cases} dS_t = l(t, S_t) a_t dW_t, \\ a_t = a_0 \exp\left(U_t\right), \end{cases} \tag{20}$$

where U_t is a centered Ornstein-Uhlenbeck process,
$dU_t = -\theta U_t dt + \nu dW_t^{\sigma}$ and $d\langle W, W^{\sigma} \rangle_t = \rho dt$.

As in the previous section, we use the normal convention for instantaneous volatilities rather than the lognormal convention. The Black-Scholes implied volatility being flat, $\mathbb{Q}_{S_t}$ is lognormal and we have $\sigma_{\mathrm{Dup}}(t, x) = \sigma_{BS} x$. In this case, we get $\mathbb{E}[a_t^2] = a_0^2 \exp\left(\frac{\nu^2}{\theta}\left(1 - e^{-2\theta t}\right)\right)$.

Remark (Exact time stepping scheme for the volatility)**.** *Using that the volatility is the exponential of an Ornstein-Uhlenbeck U_t correlated with the spot's Brownian motion W_t, we can exactly simulate the bivariate Gaussian random variable $\left(W_{t_{i+1}}, U_{t_{i+1}}\right)$. Indeed, $U_{t_{i+1}} = U_{t_i} e^{-\theta(t_{i+1}-t_i)} + \nu \int_{t_i}^{t_{i+1}} e^{\theta(u-t_{i+1})} dW_u^\sigma$, and the centered Gaussian random variable $\left(\int_{t_i}^{t_{i+1}} e^{\theta(u-t_{i+1})} dW_u^\sigma, W_{t_{i+1}} - W_{t_i} \right)$ has covariance matrix* $\begin{pmatrix} \frac{1-e^{-2\theta(t_{i+1}-t_i)}}{2\theta} & \rho\frac{1-e^{-\theta(t_{i+1}-t_i)}}{\theta} \\ \rho\frac{1-e^{-\theta(t_{i+1}-t_i)}}{\theta} & t_{i+1}-t_i \end{pmatrix}$.

In Figure 4, we present the compared results of the particle method with and without taking into account the equality shape constrained (13a) in the B-spline regression, with various sizes of the Monte Carlo sample. We observe that the equality constraint roughly results into a parallel shift in the eventual implied volatility smile, correcting the general bias of the Monte Carlo sample. In all cases, the B-spline regression is constrained to be non-negative (convex hull constraint (13b)).

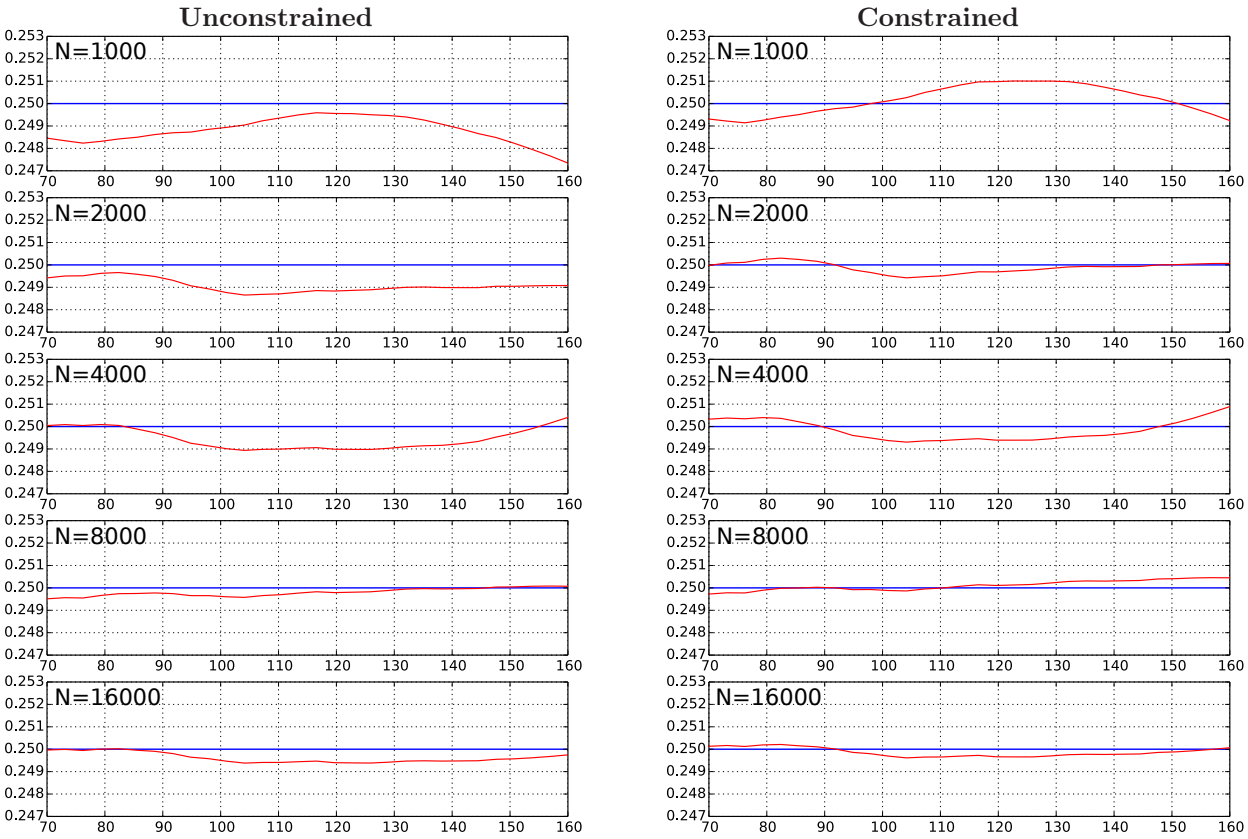


Figure 4: Implied volatility smiles of maturity $T = 1$ in Model (20), with parameters $S_0 = 100$, $a_0 = 0.2$, $\theta = 1$, $\nu = 30\%$, $\rho = -0.8$. The leverage function $l(t, x)$ is calibrated so as to recover a flat Black-Scholes implied volatility smile $\sigma_{BS} = 25\%$. The size of the Monte Carlo sample used in the calibration is indicated in the top-left corner of each figure. On the left-hand side, the B-spline regressions involved in the particle method are constrained to be non-negative. On the right-hand side, regressions are also required to satisfy the equality constraint (13a). In both cases, we use 20 knots, evenly spaced between -2.5 and 2.5 standard deviations around the forward price in the log space. The regression uses Tikhonov regularization for the second derivative ($p = 2$). The Tikhonov penalization factor is, $\sigma_{S_t}^{2p-1}/N$, as suggested in the previous section (Equation (11)).

In our numerical experiments, the quadratic inequality constraint (19) was inactive, and therefore it does not impact the results in practice, except in cases specially tailored so as to activate this constraint.

Remark (Application to the calibration of local correlation models)**.** *Another problem to which the particle method was successfully applied is the calibration of local correlation models (see e.g. [16]). Like*

in SLV models, the calibration procedure can benefit from shape constraints in the regressions ((13b) *and* (13c)*), and the other benefits of Tikhonov-regularized B-spline regression, but the equality constrained* (13a) *cannot be enforced as we do not have an explicit form for the expectation of the regressed random variable.*

4 Application to arbitrage-free completion of sparse option data

In this section, we address the problem of arbitrage-free completion of the vanilla option price surface from a sparse grid of vanilla option prices of various maturities and strikes. For the sake of simplicity, we first consider the case of a single maturity.

We propose to use a B-spline parameterization of the Radon-Nikodym derivative of the risk-neutral distribution with respect to a simple roughly calibrated base model. The base model can be normal (Bachelier), lognormal (Black and Scholes [8]) or more sophisticated like S.V.I. (Gatheral and Jacquier [14]). The resulting calibrated risk-neutral density inherits certain properties of the base model. For example, in the case of the S.V.I. prior, the calibrated density will also have fat tails.

Arbitrage conditions on vanilla option prices

As we have seen in Section 2.3, convexity, integral constraints, equalities and inequalities on values and derivatives of B-splines qualify as linear constraints. All the conditions of absence of arbitrage on call option prices fall into these categories. Namely, if $K \mapsto C_T(K)$ and $K \mapsto P_T(K)$ are the (undiscounted) call and put option prices of strike K and fixed maturity T on some given underlying, the conditions of absence of arbitrage for a given maturity are given by:

For call option prices:

- $K \mapsto C_T(K)$ is nonnegative, nonincreasing and convex on $\mathbb{R}$,
- $\lim_{K \to +\infty} C_T(K) = 0$ and $\lim_{K \to -\infty} C'_T(K) = -1$.

For put option prices:

- $K \mapsto P_T(K)$ is nonnegative, nondecreasing and convex on $\mathbb{R}$,
- $\lim_{K \to -\infty} P_T(K) = 0$ and $\lim_{K \to +\infty} P'_T(K) = 1$.

All these conditions put together mean that the second derivative of the (undiscounted) call or put option price as a function of the strike in the sense of distributions is a probability distribution. This probability distribution is the risk-neutral probability of the underlying asset at maturity[1]. Arbitrage-free interpolation of vanilla option prices amounts to a density estimation problem.

The forward price F_T (the risk-neutral expectation of the underlying) must be equal to $\lim_{K \to -\infty} C_T(K) + K$ and to $\lim_{K \to +\infty} K - P_T(K)$. This amounts to $C_T(0) = F_T$ in the case where the underlying can only be nonnegative (such as a stock price). Bid-ask spreads on the forward price and call option prices, as well as firm equality constraints on these quantities qualify as linear constraints. Hence, there is a temptation to use a B-spline parameterization of the call option price as a function of strike (as in [22] and [13]). However, the resulting probability distribution would have compact support which is unrealistic. It has also been proposed to directly use a spline parameterization of the risk-neutral probability distribution [24], which has an identical flaw: the resulting density has compact support.

B-spline parameterization of the Radon-Nikodym derivative with respect to a base model

The method of our choice is to start from a prior (L^1) probability distribution, $\mathbb{Q}^0_{S_T}$. For example, $\mathbb{Q}^0_{S_T}$ can correspond to a rough calibration of the Black-Scholes model. We then use a B-spline parameterization of the Radon-Nikodym derivative of the risk-neutral distribution of S_T, $\mathbb{Q}_{S_T}$ with respect to $\mathbb{Q}^0_{S_T}$. For the sake of brevity, we will use the shorter notation $\mathbb{Q}^0_T := \mathbb{Q}^0_{S_T}$ and $\mathbb{Q}_T := \mathbb{Q}_{S_T}$.

$$\mathbb{Q}_T = f_w \mathbb{Q}^0_T = \left(\sum_{i=0}^{k+n} w_i b^{\Gamma}_{i,n} \right) \mathbb{Q}^0_T.$$

[1] Actually, it is the so-called forward-probability of the underlying asset, which coincides with the risk-neutral density if interest rates are assumed to be deterministic. We will not make the distinction in the rest of the paper.

We truncate the B-spline to 0th order (constant) extrapolation so that the resulting density is also ensured to be L^1. Constraints on f_w are

$$f_w \geq 0 \qquad \text{and} \qquad \int_{\mathbb{R}} f_w(x) d\mathbb{Q}_T^0(x) = 1.$$

The forward price is given by

$$F_T = \int_{\mathbb{R}} x f_w(x) d\mathbb{Q}_T^0(x),$$

while the undiscounted call and put option prices of strike K are given by

$$C_T(K) = \int_K^{+\infty} (x-K) f_w(x) d\mathbb{Q}_T^0(x) \qquad \text{and} \qquad P_T(K) = \int_{-\infty}^K (K-x) f_w(x) d\mathbb{Q}_T^0(x).$$

All these quantities happen to be linear forms of the loadings w. As a result, problems such as

- the least-square fit of (mid) vanilla option prices under these constraints and a firm equality constraint on the forward,
- minimization of the mean-square second derivative of f_w under the firm constraint of yielding a price within bid-ask for each listed strike and for the forward price,

all qualify as second-order cone programs.

Choice of the base model

The advantage of lognormal base models is that integrals of the form $\int_a^b x^n d\mathbb{Q}_T^0(x)$ have closed-form expressions. However, this base model fails to account for implied volatility skew. The consequence is that the calibrated Radon-Nikodym density typically either explodes or almost vanishes in the wings and therefore, it is not very well approximated by piecewise polynomials. The advantage of more sophisticated models such as the SVI parameterization is that it can account for general features of the implied risk-neutral density of S_T and thus more regular results for the Radon-Nikodym derivative. However, there is no closed-form expression for quantities of the form $\int_a^b x^n d\mathbb{Q}_T^0(x)$ if $n > 1$, and one has to resort to approximate quadrature methods.

In Figure 5, we have calibrated a B-spline parameterization of the Radon-Nikodym derivative $\frac{d\mathbb{Q}_T}{d\mathbb{Q}_T^0}$ where $\mathbb{Q}_T$ is the risk-neutral probability distribution of the VIX volatility index on October 16th, 2013 seen from May 6th, 2013.

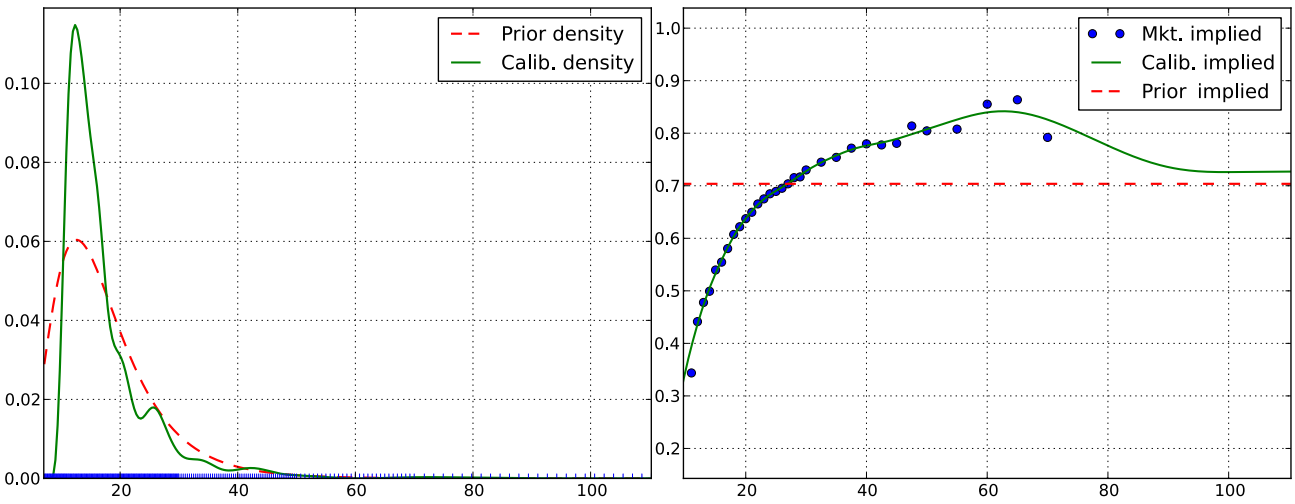


Figure 5: Risk-neutral density (left) and implied volatility smile (right) of the VIX index for expiry October 16th, 2013 as seen on May 6th, 2013. The red dotted curve corresponds to the Black-Scholes base model, and the green curve to the calibrated implied risk-neutral density.

In Figure 6, we have calibrated a B-spline parameterization of the Radon-Nikodym derivative $\frac{d\mathbb{Q}_T}{d\mathbb{Q}_T^0}$ where $\mathbb{Q}_T$ is the risk-neutral probability distribution of the Dow Jones index for maturity December 20th, 2014 as seen on April 25th, 2014 around 11am.

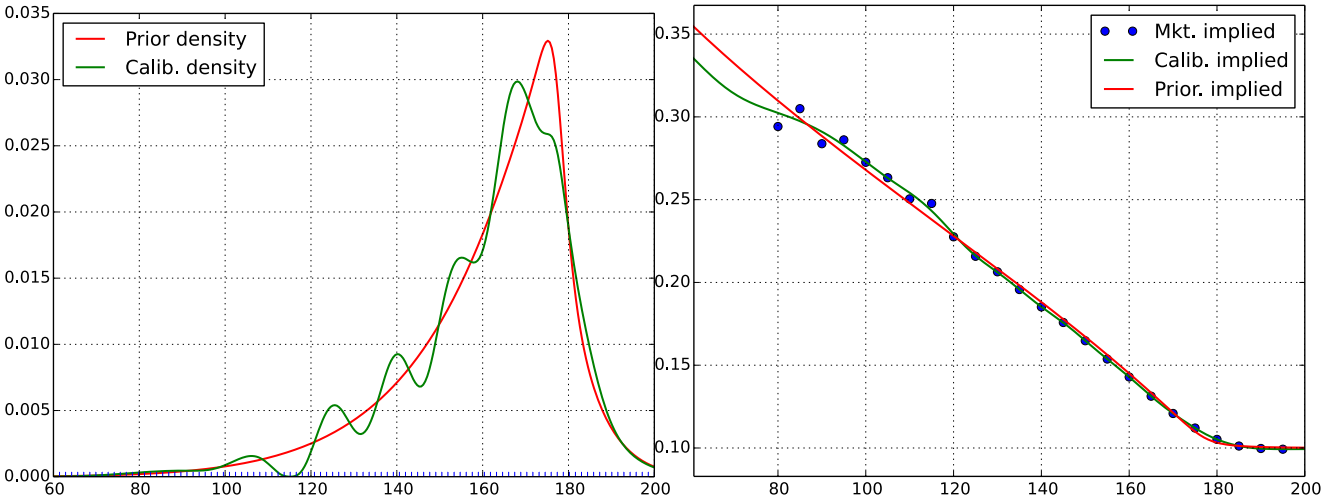


Figure 6: Risk-neutral density (left) and implied volatility smile (right) of the Dow Jones index for expiry December 20th, 2014 as seen on April 25th, 2014. The red dotted curve corresponds to the roughly calibrated SVI base model, and the green curve to the calibrated implied risk-neutral density. The SVI parameters are $a = 6.52e-3, b = 0.036, \rho = -0.99, m = 0.091$ and $\sigma_0 = 0.027$.

Eventually, we obtain a parameterization of the Radon-Nikodym density that also gives a smooth implied volatility smile. The volatility smile resulting from the procedure has the same asymptotics as the ones of the prior model, as the parameterized Radon-Nikodym derivative has a flat extrapolation. The prior model is typically parametric like Black-Scholes or the SVI parameterization which is flexible and often results in a decent fit. However, SVI cannot account for potential local features of the implied risk-neutral distribution such as bimodality. Our approach improves the fit of the parametric method while preserving the asymptotics and absence of arbitrage.

Calendar arbitrage conditions

So far, we have only addressed the case of a single maturity. If one assumes that rates are deterministic, the calendar arbitrage constraint is that undiscounted call and put option prices of constant relative strike (with respect to the forward price of the same maturity) is a non-decreasing function of maturity.

If we are given a sequence of listed maturities, we can enforce this condition on a fine grid of relative strikes, which amounts to a finite set of linear constraints. The arbitrage-free calibration of the implied volatility surface then amounts to a second-order cone program. However, this is not entirely satisfactory:

- On the one hand, we need the calendar arbitrage condition to be satisfied at any point lower than the first knot or higher than the last knot. If $0 \leq T_1 \leq T_2$ are two maturities and x is a fixed relative strike, we denote $K_1 := xF_{T_1}$ and $K_2 := xF_{T_2}$. Now, if x is such that K_1 and K_2 are both greater than the maximum knot, we have

$$C_{T_1}(K_1) = \int_{K_1}^{\infty} w_{T_1}(x - K_1) d\mathbb{Q}_{T_1}^0(x) \qquad \text{and} \qquad C_{T_2}(K_2) = \int_{K_2}^{\infty} w_{T_2}(x - K_2) d\mathbb{Q}_{T_2}^0(x),$$

  where w_{T_1} and w_{T_2} are the coefficients of the order-zero extrapolating B-spline basis function in the parameterizations of the Radon-Nikodym derivative for maturity T_1 and T_2 respectively. If the base model is assumed to be free of static arbitrage, $w_{T_2} \geq w_{T_1}$ is a sufficient condition to ensure that $C_{T_2}(K_2) \geq C_{T_1}(K_1)$. Regarding the low-strike extrapolation, the same argument holds when working with the put option prices.

- On the other hand, we need to ensure absence of calendar arbitrage in the interpolation region. In practice, we can settle for enforcing inequalities $C_{T_1}(xF_{T_1}) \leq C_{T_2}(xF_{T_2})$ only on a fine grid

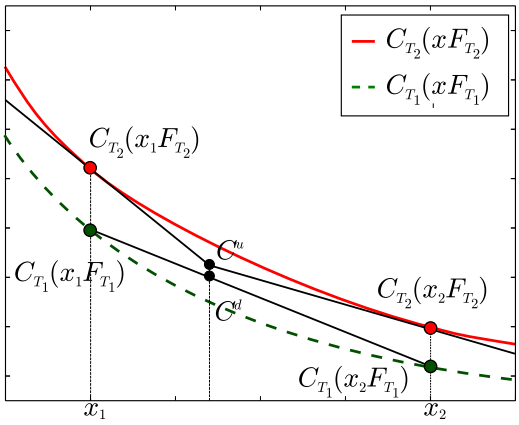


Figure 7: C^u is the ordinate of the interpolation between the tangents of the call option price of maturity T_2 as a function of the relative strike x. C^d is the value of the linear interpolation between $C_{T_1}(x_1F_{T_1})$ and $C_{T_1}(x_2F_{T_1})$ at the corresponding abscissa. $C^u \geq C^d$ is a sufficient condition for the absence of calendar arbitrage on $[x_1, x_2]$. The same equivalent argument holds for the Put option prices.

of values of x, for each couple of subsequent listed maturities ($T_1 < T_2$). However, it can still happen that the calendar arbitrage occurs between two points of the grid. A sufficient condition to ensure absence of calendar arbitrage in the interpolation is detailed in Figure 7. This second-order inequality condition is rather strong and it is the only nonlinear constraint. Therefore, for the numerical experiments, we settled for the fine grid condition.

This discussion shows that the calibration of the Radon-Nikodym density with respect to a prior model, under the constraints of absence of arbitrage amounts to a global second-order cone program. Second order cone programs can be solved in minimum running time using available software such as CVXOPT [5], Mosek [4] or CPLEX [1]. If the base model is Black-Scholes, the pricing of call and put options in the calibrated model remains explicit. The risk-neutral density in the calibrated model also has a closed-form expression. Hence, the denominator in the local volatility stripping formula (15) has a closed-form expression.

Smoothness in the time direction

In [6], Andreasen and Huge proposed an interpolation and extrapolation method which is guaranteed to yield arbitrage-free surfaces. Nonetheless, the resulting implied volatility surface fails to be differentiable in time at listed market maturities and thus, the corresponding local volatility has discontinuities in the time direction. Even though it is not formally an arbitrage, one could take advantage of a market maker using this method in markets where listed maturities are sliding with the current date, like foreign exchange markets. If the discontinuity is positive one can sell a "calendar butterfly" $C_{T+h} + C_{T-h} - 2C_T$ where T is a given market maturity and h is 1 day. One day later, all the tenors of the corresponding options will be lower than the new listed market maturity and the calendar butterfly will be valued at (almost) 0 by the market maker if he has recalibrated his model. One can then buy it back for a much lower price. In the case where the discontinuity is negative, we can apply the opposite strategy. In short, in option markets with sliding listing maturity, market makers re-calibrating their models every day expose themselves to "calendar butterfly" strategies if their calibration procedure for the implied volatility surface has discontinuities in the time direction.

Therefore, one should, wherever possible, produce implied volatility surfaces that are C^1 in the time direction. This can be ensured in our method by adding a term in the regularity penalization corresponding to the second time-derivative of loadings of the B-spline basis functions.

Choice of the base model

Like in the case of a single maturity slice, we may choose a simple parametric model like Bachelier or Black-Scholes as a base model $\mathbb{Q}_t^0$, because quantities of the form $\int_a^b x^n d\mathbb{Q}_t^0(x)$, $(a,b) \in \mathbb{R}^2$ have closed-form expressions for any maturity t. For example, in Figure 8, we display the results of the global calibration to a sparse grid of vanilla option prices. Some of the maturities do not have any listed option prices. We can add as many of those intermediate maturities as necessary to obtain a fine grid in the time direction. The base model is the Black & Scholes model with volatility equal to 20%.

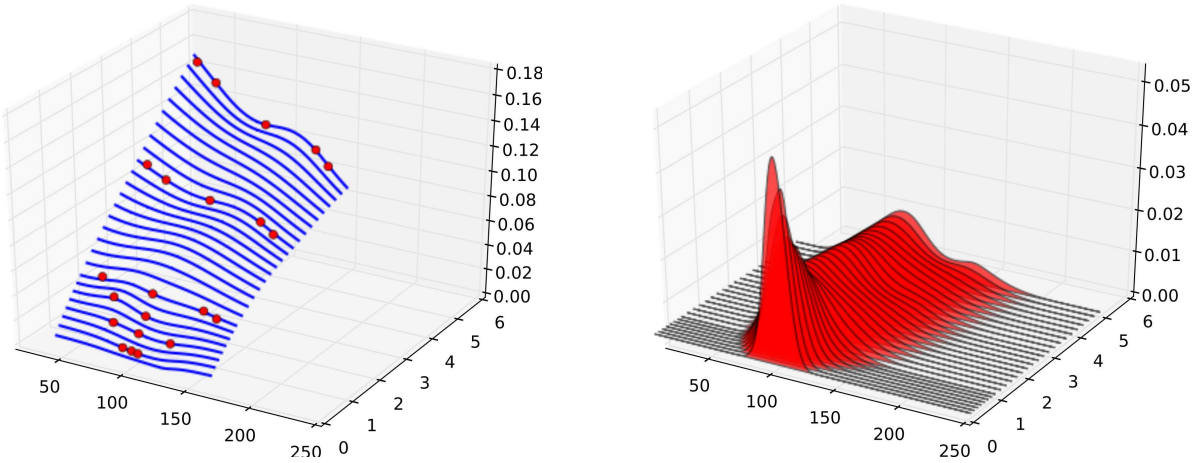


Figure 8: Calibration of Radon-Nikodym densities from sparse vanilla option data. On the left-hand side, we plot the sequence of implied total lognormal variances (input data and interpolation). On the right-hand side, we show the calibrated risk-neutral densities for the same maturities.

However, in the case where the market smile is skewed, the Radon-Nikodym derivatives $\frac{d\mathbb{Q}_t}{d\mathbb{Q}_t^0}$ will either explode or vanish far in or out of the money. Therefore, we need a parametric base model, flexible enough to account for general features of the volatility smile like fat-tailness and skewness, while remaining globally arbitrage-free. Such a parametric form was discovered by Gatheral and Jacquier [14]. They devised a large family of parametric models, which amounts to a usual SVI parameterization at each maturity slice. Although, it is still unable to account for local features of the volatility smiles like multi-modality. In this context, our Radon-Nikodym parameterization method can be seen as a multiplicative residual fitting algorithm for the SVI model, which still guarantees absence of static arbitrage.

The global SVI prior (SSVI [14])

We denote the log-moneyness by $k := \log(x/F_t)$, where F_t is the forward price of maturity t. The implied Black-Scholes total variance is denoted by $w(t,k) := \sigma_{\text{BS}}^2(t,x)t$. The Black-Scholes formula then gives $C_{BS}(k) = F_t\left(\mathcal{N}(d_+(t,k)) - e^k\mathcal{N}(d_-(t,k))\right)$, where $d_\pm := -\frac{k}{\sqrt{w(t,k)}} \pm \frac{\sqrt{w(t,k)}}{2}$. If we define $g(t,k) := \left(1 - \frac{kw'}{2w}\right)^2 - \frac{g'}{4}\left(\frac{1}{w} + \frac{1}{4}\right) + \frac{w''}{2}$, the risk-neutral density of the spot at maturity t is

$$p(t,x) = \frac{1}{x}\frac{g(t,k)}{\sqrt{2\pi w(t,k)}}\exp\left(-\frac{d_-(t,k)^2}{2}\right) \qquad (k = \log(x/F_t)).$$

For a given maturity, Gatheral and Jacquier proposed to use the following parametric form

$$w(t,k) = \Delta_t + \frac{\theta_t}{2}\left(1 + \zeta_t\rho k + \sqrt{(\zeta_t k + \rho)^2 + (1-\rho^2)}\right),$$

which coincides with a special case of the original SVI parameterization, up to a change of variable. Then ζ_t is assumed to be a function of θ_t, that is $\zeta_t = \phi(\theta_t)$. It is proved that this so-called SSVI parameterization is free of butterfly arbitrage if for all $\theta > 0$,

$$\theta\phi(\theta)(1+|\rho|) < 4 \quad \text{and} \quad \theta\phi(\theta)^2(1+|\rho|) \leq 4.$$

The absence of calendar arbitrage is ensured by the following set of conditions

$$\partial_t \theta \geq 0,\ \partial_t \Delta \geq 0 \text{ and } \Delta_t \geq 0 \quad \text{for all } t \geq 0,$$
$$0 \leq \partial_\theta(\theta\phi(\theta)) \leq \tfrac{1}{\rho^2}\left(1 + \sqrt{1-\rho^2}\right)\phi(\theta) \quad \text{for all } \theta > 0.$$

Our choice of parametric global SSVI

We used the following functional forms for $\Delta(t)$, $\theta(t)$ and $\phi(\theta)$:

$$\Delta(t) = Ct, \quad \phi(\theta) \equiv \eta\theta^{-\gamma} \quad \text{and} \quad \theta(t) = Kt \quad \text{with} \quad C, \eta, K > 0,\ \gamma \in (0,1) \text{ and } \rho \in (-1,1).$$

Eventually, the global implied volatility surface is determined by 5 parameters, $(C, K, \rho, \eta, \gamma)$. It is clearly not sufficient for an accurate calibration of the implied volatility surface but it can already account for its general shape. In [14], Gatheral and Jacquier start from this parametric form, and then achieve a better fit by relaxing the rigid link between different maturity slices. We don't need to do so because we only use this parametric shape as a base model in our spline-based Radon-Nikodym calibration procedure.

Numerical experiments with SSVI global prior model

In Figure 9, we display the results of the global calibration to the vanilla option prices on the Dow Jones index as of May, 28th 2014.

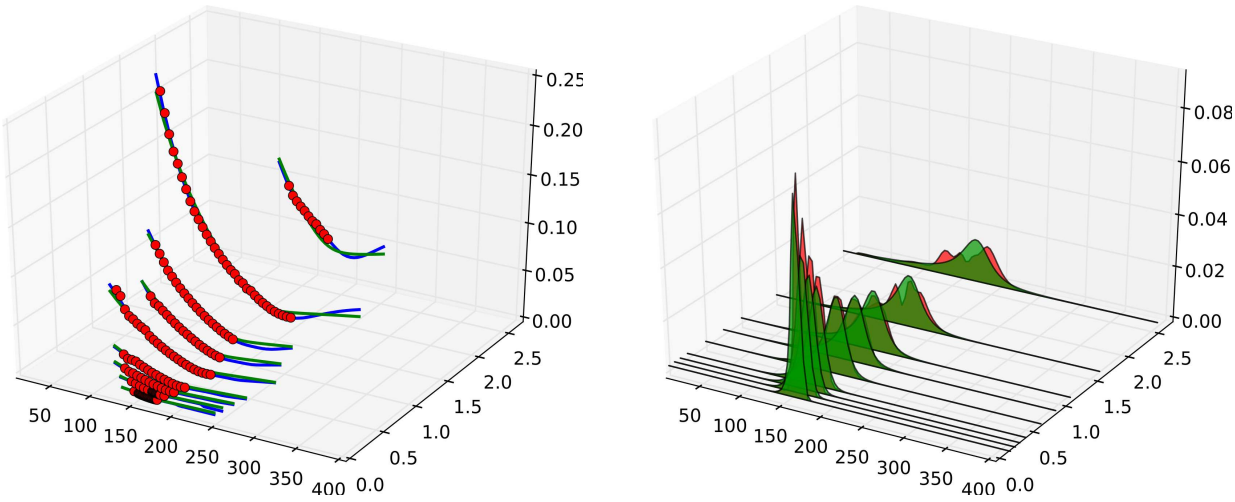


Figure 9: Calibration of Radon-Nikodym densities from sparse vanilla option data on the Dow-Jones index. On the left-hand side, we plot the sequences of implied total lognormal variances (Input market data (red circles), interpolation (blue lines) and prior global SSVI fit (green lines)). On the right-hand side, we show the calibrated risk-neutral densities for the same maturities. The parameters for the roughly calibrated SSVI form are: $C = 0.0$, $\rho = -0.55$, $K = 0.019$, $\eta = 2.72$ and $\gamma = 0.25$.

Singular measures

Even though the method proposed here relies on parameterizing the Radon-Nikodym derivatives of the marginals of the risk-neutral probability with respect to a prior model, it does not correspond to a global continuous change of measure. Indeed, if one calibrates (*e.g.*) a local volatility model based on our interpolation, it will have a different volatility process from the prior model, while a continuous change of measure would only result in a change of drift. Therefore, the corresponding change of measure is singular.

5 Kolmogorov forward PDE for the Radon-Nikodym derivative

As we have seen earlier, parameterizing the Radon-Nikodym derivative with respect to a simple base model allows to account for a prior belief on the risk-neutral distribution tails. If we believe in fat tails, we can use a fat-tailed prior model and the calibrated risk-neutral density will have the same rate of decay as the prior density. Moreover, as it is a linear transformation of the price space, conditions of absence of arbitrage remain linear conditions.

In this section, we show that the Radon-Nikodym derivative is also a convenient space for the numerical treatment of the Kolmogorov forward PDE with the Galerkin method using B-spline finite elements.

We consider the case of a simple S.D.E. $dS_t = \sqrt{v(t,S_t)}dW_t$ for some instantaneous variance $v(t,x)$ and deterministic initial condition S_0. The base model is assumed to be driven by the S.D.E. $dS_t = \sqrt{v_0(t,S_t)}dW_t$ where $v_0(0,S_0) = v(0,S_0)$. The Kolmogorov forward equation for the risk-neutral density $\phi(t,x)$ and the prior density ϕ_0 are

$$\partial_t \phi = \frac{1}{2}\partial^2_{xx}\left(v\phi\right) \qquad \text{and} \qquad \partial_t \phi_0 = \frac{1}{2}\partial^2_{xx}\left(v_0\phi_0\right).$$

If we denote by $f(t,x) := \frac{d\mathbb{Q}_t}{d\mathbb{Q}^0_t} = \frac{\phi(t,x)}{\phi_0(t,x)}$ the Radon-Nikodym derivative of the risk-neutral probability distribution of S_t with respect to the base model, we obtain a linear PDE for f,

$$\partial_t f = \frac{1}{2}\frac{\partial^2_{xx}(fv\phi_0)}{\phi_0} - f\frac{1}{2}\frac{\partial^2_{xx}(v_0\phi_0)}{\phi_0}, \qquad \text{and} \qquad f(0,x) \equiv 1. \tag{21}$$

The main benefit of this formulation with respect to the usual Kolmogorov forward PDE is that the initial condition is constant rather than being a Dirac mass. Hence, we do not need to tackle the numerical approximation of the Dirac mass with the basis functions. The formulation in terms of Radon-Nikodym derivative allows us to work with more regular functions.

In the special case of a Bachelier base model (*i.e.*, constant instantaneous variance v_0), $\phi_0(t,x) = \frac{1}{\sqrt{2\pi v_0 t}}\exp\left(-\frac{(x-S_0)^2}{2v_0 t}\right)$ and the terms $\frac{\partial_x(\phi_0)}{\phi_0}$ and $\frac{\partial_{xx}(\phi_0)}{\phi_0}$ involved in PDE (21) are polynomials of order 1 and 2 respectively.

Variational formulation

We first derive a spatially weak formulation of PDE (21). Multiplying by a test function u of the space variable x and integrating over $\mathbb{R}$, we get

$$\partial_t \int fu = \frac{1}{2}\int v\partial^2_{xx}(f)u + \int\left(\frac{\partial_x(v\phi_0)}{\phi_0}\right)\partial_x(f)u + \int\frac{1}{2}\left(\frac{\partial^2_{xx}((v-v_0)\phi_0)}{\phi_0}\right)fu.$$

Using an integration by parts, we get the weak formulation of the PDE

$$\partial_t \int fu = -\frac{1}{2}\int v\partial_x(f)\partial_x(u) + \int \underbrace{\left(\frac{\partial_x(\phi_0)}{\phi_0} + \frac{1}{2}\partial_x(v)\right)}_{=:c}\partial_x(f)u + \int \underbrace{\frac{1}{2}\left(\frac{\partial^2_{xx}((v-v_0)\phi_0)}{\phi_0}\right)}_{=:e}fu. \tag{22}$$

B-spline finite elements

We plug a B-spline or order n, $f(t,x) = \sum_{j=0}^{k+n} w_j(t)b^\Gamma_{j,n}(x)$ into Equation (22) and we obtain

$$\partial_t \sum_{j=0}^{k+n+1} w_j(t)\int b^\Gamma_{j,n}u = -\frac{1}{2}\sum_{j=0}^{k+n+1} w_j(t)\int v\partial_x\left(b^\Gamma_{j,n}\right)\partial_x(u) + \sum_{j=0}^{k+n+1} w_j(t)\int c\partial_x\left(b^\Gamma_{j,n}\right)u + \sum_{j=0}^{k+n+1} w_j(t)\int eb^\Gamma_{j,n}u,$$

that is, $\partial_t \sum_{j=0}^{k+n+1} w_j(t)\int b^\Gamma_{j,n}u = \sum_{j=0}^{k+n+1} w_j(t)\left(-\frac{1}{2}\int v\partial_x\left(b^\Gamma_{j,n}\right)\partial_x(u) + \int c\partial_x\left(b^\Gamma_{j,n}\right)u + \int eb^\Gamma_{j,n}u\right)$. We plug $u = b^\Gamma_{i,n}$ for $n < i < k$ (the basis functions of compact support) into (22), we get a matrix equation

$A\partial_t w = B(t)w$, where A is defined by $A_{ij} = \int b^\Gamma_{j,n} b^\Gamma_{i,n}$, and is a band matrix independent of t, and B is defined by

$$B(t)_{ij} = -\frac{1}{2}\int v\partial_x \left(b^\Gamma_{j,n}\right)\partial_x(b^\Gamma_{i,n}) + \int c\partial_x \left(b^\Gamma_{j,n}\right) b^\Gamma_{i,n} + \int e b^\Gamma_{j,n} b^\Gamma_{i,n}, \qquad n < i < k,$$

which is time-dependent (because of the time dependence of v, c and e.) If p is the truncation order, we need to have as much as $2(p+1)$ additional linear equations or boundary conditions to obtain a full-rank system. For example, in the case of a flat extrapolation ($p = 0$), we need two boundary conditions. We can use the integral constraints

$$\int_{\mathbb{R}} f(x)\phi_0(x)dx = 1 \qquad \text{and} \qquad \int_{\mathbb{R}} xf(x)\phi_0(x)dx = S_0. \tag{23}$$

For higher order extrapolation, additional boundary conditions must be added to the system. In the end, we obtain a $(k + 2p - n + 1)$-dimensional ordinary differential equation. This can be solved using an explicit or implicit time stepping scheme.

Integral constraints, stability and singly bordered band diagonal linear systems

In place of boundary conditions, the integral constraints (23) do not preserve the banded form of the system. The eventual system is *singly bordered band diagonal*. Still, using a specialization of Gauss elimination to this specific case, we can take advantage of the sparsity with a $O(nm^2)$ complexity. We refer to [15] for more details on Gauss elimination for sparse matrices.

Furthermore, in the stepping algorithm, rather than explicitly solving the linear system, one can use a least-square minimization under firm non-negativity constraints on the B-spline coefficients for the Radon-Nikodym derivative. In this setting, the time stepping is a quadratic program, which is slightly more expansive than solving the linear system. The upside is that together with the integral conditions (23), the non-negativity constraint forces the result to be a valid Radon-Nikodym density, and therefore to have an $L^1(\mathbb{P}_X)$ norm equal to 1 at every time step. In this setting, both the explicit and implicit versions of the scheme are unconditionally stable.

The time-independent case

A PDE of the form $\partial_t f = \mathcal{L}_x f$ where $\mathcal{L}_x$ is a linear differential operator independent of t can be seen as an infinite-dimensional ordinary differential equation whose solution is directly given by the exponential of the operator $\mathcal{L}_x$: $f(t,\cdot) \equiv \exp(\mathcal{L}_x) f_0(\cdot)$ where $f_0(\cdot) \equiv f(0,\cdot)$ is the initial condition. A space-discretized version of the PDE can be solved in the same way with a matrix exponential. This has been used for option pricing with time-independent local volatility models by Albanese and Trovato in [3]. However, even if the local variance v is time-independent, the forward PDE for the Radon-Nikodym derivative (21) is not of the required form, unless we use a time-independent base density ϕ_0.

6 Conclusion and future research

In this article, we have applied techniques based on shape-constrained B-splines for two problems arising in volatility modeling.

The first one is the application to the recently developed particle method for the calibration the leverage function in SLV models. The regression problems arising in this case were typically tackled by non-parametric regression. Using shape-constrained B-splines, we could exploit additional information coming from the knowledge of the marginal distributions in addition to the bivariate sample. This led us to an improved accuracy with the same sample size.

The second problem we tackled is the arbitrage-free interpolation of the implied volatility surface. In the price space, arbitrage conditions are simply formulated but naive interpolation methods yield unrealistic implied volatility surfaces. Interpolating in the implied volatility space poses challenges as non-arbitrage constraints on implied volatility are non-linear. In this paper, we have shown that using a prior roughly calibrated model, a B-spline parameterization of the Radon-Nikodym density allows

us to formulate the calibration problem as a second-order cone program and produces smooth implied volatility surfaces.

This article leaves some questions open. A first interesting problem is whether the necessary compatibility conditions (13a), (13b) and (13c) between the marginals and the conditional expectation are sufficient to ensure the existence of a corresponding joint distribution. Another subject which has not been addressed in this paper is the problem of optimal knot selection for B-spline regression in the context of the particle method, as knots were generally evenly spaced on a certain range of standard deviations. A new approach to adaptive knot selection for multivariate B-spline is the subject of an ongoing research article by the author.

A The Abergel-Tachet PDE

Let us consider the SDE (16) in the case where the stochastic volatility process a_t is itself a one-dimensional Itô process.

$$\begin{cases} dS_t = a_t l(t, S_t) dW_t, \\ da_t = \beta(t, a_t)dt + \nu(t, a_t)dW_t^\sigma, \\ d\langle W, W^\sigma \rangle_t = \rho dt. \end{cases}$$

Following Abergel and Tachet [2] (except that we use normal conventions for volatilities rather than lognormal), the Kolmogorov forward PDE for the joint density p of (S_t, a_t) is

$$\begin{cases} \partial_t p = \frac{1}{2}\partial^2_{xx}((ly)^2 p) + \rho\partial^2_{xy}((ly)\nu p) + \frac{1}{2}\partial^2_{yy}(\nu^2 p) - \partial_y(\beta p), \\ p(0, x, y) = \delta_{S_0, a_0}(x, y). \end{cases}$$

Defining $q := \int p dy$, we get $\partial_t q = \frac{1}{2}\partial^2_{xx}\left(l^2 \int y^2 p dy\right)$. Then, identifying the terms with the Dupire equation $\partial_t q = \frac{1}{2}\partial^2_{xx}\left(\sigma^2_{\text{Dup}} q\right)$, we recover the calibration condition (17), that is, $l^2 = \frac{\sigma^2_{\text{Dup}} \int p dy}{\int y^2 p dy}$. The PDE corresponding to the McKean stochastic integro-differential equation (18) is then

$$\begin{cases} \partial_t p = \frac{1}{2}\partial^2_{xx}\left(\frac{y^2 \int p dy}{\int y^2 p dy}\sigma^2_{\text{Dup}} p\right) + \rho\partial^2_{xy}\left(\frac{y\sqrt{\int p dy}}{\sqrt{\int y^2 p dy}}\sigma_{\text{Dup}}\nu p\right) + \frac{1}{2}\partial^2_{yy}(\nu^2 p) - \partial_y(\beta p), \\ p(0, x, y) = \delta_{S_0, a_0}(x, y). \end{cases}$$

The author is grateful to Daniel Andor, Bruno Dupire, Julien Guyon, Alexey Polishchuk and Stephen Taylor for fruitful discussions and their remarks and comments on this work.

References

[1] IBM ILOG CPLEX Optimizer. http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/, 2010.

[2] Frédéric Abergel and Rémi Tachet. A nonlinear partial integro-differential equation from mathematical finance. *Discrete and Continuous Dynamical Systems-Series A*, 27:907–917, 2010.

[3] Claudio Albanese and Manlio Trovato. Monetary policy risk and CMS spreads. *Preprint*, 2007.

[4] Erling D. Andersen and Knud D. Andersen. The MOSEK optimization software. http://www.mosek.com/.

[5] Martin S. Andersen, Joachim Dahl, and Lieven Vandenberghe. CVXOPT: A Python package for convex optimization, Version 1.1.6. Available at cvxopt.org, 2013.

[6] Jesper Andreasen and Brian Huge. Volatility interpolation. *Risk*, pages 76–79, 2012.

[7] Lorenzo Bergomi. Smile dynamics II. *Risk*, 18:67–73, 2005.

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

[9] Anthonie W. Van der Stoep, Lech A. Grzelak, and Cornelis W. Oosterlee. The Heston stochastic-local volatility model: efficient Monte Carlo simulation. *Preprint*, 2013.

[10] Bruno Dupire. Pricing with a smile. *Risk*, 7:18–20, 1994.

[11] Bruno Dupire. A unified theory of volatility. *Paribas working paper*, 1996.

[12] Adrien Escande, Nicolas Mansard, and Pierre-Brice Wieber. Hierarchical quadratic programming. *Preprint*, 2012.

[13] Matthias R. Fengler. Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance*, 9(4):417–428, 2009.

[14] Jim Gatheral and Antoine Jacquier. Arbitrage-free S.V.I. volatility surfaces. *Preprint*, 2012.

[15] Gene H. Golub and Charles F. Van Loan. *Matrix Computations (3rd Ed.)*. The Johns Hopkins University Press, Baltimore, MD, USA, 1996.

[16] Julien Guyon. Local correlation families. *Risk*, 2014.

[17] Julien Guyon and Pierre Henry-Labordère. Being particular about calibration. *Risk*, pages 92–97, 2012.

[18] István Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probability theory and related fields*, 71:501–516, 1986.

[19] Patrick S. Hagan, Deep Kumar, Andrew S. Lesniewski, and Diana E. Woodward. Managing smile risk. *Wilmott magazine*, 2002.

[20] Xuming He and Pin Ng. COBS: Qualitatively constrained smoothing via Linear programming. *Computational Statistics*, 14(3):315–337, 1999.

[21] Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.

[22] Márcio P. Laurini. Imposing no-arbitrage conditions in implied volatility surfaces using constrained smoothing splines. *Applied Stochastic Models in Business and Industry*, 27(6):649–659, 2011.

[23] Dilip Madan, Yong Ren, and Michael Qian Qian. Calibrating and pricing with embedded local volatility models. *Risk*, 2007.

[24] Ana Margarida Monteiro, Reha H. Tütüncü, and Luís N. Vicente. Recovering risk-neutral probability density functions from options prices using cubic splines. *European Journal of Operational Research*, 187(2):525–542, 2008.

[25] David Papp and Farid Alizadeh. Shape constrained estimation using nonnegative splines. *Journal of Computational and Graphical Statistics*, (to appear), 2012.

[26] Larry L. Schumaker. *Spline Functions: Basic Theory, 2nd Edition*. Cambridge Mathematical Library, 2007.

[27] Louis O. Scott. Option pricing when the variance changes randomly: theory, estimation, and applications. *Journal of Finance and Quantitative Analysis*, 22(4):419–438, 1987.

[28] Alan H. Vermeulen, Richard H. Bartels, and Glenn R. Heppler. Integrating products of B-splines. *SIAM Journal on Scientific and Statistical Computing*, 13(4):1025–1038, 1992.