

MONTE CARLO APPROXIMATIONS OF AMERICAN OPTIONS THAT PRESERVE MONOTONICITY AND CONVEXITY

Pierre del Moral, Bruno Rémillard, Sylvain Rubenthaler

▶ **To cite this version:**

Pierre del Moral, Bruno Rémillard, Sylvain Rubenthaler. MONTE CARLO APPROXIMATIONS OF AMERICAN OPTIONS THAT PRESERVE MONOTONICITY AND CONVEXITY. Numerical Methods in Finance, Bordeaux June 2011, Jun 2011, Bordeaux, France. pp.115-143, 10.1007/978-3-642-25746-9_4 . hal-00755423

HAL Id: hal-00755423

https://hal.univ-cotedazur.fr/hal-00755423

Submitted on 21 Nov 2012

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.

MONTE CARLO APPROXIMATIONS OF AMERICAN OPTIONS THAT PRESERVE MONOTONICITY AND CONVEXITY

PIERRE DEL MORAL, BRUNO RÉMILLARD, SYLVAIN RUBENTHALER

Abstract. It can be shown that when the payoff function is convex and decreasing (respectively increasing) with respect to the underlying (multidimensional) assets, then the same is true for the value of the associated American option, provided some conditions are satisfied. In such a case, all Monte Carlo methods proposed so far in the literature do not preserve the convexity or monotonicity properties. In this paper, we propose a method of approximation for American options which can preserve both convexity and monotonicity. The resulting values can then be used to define exercise times and can also be used in combination with primal-dual methods to get sharper bounds. Other application of the algorithm include finding optimal hedging strategies.

1. Introduction

Evaluation of American options is a central problem in financial engineering. Many ways of tackling the problem have been proposed so far, mostly for options on a single asset. One of the first method was proposed by Brennan and Schwartz (1977), and is based on numerical solutions of partial differential equations. Then tree-based methods were first introduced by Cox et al. (1979) for American options. Although Monte Carlo methods were proposed by Boyle (1977) for European options, it seems that the first simulation-based method is the one of Tilley (1993). After that paper, Monte Carlo methods increased in popularity and many papers improved Tilley's results, e.g. Carriere (1996), Broadie and Glasserman (1997), Longstaff and Schwartz (2001), Broadie and Glasserman (2004). Recently, two-stage Monte Carlo methods were developed by Rogers (2002), Andersen and Broadie (2004) and Haugh and Kogan (2004). For an interesting review of Monte Carlo methods, see Fu et al. (2001).

Date: October 30, 2010.

Key words and phrases. American option, American option, simulation, exercise region, Snell envelope.

Funding in partial support of this work was provided by the Natural Sciences and Engineering Research Council of Canada and the Fonds québécois de la recherche sur la nature et les technologies.

However, even if in some interesting cases, e.g. American call-on-max or American put-on-max options of dividend paying assets, where the underlying assets follows geometric Brownian motions, the value of an American option possesses convexity or monotonicity properties, the Monte Carlo methods described above do not preserve these properties. For a related work, see Laprise et al. (2006).

In what follows, a Monte Carlo approach is proposed so that convexity and monotonicity are preserved. It is shown that the algorithm is quite precise for American options on a small number of assets. When the number of assets is large then the proposed method can be used to implement primal-dual methods, as in Andersen and Broadie (2004).

In Section 2, one states the definition of the Snell envelope, together with a review of existing methods for pricing American options, in particular the methods of Carriere (1996), Longstaff and Schwartz (2001) and Andersen and Broadie (2004). Properties of options prices in terms of assets dynamics and payoff functions are then studied in Section 3.1, and the algorithm for pricing an American option is described in Section 3.2. Finally, implementation issues are presented in Section 4, and a brief discussion of the results is provided in Section 5.

2. A brief review of algorithms for valuation of American options

As it is usually assumed, one wants to calculate the value of a Bermudan option, that is an option that can be exercised at given fixed periods, instead of a real American option, which can be exercised anytime. For simplicity, assume that the exercise periods are $0, 1 \ldots n$.

Suppose that, under the risk neutral measure, the value of the assets, including possibly stochastic volatility factors, is modeled by a discrete time Markov chain $(X_k)_{k\geq 0}$, and that the actualized payoff Z_k at period k, given by $Z_k = f_k(X_k)$, is integrable for any $k \in \{0, \ldots, n\}$. For example, if the interest rate is constant and given by r, then

$$f_k(x) = f_k\left(x^{(1)}, \ldots, x^{(d)}\right) = e^{-rk} \max\left(\max_{1\leq j\leq d} x^{(j)} - K, 0\right)$$

is the actualized payoff of a so-called call-on-max option.

In what follows, one first states the solution of the American option problem, in terms of the Snell envelope in the Probability literature. In Operations Research, such a problem is called a dynamic programming problem. Then one discusses some deterministic and stochastic algorithms for calculating the Snell envelope or the exercise region.

2.1. **Snell envelope.** For a given filtration $\mathbb{F}$, the value $U_0(x)$ at period 0 of the "American" option with $X_0 = x$, is given by

$$U_0(x) = \sup_{\tau \in \mathcal{T}_{0,n}} E(Z_\tau | X_0 = x),$$

where $\mathcal{T}_{k,n}$ stands for the set of all $\mathbb{F}$ stopping times with values in $\{k, \ldots, n\}$, $0 \leq k \leq n$. The sequence of functions

$$U_k(x) = \sup_{\tau \in \mathcal{T}_{k,n}} E(Z_\tau | X_k = x), \qquad k \in \{0, \ldots, n\},$$

is called the Snell envelope.

It is well-known, e.g. Neveu (1975), that the Snell envelope can be calculated in a recursive way. In fact, $U_n \equiv V_n \equiv f_n$, and

$$\begin{array}{rcll} V_k(x) & = & E\{U_{k+1}(X_{k+1}) | X_k = x\}, & \\ U_k(x) & = & f_k(x) \vee V_k(x) = \max\{f_k(x), V_k(x)\}, & k = n-1, \ldots, 0. \end{array}$$

Note that U_k also has the representation

$$U_k(x) = E(Z_{\tau_k^*} | X_k = x),$$

where $\tau_k^* = \min\{j \geq k\,;\, f_j(X_j) = U_j(X_j)\}$, $k = 0, 1, \ldots, n$. Also

$$\tau_k^* = \min\{j \geq k\,;\, f_j(X_j) \geq V_j(X_j)\}, \quad k = 0, 1, \ldots, n.$$

An equivalent approach to solving the problem of American options is to look at the exercise region $\mathcal{E}$, defined as

$$\mathcal{E} = \{(k, x)\,;\, f_k(x) = U_k(x)\} = \{(k, x)\,;\, f_k(x) \geq V_k(x)\}.$$

In fact, knowing $\mathcal{E}$ suffices to evaluate the option, since in that case, one can obtain the optimal stopping times τ_k^* viz.

$$\tau_k^* = \min\{j \geq k\,;\, (j, X_j) \in \mathcal{E}\}, \quad k = 0, 1, \ldots, n.$$

In some situations, the boundary of the exercise region is quite simple, e.g. Broadie and Detemple (1997). For example, for the American call option in a Black-Scholes setting,

$$\mathcal{E} = \bigcup_{k=0}^{n} \{k\} \times [e_k, \infty),$$

for some $e_0, \ldots, e_n$, since $(k, x) \in \mathcal{E}$ implies $(k, x') \in \mathcal{E}$, for any $x' \geq x$. Therefore it suffices to know the e_k's to evaluate the option.

The boundary of the exercise region for the American put option in a Black-Scholes setting is also quite simple, and is given by

$$\mathcal{E} = \bigcup_{k=0}^{n} \{k\} \times [0, e_k],$$

for some $e_0, \ldots, e_n$, since $(k, x) \in \mathcal{E}$ implies $(k, x') \in \mathcal{E}$, for any $x' \leq x$. However, in general, the exercise region is more complex. For more details, see Broadie and Detemple (1997).

2.2. **Classes of algorithms for valuation of American options.** Algorithms for evaluating American options can be divided in three classes.

(i) Tree-based methods: For these algorithms, values of U are estimated at a finite number of points, possibly random. In most cases, the underlying process X is replaced by a finite state Markov chain $\left(\hat{X}_k\right)_{k\geq 0}$. The main representatives are the algorithms based on the binomial and multinomial methods, initiated by Cox et al. (1979), and more recently the finite Markov chain methods, e.g. Duan and Simonato (2001), Duan et al. (2004), **?** and Bally et al. (2005). In the latter, a powerful optimal method for choosing the approximation, known as quantization, is discussed. See also Kargin (2005) for a similar approach that can be easily adapted

to Bermudan and American options. Broadly speaking, the stochastic method proposed by Broadie and Glasserman (2004) can be seen as a random version of the above algorithms. In all cases, theses algorithms are designed to approximate the value of $U_0(x)$ for a fixed x, not for all x.

(ii) Snell's method: In these algorithms, the Snell envelope U is calculated. The main representatives are those based on partial differential equations, e.g. Brennan and Schwartz (1977), and Carriere's algorithm, e.g. Carriere (1996), which is a simulation-based method. The latter will be described in more details below.

(iii) Exercise region method: For these algorithms, $\mathcal{E}$ is estimated. The first algorithm in this direction appeared in Tilley (1993) for options based on a single asset. It was later improved by Longstaff and Schwartz (2001). Almost immediately, several papers appeared that validated the Longstaff-Schwartz method, e.g. Clément et al. (2002). Other exercise region methods include the "primal-dual" methods of Rogers (2002) and Andersen and Broadie (2004). These methods are also discussed in more details below. For other methods for a single asset based on the exercise boundary, see e.g. Broadie et al. (2000).

The (random) algorithms of Carriere (1996) and Longstaff and Schwartz (2001), all have in common the following features. They both rely on the simulation of a large number of paths, and conditional expectations given $\mathcal{F}_k$ are approximated by functions in some given classes $\mathcal{A}_k$, $k = 0, \ldots, n-1$.

These algorithms, together with the dual approach algorithm, e.g. Rogers (2002) and Andersen and Broadie (2004), are detailed next.

2.3. **Carriere algorithm.** Start by simulating N paths, denoted by

$$\mathbf{S}_1 = (X_{1,0}, \ldots, X_{1,n}), \ldots, \mathbf{S}_N = (X_{N,0}, \ldots, X_{N,n}).$$

Then, based on approximating classes $\mathcal{A}_0, \ldots, \mathcal{A}_{n-1}$, estimate V by $\hat{V}$ and U by $\hat{U}$ using local regression methods. More precisely, set $\hat{V}_n = \hat{U}_n = f_n$, and then, for all $n-1, \ldots, 0$, let $\hat{V}_k$ and $\hat{U}_k$ be defined by

$$\hat{V}_k = \arg\min_{g \in \mathcal{A}_k} \sum_{i=1}^{N} \Big\{ \hat{U}_{k+1}(X_{i,k+1}) - g(X_{i,k}) \Big\}^2, \quad \hat{U}_k = f_k \vee \hat{V}_k, \tag{2.1}$$

provided the solution of the minimization problem (2.1) exists.

In Carriere (1996), $\mathcal{A}_k$ is the set of all q-splines.

Remark 2.1. In general, if $\mathcal{A}_k$ consists of all possible linear combinations of a finite number d_k of basis functions, then $\hat{V}_k$ can be found using regression, and the computations can be quite fast. Unfortunately, if d_k is small compared to N, then the approximation $\hat{V}_k$ of V_k can be poor. However if V_k is convex or monotone, $\hat{V}_k$ will not inherit the same properties.

2.4. **Longstaff-Schwartz algorithm.** Using approximating classes $\mathcal{A}_0$, …, $\mathcal{A}_{n-1}$, estimate $\mathcal{E}$ by $\widehat{\mathcal{E}}$, where $\widehat{\mathcal{E}} = \bigcup_{k=1} k \times E_k$ and $E_k = \{x\,;\, f_k(x) \geq \tilde{V}_k(x)\}$, with $\tilde{V}_n = f_n$, $\hat{\tau}_n \equiv n$, and for all $k = n-1, \ldots, 0$,

$$\tilde{V}_k = \arg\min_{g \in \mathcal{A}_k} \sum_{i=1}^{N} \left\{ f_{\hat{\tau}_{k+1}}(X_{i,\hat{\tau}_{k+1}}) - g(X_{i,k}) \right\}^2, \tag{2.2}$$

where for any path $\mathbf{S} = (X_0, \ldots, X_n)$,

$$\hat{\tau}_k(\mathbf{S}) = \begin{cases} k & \text{if} \quad f_k(X_k) \geq \tilde{V}_k(X_k), \\ \hat{\tau}_{k+1}(\mathbf{S}) & \text{if} \quad f_k(X_k) < \tilde{V}_k(X_k), \end{cases}$$

provided the solution of the minimization problems (2.2) exist. Then, the estimation of $U_0(X_0)$ is given by

$$\hat{U}_0(X_0) = \frac{1}{N} \sum_{i=1}^{N} f_{\hat{\tau}_0}(X_{i,\hat{\tau}_0}). \tag{2.3}$$

As before, the same conclusions as those in Remark 2.1 apply.

Remark 2.2. The Carriere and Longstaff-Schwartz algorithms may seem equivalent, but they are not. In the latter one, emphasis is on stopping times. Instead of estimating U_k by

$f_k \vee \tilde{V}_k$, as in Carriere's algorithm, $U_k(X_{i,k})$ is estimated by

$$\hat{U}_k(X_{i,k}) = \frac{1}{N}\sum_{i=1}^{N} f_{\hat{\tau}_k}(X_{i,\hat{\tau}_k}), \qquad k \in \{0,\dots,n\}. \tag{2.4}$$

So, in the Longstaff-Schwartz algorithm, U_k is estimated at sample points, not all points.

2.5. **Primal-dual approach.** As seen before, to approximate the value of an American option, one can estimate directly $\mathcal{E}$, as proposed in Section 2.4, or approximate first U or V, and then approximate $\mathcal{E}$, as proposed in Section 2.3 or Section 3.2.

In Andersen and Broadie (2004), the authors proposes a two-stage approach: one for obtaining a lower bound for $U_0(X_0)$ and the other one for an upper bound. Suppose that $\hat{\mathcal{E}}$ is an estimation of $\mathcal{E}$, e.g. $\hat{\mathcal{E}}$ is obtained by the methods discussed previously.

2.5.1. *Algorithm for the lower bound.*

A1: For each $i = 1,\dots,N_0$, simulate a path $\mathbf{S}_i = (X_{i,0},\dots,X_{i,n})$ starting from X_0;

A2: Define the entrance time of $\mathbf{S}_i$ in $\widehat{\mathcal{E}}$ as

$$\tau_{\widehat{\mathcal{E}},i} = \min\left\{ j \geq 0\,;\, (j, X_{i,j}) \in \widehat{\mathcal{E}} \right\};$$

A3: Calculate

$$\overline{LB} = \frac{1}{N_0}\sum_{i=1}^{N_0} f_{\tau_{\widehat{\mathcal{E}},i}}\left(X_{i,\tau_{\widehat{\mathcal{E}},i}}\right). \tag{2.5}$$

Then $\overline{LB}$ is a pointwise estimation of a lower bound for $U_0(X_0)$. One could also calculate the standard deviation $\hat{\sigma}_{LB}$ associated with the data $LB_i = f_{\tau_{\widehat{\mathcal{E}},i}}\left(X_{i,\tau_{\widehat{\mathcal{E}},i}}\right)$, $1 \leq i \leq N_0$.

One can assume that $(0, X_0) \notin \widehat{\mathcal{E}}$, for otherwise, $\hat{U}_0(X_0) = f_0(X_0)$.

Remark 2.3. As emphasized by Andersen and Broadie (2004), the last procedure yields a lower bound for $U_0(X_0)$. In fact, even if one can approximate U and V, to estimate $U_0(X_0)$ more precisely, it is recommended to calculate (2.5) in addition, since errors in U or $\widehat{\mathcal{E}}$ often transform in smaller errors for $\tau_{\widehat{\mathcal{E}}}$, as many simulations showed. Unfortunately, so far there is no mathematical justification of that property.

For the upper bound, Rogers (2002) showed that for any martingale M,

$$U_0(X_0) \leq M_0 + E\left[\max_{0\leq k\leq n}\{f_k(X_k) - M_k\}\right],$$

with equality for the (unique) martingale M associated with the Doob-Meyer decomposition of the supermartingale $U_k(X_k)$.

Remark 2.4. In fact, the starting point of the martingale is not important. For simplicity, one could restrict ourselves to martingales starting from $M_0 = 0$. In that case, the equality holds for $M_k = M'_k - M'_0$, where M' is the (unique) martingale M associated with the Doob-Meyer decomposition of the supermartingale $U_k(X_k)$.

Based on $\widehat{\mathcal{E}}$, Andersen and Broadie (2004) suggests a Monte Carlo algorithm to construct a martingale M, close to the optimal Doob-Meyer martingale. For all $k \in \{0,\ldots,n\}$, let l_k be the indicator function of the set $\left\{(k, X_k) \in \widehat{\mathcal{E}}\right\}$, let $\tau_{\widehat{\mathcal{E}},k} = \min\{j \geq k; (j, X_j) \in \widehat{\mathcal{E}}\}$ and define $L_k = E\left(Z_{\tau_{\widehat{\mathcal{E}},k}}|\mathcal{F}_k\right)$. Further set set $M_0 = L_0 = \hat{U}_0(X_0)$, and for any $1 \leq k \leq n$, set

$$M_k = M_{k-1} + L_k - L_{k-1} - l_{k-1}E\left(L_k - L_{k-1}|\mathcal{F}_{k-1}\right).$$

Note that the last equation can be written in the much simpler form

$$M_k = M_{k-1} + L_k - E(L_k|\mathcal{F}_{k-1}) \tag{2.6}$$

$$= M_{k-1} + E\left\{f_{\tau_k}(X_{\tau_k})|\mathcal{F}_k\right\} - E\left\{f_{\tau_k}(X_{\tau_k})|\mathcal{F}_{k-1}\right\}. \tag{2.7}$$

Since the conditional expectations must be estimated, they suggest a procedure to approximate M. However, based on (2.6)–(2.7), the following modification seems more natural.

For each simulated path $\mathbf{S}_i$, $i \in \{1,\ldots,N_1\}$, set $M_{i,0} = 0$, and repeat the following steps, for each $k = 1,\ldots,n$:

B1:
- For each $j = 1,\ldots,N_2$, simulate a new sub-path $\mathbf{S}_{i,j} = \{X_{i,j,\ell}\}_{\ell=k}^{n}$, starting from $X_{i,k}$;
- Calculate $\hat{L}_{i,k} = \frac{1}{N_2}\sum_{j=1}^{N_2} f_{\tau_{\widehat{\mathcal{E}},k}}\left(X_{i,j,\tau_{\widehat{\mathcal{E}},k}}\right)$;

B2: – For each $j = 1, \ldots, N_3$, simulate a new sub-path $\mathbf{S}_{i,j} = \{X_{i,j,\ell}\}_{\ell=k-1}^{n}$, starting from $X_{i,k-1}$;

– Calculate $\hat{E}(L_{i,k}|\mathcal{F}_{k-1}) = \frac{1}{N_3}\sum_{j=1}^{N_3} f_{\tau_{\widehat{\mathcal{E}},k}}\left(X_{i,j,\tau_{\widehat{\mathcal{E}},k}}\right)$;

B3: Set $M_{i,k} = M_{i,k-1} + \hat{L}_{i,k} - \hat{E}(L_{i,k}|\mathcal{F}_{k-1})$.

Finally, the upper bound $\overline{UB}$ for $U_0(X_0)$ is then approximated by

$$\overline{UB} = \frac{1}{N_1}\sum_{i=1}^{N_1}\left[\max_{0\le k\le d}\{f_k(X_{i,k}) - M_{i,k}\}\right]. \tag{2.8}$$

One could also calculate the standard deviation $\hat{\sigma}_{UB}$ associated with the data $UB_i = \max_{0\le k\le d}\{f_k(X_{i,k}) - M_{i,k}\}$, $1 \le i \le N_1$.

Because the lower and upper bounds estimations are conditionally independent given $\widehat{\mathcal{E}}$, a 95% percent confidence interval for U_0 is given by

$$\left[\overline{LB} - 1.9545\frac{\hat{\sigma}_{LB}}{\sqrt{N_0}}, \overline{UB} + 1.9545\frac{\hat{\sigma}_{UB}}{\sqrt{N_1}}\right]. \tag{2.9}$$

Remark 2.5. In Andersen and Broadie (2004), it is argued that the upper bound constructed by their Monte Carlo methods will always be greater that the Monte Carlo lower bound, i.e., the value of (2.5) should be smaller than the value of (2.8). If true, their reasoning should apply as well to the modified algorithm described above. However their proof is clearly incorrect[1]. As shown in Tables 1–2, it is possible that the (pointwise) Monte Carlo upper bound is smaller that the (pointwise) Monte Carlo lower bound.

3. Approximation of the Snell envelope

From now on, assume that the Markov chain $(X_k)_{k\ge 0}$ takes values in a convex subset $\mathfrak{X}$ of $[0,\infty)^d$. Note that this setting excludes finite state space Markov chains, which can be treated much more easily. As customary, when $s, t \in \mathbb{R}^d$, $s \le t$ means that $s^{(j)} \le t^{(j)}$ for all $j = 1, \ldots, d$. For $x \in \mathfrak{X}$, let $\|x\| = \max_{1\le i\le d}\left|x^{(i)}\right|$. Finally, recall that a real valued

[1]If true, their argument would lead to the following (false) result: if $(\epsilon_k)_{k=1}^n$ are martingale differences and if τ is a random time with values in $\{1, \ldots, n\}$, τ not necessarily a stopping time, then $E(\epsilon_\tau) = 0$.

function g defined on $\mathfrak{X}$ is non decreasing if for any $x, x' \in \mathfrak{X}$, $x \leq x'$ implies $g(x) \leq g(x')$; conversely, g is non increasing if $-g$ is non decreasing.

Throughout the rest of the section, one has to make some hypotheses. The first assumption is related to the law of the Markov chain.

Assumption 1. *For all $1 \leq k \leq n$, X_k has the following representation*

$$X_k = \pi_k(X_{k-1}, Y_k), \quad Y_k \in \mathcal{Y}, \tag{3.1}$$

where Y_k has law μ_k, is independent of $\mathcal{F}_{k-1}$, and $x \mapsto \pi_k(x, y)$ is continuous on $\mathfrak{X}$, for any fixed $y \in \mathcal{Y}$.

Since Monte Carlo simulations play an essential role in the proposed methodology, the following assumption is needed too.

Assumption 2. *For each $k \in \{1, \ldots, n\}$, $f_k(X_k)$ is integrable.*

That assumption insures that the Snell envelope is well defined.

Assumption 3. *For all $0 \leq j \leq k \leq n$, the mappings $x \mapsto E\{f_k(X_k)|X_j = x\}$ and $x \mapsto f_k(x)$ are continuous.*

In the next section, desirable properties like convexity and monotonicity are studied for American options. The main result is that when the payoff has nice properties, then they are inherited by U and V as well. Next, in Section 3.2, the proposed algorithm is described.

3.1. **Properties of U and V.** The first two propositions are about monotonicity and continuity.

Proposition 3.1. *Assume that for every $k \in \{0, \ldots, n\}$, f_k is non decreasing (respectively non increasing) and $\pi_k(\cdot, y)$ is non decreasing for any $y \in \mathcal{Y}$. Then, for every $k \in \{0, \ldots, n\}$, U_k and V_k are non decreasing (respectively non increasing).*

If in addition, f_k is continuous for any $k \in \{0, \dots, n\}$, then U_k and V_k are also continuous for all $k \in \{0, \dots, n\}$.

When the monotonicity properties do not hold for π_k, one needs another assumption to get the continuity of the Snell envelope.

Proposition 3.2. *Under Assumptions 2 and 3, V_k and U_k are continuous for all $k \in \{0, \dots, n\}$.*

The next proposition deals with convexity of the Snell envelope.

Proposition 3.3. *Assume that for every $k \in \{0, \dots, n\}$, f_k is convex and non decreasing, and $\pi_k(\cdot, y)$ is convex and non decreasing for any $y \in \mathcal{Y}$. Then U_k and V_k are convex and non decreasing, for all $k \in \{0, \dots, n\}$.*

If f_k is convex and non increasing, and if $\pi_k(\cdot, y)$ is concave and non decreasing for any $y \in \mathcal{Y}$ and any $k \in \{0, \dots, n\}$, then U_k and V_k are also convex and non increasing, for all $k \in \{0, \dots, n\}$.

The proofs of Propositions 3.1 and 3.3 follow readily from Proposition A.1, since convexity, continuity and monotonicity are preserved by taking the max, that is, if g and h are convex and non decreasing (respectively non increasing), then $\max(g, h) = g \vee h$ is also convex and non decreasing (respectively non increasing). Clearly, if g and h are continuous, then $g \vee h$ is continuous. Proposition 3.2 is proven in Section B.1

3.2. **Description of the algorithm and justification.** Before describing the algorithm, one has to define what one means by a partition.

Definition 3.4. A partition $\mathcal{P}$ of a compact convex set K, is any finite set $\mathcal{P} = \{S_1, \dots, S_m\}$ of simplexes with disjoint non empty interiors, so that $K = \bigcup_{j=1}^{m} S_j$. The set of vertices of the partition $\mathcal{P}$ is denoted by $\mathcal{V}(\mathcal{P})$.

Note that K is then the convex hull generated by $\mathcal{V}(\mathcal{P})$.

The algorithm is based on Monte Carlo simulations, combined with a sequence of approximations on compact sets $R_0, \ldots, R_{n-1}$, determined by partitions $\mathcal{P}_0, \ldots, \mathcal{P}_{n-1}$.

The idea behind the algorithm is quite simple. Given approximations $\tilde{V}_k, \ldots, \tilde{V}_n$ of $V_k, \ldots, V_n$, one first get $\hat{V}_{k-1}$ by estimating V_{k-1} at every vertices $x \in \mathcal{V}(\mathcal{P}_{k-1})$, using Monte Carlo simulations, and then, one uses a linear interpolation of $\hat{V}_{k-1}$, to define $\tilde{V}_{k-1}$ at any point $x \in R_{k-1}$. Then $\tilde{V}_{k-1}$ may be extended to all of $\mathfrak{X}$ using projections.

More precisely, one may proceed through the following steps.

3.2.1. *Algorithm.*

C1: Set $\tilde{V}_n = \tilde{U}_n = \tilde{f}_n$, where $\tilde{f}_n = f_n$ on R_n and for any $x \in R_n^c$, $\tilde{f}_n(x) = \tilde{f}_n(x')$, where x' is the unique point in R_n so that $d(x, R_n) = d(x, x')$;

C2: For $k = n, \ldots, 1$, generate $\xi_{1,k}, \ldots, \xi_{N_k,k}$ according to μ_k;

C3: For every $x \in \mathcal{V}(\mathcal{P}_{k-1})$, define

$$\hat{V}_{k-1}(x) = \frac{1}{N_k} \sum_{i=1}^{N_k} \tilde{U}_k \left\{ \pi_k(x, \xi_{i,k}) \right\},$$

where $\tilde{U}_k = \max\left(\tilde{f}_k, \tilde{V}_k\right)$, and where $\tilde{f}_k = f_k$ on R_k and for any $x \in R_k^c$, $\tilde{f}_k(x) = \tilde{f}_k(x')$, where x' is the unique point in R_k so that $d(x, R_n k) = d(x, x')$;

C4: Interpolate linearly $\hat{V}_{k-1}$ over R_{k-1}, as in Definition C.1, and call it $\tilde{V}_{k-1}$.

For the precise meaning of linear interpolation and ways to implement it, see Section C.

Remark 3.1. First, contrary to most Monte Carlo algorithms, one does not simulate trajectories of the price process S. This is the key to preserve monotonicity and convexity, as illustrated in Figure 1.Next, in the case where $\mu_1 = \ldots, \mu_n = \mu$ and $N_1 = \ldots = N_n = N$, one could take the same set of random points $\xi_1, \ldots, \xi_N$ from distribution μ, thus reducing calculations. Last but not least, note that since f_k is not interpolated, which would induce extraneous errors, our algorithm cannot be interpreted as generating a tree-based algorithm.

As a by-product of the method, one obtains an estimation $\widehat{\mathcal{E}}$ of the exercise region $\mathcal{E}$. Such approximations are illustrated in Figures 4 and 6.

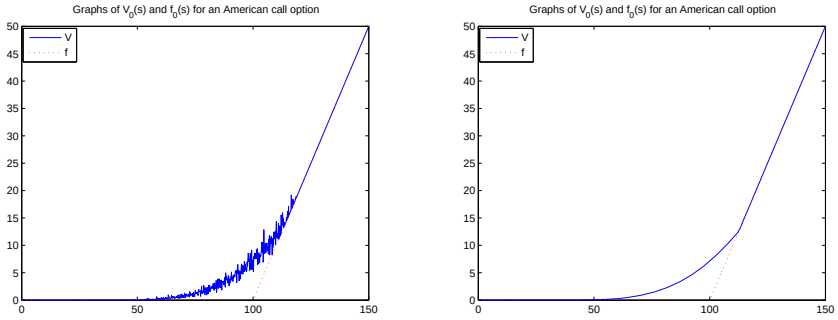


FIGURE 1. Full Monte Carlo method vs suggested Monte Carlo method with N=100

The following results show that the algorithm produces good uniform approximations of the Snell envelope U.

Theorem 3.5. *Suppose that for all $k \in \{0, \dots, n\}$, f_k, V_k and U_k are continuous. Let R_0 be a given compact convex subset of $\mathfrak{X}$. Let $\epsilon > 0$ be given. Then one can find compact convex sets $R_1, \dots, R_{n-1} \subset \mathfrak{X}$, partitions $\mathcal{P}_0, \dots \mathcal{P}_{n-1}$ generating respectively $R_0, \dots, R_{n-1}$, such that almost surely, one can find integers $N_{10}, \dots, N_{n0}$, so that for the simple interpolation method,*

$$\sup_{x \in \mathbb{R}_k} |U_k(x) - \tilde{U}_k(x)| = \|U_k - \tilde{U}_k\|_{R_k} \leq \|V_k - \tilde{V}_k\|_{R_k} < \epsilon,$$

whenever $N_1 \geq N_{10}, \dots, N_n \geq N_{n0}$.

The next result shows that using convex interpolations, as in Definition C.2, one can preserve the convexity of the Snell envelope.

Corollary 3.6. *Suppose in addition that for each k, f_k is monotone and $\pi_k(\cdot, y)$ is non decreasing for any fixed $y \in \mathcal{Y}$. If the quick interpolation method is used, then Theorem 3.5 holds true and the functions $\tilde{V}_k$ and $\tilde{U}_k$ have the same monotonicity as the U_k's and*

V_k's when restricted to R_k. If a convex interpolation on a grid is used instead of a linear interpolation, then the conclusion of Theorem 3.5 holds true and the functions $\tilde{V}_k$ and $\tilde{U}_k$ have the same monotonicity as the U_k's and V_k's when restricted to R_k. If in addition f_k is convex and non decreasing and $\pi_k(\cdot, y)$ is convex and non decreasing for any fixed y, then the U_k's and V_k's are convex when restricted to R_k. The same holds true if f_k is convex and non increasing and $\pi_k(\cdot, y)$ is concave and non decreasing, for any fixed y.

The proofs of these results are given is Appendix B.

4. Implementation issues

The first example of application is the classical Black-Scholes-Merton setting. Examples of calculations are provided for the case of an American call-on-max option on one and two assets. The second example is the N-GARCH model, as studied in Duan and Simonato (2001) and Duan et al. (2003). Examples of calculations are given for the American put option. In each numerical example, estimated lower and upper bounds are calculated, using the modified Andersen-Broadie algorithm described in Section 2.5.

4.1. **Geometric Brownian motion.** Suppose that under the risk neutral measure Q, the model satisfies

$$dX_i(t) = (r - \delta_i)X_i(t)dt + \sigma_i X_i(t)dW_i(t), \tag{4.1}$$

where $W_1, \ldots, W_d$ are dependent Brownian motions with correlation matric ρ. Setting $\alpha_i = \left(r - \delta_i - \sigma_i^2/2\right)\Delta$ and $\beta_i = \sigma_i \Delta^{1/2}$, then the conditional law of X_t given $X_{t-1} = x$ is log-normal, with mean vector $\log(x) + \alpha$, and covariance matrix C, where $\log(x)$ stands for the vector with components $\log(x_i)$, $1 \le i \le d$, and $C_{ij} = \beta_i \beta_j \rho_{ij}$, $1 \le i, j \le d$. It follows that Assumption 1 is satisfied with $Y_k \sim N(\alpha, C)$ and $\pi_k^{(j)}(x, y) = x_j e^{y_j}$, $j = 1, \ldots, d$. Furthermore, for each fixed $y = (y_1, \ldots, y_d)$, $\pi_k(\cdot, y)$ is continuous, non decreasing, convex and concave. Since $x \mapsto \max_{1 \le j \le d} x_j$ is non decreasing and convex and $x \mapsto \min_{1 \le j \le d} x_j$ is

non decreasing and concave, it follows that for the call-on-max, U_k and V_k are convex and non decreasing, while for the put-on-min, U_k and V_k are convex and non increasing.

4.1.1. *Numerical illustration for the American call option.* For this example, S_t satisfies (4.1) with the same parameters as in Andersen and Broadie (2004), i.e., $r = 0.05$, $\delta = 0.1$, $\sigma = 0.2$, $T = 3$ and $K = 100$. They did their calculations assuming that $S_0 = 100$, that is the option is at-the-money. For this reason, one could take $R_0 = \cdots = R_n = R = [23, 230]$, so that $P(S_T \in R | S_0 = 100) \geq 0.997$. The Monte Carlo algorithm described in Section 3.2 was implemented using $M = 1000$ equally spaced vertices in R. Also, one used $N_{10} = \cdots = N_{n0} = N = 10000$ new trajectories each time.

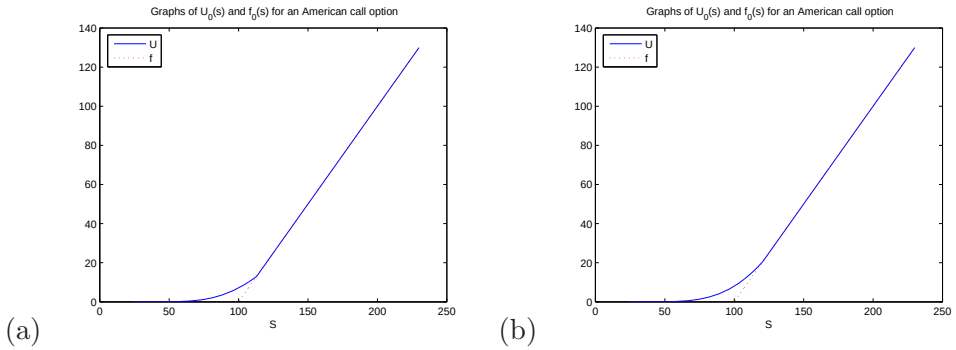
As argued previously, for each $k = 0, \ldots, n$, U_k and V_k are convex and non-decreasing. Since $d = 1$, one can take take the convex interpolation described in Example 1 of Appendix C. It follows from Corollary 3.6 that $\tilde{U}_k$ and $\tilde{V}_k$ are also convex and non decreasing.

Table 1 reports the results of the simulations for $n = 2$ and $n = 10$ exercise periods. In each case, the primal-dual algorithm was implemented with $N_0 = 2000000$, $N_1 = 1500$, $N_2 = N_3 = 10000$. The corresponding estimated values of $\tilde{U}_0$ are displayed in Figure 2.

TABLE 1. Estimated values of the call option with initial value $S_0 = 100$ and strike price $K = 100$.

Method	$n = 2$	$n = 10$
Monte Carlo	7.027575	7.950981
Lower bound	7.179106 (0.009548)	7.980827 (0.008106)
Upper bound	7.173063 (0.004365)	7.973237 (0.007579)
95% Confidence interval	[7.1604, 7.1816]	[7.9650, 7.9881]
Binomial tree	7.18	7.98

4.1.2. *Numerical illustration for the American call-on-max option on two assets.* For this example, $S_t = \left(S_t^{(1)}, S_t^{(0)}\right)$ satisfies (4.1) with the same parameters as in Andersen and Broadie (2004), i.e., $r = 0.05$, $\delta = 0.1$, $\sigma = 0.2$, $\rho = I_2$, $T = 3$ and $K = 100$. They did their calculations assuming that $S_0^{(1)} = S_0^{(0)} = 100$, that is the option is at-the-money.



(a) (b)

FIGURE 2. Estimated value of $U_0(s)$ for an American call option for $s \in [25, 230]$, with $n = 2$ exercise periods (panel a) and $n = 10$ exercise periods (panel b).

For this reason, one could take $R_0 = \cdots = R_n = R = [23, 230]^2$, so that $P(S_T \in R | S_0 = 100) \geq 0.995$. The Monte Carlo algorithm described in Section 3.2 was implemented using $M = 200^2$ vertices corresponding to equally spaced points in each coordinate. Also, one used $N_{10} = \cdots = N_{n0} = N = 10000$ new trajectories each time.

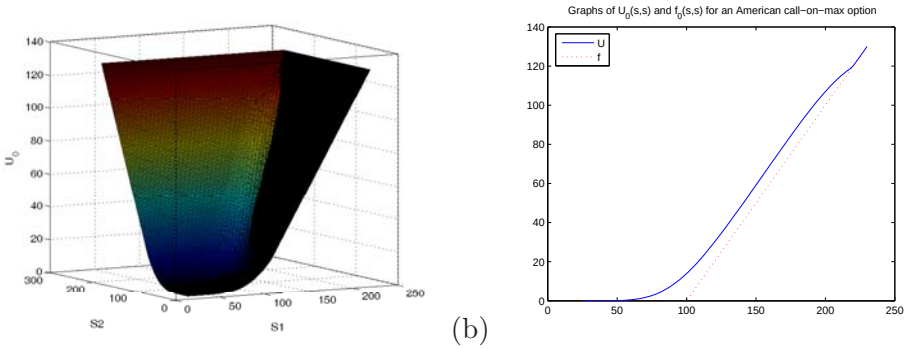
As argued previously, for each $k = 0, \ldots, n$, U_k and V_k are convex and non-decreasing. However, due to computation time constraints, the interpolation method used was only (locally) convex in each of the 199^2 sub-rectangles, using the the interpolation described in Example 2 of Appendix C.1. Therefore the estimated functions $\tilde{U}_k$ and $\tilde{V}_k$ are not necessarily convex over R. However, they are non decreasing, using Corollary 3.6.

Table 2 reports the results of the simulations for $n = 9$ exercise periods, as in Andersen and Broadie (2004), except that they only consider initial values $S_0 = (90, 90)$, $S_0 = (100, 100)$ and $S_0 = (110, 110)$. In each case, the primal-dual algorithm was implemented with the same number of iterations, i.e., $N_0 = 2000000$, $N_1 = 1500$, $N_2 = N_3 = 10000$. The results are quite similar. However, one obtains as a bonus the estimation of U_k over $R = [25, 230]^2$. This is illustrated in Figure 3, where the values along the diagonal are also displayed. One can remark that around the diagonal, the $\hat{U}_0$ is not convex, due to the fact that the convex interpolation was not used.

It is quite interesting to plot the estimated exercise region at time 0 given by the set of all points $s = (s_1, s_2) \in [25, 230]^2$ such that $\tilde{V}_0(s) \leq f_0(s) = \max\{0, \max(s_1, s_2) - 100)\}$. This is done in Figure 4. It is quite surprising to see that along the diagonal (s, s), the option is not exercised unless the $s > 220$. The values $S_0 = (160, 160)$ and $S_0 = (200, 200)$ were included in Table 2 to show that the option is not exercised at time 0 for these values, even if they are quite large compared to the strike price $K = 100$. Looking at the confidence intervals, one sees that $60 < 69.4746$ and $100 < 107.3452$.

TABLE 2. Estimated values of the call-on-max option with several initial values S_0 and strike price $K = 100$.

S_0	Method	$\tilde{U}_0(S_0, S_0)$
90	Monte Carlo	8.095579
	Lower bound	8.068165 (0.008518)
	Upper bound	8.070320 (0.008704)
	95% Confidence Interval	[8.0515, 8.0873]
	Binomial tree	8.075
100	Monte Carlo	13.900166
	Lower bound	13.912179 (0.010687)
	Upper bound	13.904105 (0.009686)
	95% Confidence Interval	[13.8913, 13.9230]
	Binomial tree	13.902
110	Monte Carlo	21.300620
	Lower bound	21.334454 (0.012411)
	Upper bound	21.338053 (0.011061)
	95% Confidence Interval	[21.3102, 21.3597]
	Binomial tree	21.345
160	Monte Carlo	69.244785
	Lower bound	69.507220 (0.016665)
	Upper bound	69.519498 (0.011474)
	95% Confidence Interval	[69.4746, 69.5419]
	Binomial tree	
200	Monte Carlo	106.990926
	Lower bound	107.372955 (0.014175)
	Upper bound	107.370157 (0.008543)
	95% Confidence Interval	[107.3452, 107.3869]
	Binomial tree	



(a) (b)

FIGURE 3. Estimated value of a call-on-max American option at period 0 for $s \in [25, 230]^2$ (panel a) and along its diagonal (panel b), with $n = 9$ exercise periods.

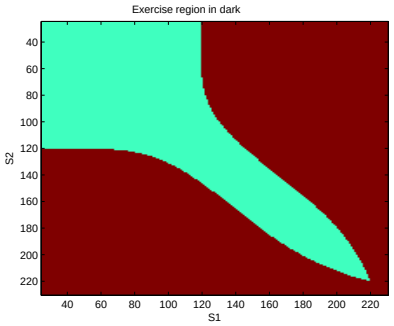


FIGURE 4. Estimated exercice region at period 0 for a call-on-max American option with 9 exercise periods.

4.2. **N-GARCH models.** As a last example of application, consider the following N-GARCH model under the martingale measure Q:

$$\begin{aligned} S_k &= S_{k-1} e^{r-\delta-h_k/2+h_k^{1/2} Y_k} \\ h_k &= \beta_0 + \beta_1 h_{k-1} + \beta_2 h_{k-1}(Y_{k-1} - \theta)^2, \end{aligned}$$

where $Y_k \sim N(0,1)$. Then $\{S_k\}_{k\geq 0}$ or $\{(S_k, h_k)\}_{k\geq 0}$ are not Markov processes. However, it is easy to check that $\{X_k = (S_k, h_{k+1})\}_{k\geq 0}$ is a Markov chain. In fact, Assumption 1 holds with $\pi_k(s,h,y) = \left(se^{r-\delta-h/2+h^{1/2}y}, \beta_0 + \beta_1 h + \beta_2 h(y-\theta)^2\right)$. It follows that for any given $y \in \mathbb{R}$, $\pi_k(s,h,y)$ is monotone in s but it is not monotone in h, nor it is convex or concave. It

is well-known, e.g., Duan and Simonato (2001), that under the condition $\beta_1+\beta_2(1+\theta^2) < 1$, X_k has a stationary distribution and $\lim_{k\to\infty} E(H_k) \to h^\star = \beta_0\{1-\beta_1-\beta_2(1+\theta^2)\}^{-1}$.

In Duan and Simonato (2001) and Duan et al. (2003), the authors proposed two kinds of approximations for the American put option on S_k. They calculated the option prices for $S_0 = 50$ for three strike prices $K \in \{45, 50, 55\}$, three volatility levels $h_1 \in \{0.8h^\star, h^\star, 1.2h^\star\}$ and four exercise periods $n \in \{10, 30, 90, 270\}$. Here, under similar setting, option prices for $n = 10$ and $n = 30$ exercise periods are estimated using the Monte Carlo methodology and the primal-dual algorithm. However, since there there was almost no difference in option prices when $K = 45$ or $K = 55$, only the case $K = 50$ was considered. The Monte Carlo estimation was based on $N = 10000$ for each time steps. In the case $n = 10$, the rectangle $[45, 55] \times [5\beta_0, 45\beta_0]$ was partitioned into 300×100 points and the same number of points was used to partition $[40, 65] \times [5\beta_0, 50\beta_0]$ for the case $n = 30$. These rectangles, obtained through Monte Carlo simulations, correspond to approximate 95% confidence regions for X_{10} and X_{30} respectively. As before, the number of iterations used for the primal-dual algorithm were $N_0 = 2000000$, $N_1 = 1500$, $N_2 = N_3 = 10000$. Note that by Corollary 3.6, $\tilde{U}_k$ is non-increasing in s. This is reflected in Figure 5, where the values of $\tilde{U}_0(s, h)$ are displayed. The corresponding exercise regions are shown in Figure 6.

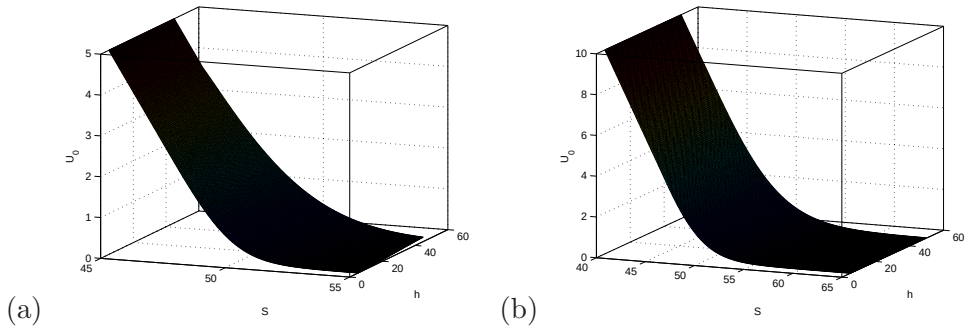
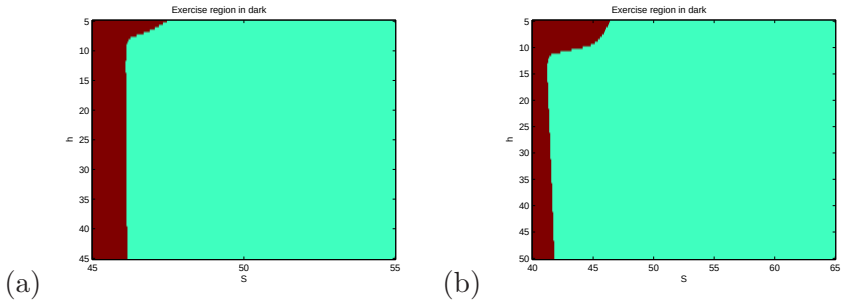


FIGURE 5. Estimated value of U_0 for an American put option based on a N-GARCH model, with 10 exercise periods (panel a) and 30 exercise periods (panel b).

TABLE 3. Estimated values of the American put options at time 0 for $n = 10$ and $n = 30$ exercise periods, with initial asset value $S_0 = K = 50$, three initial volatilities h_1 and strike price $K = 50$.

$h_1/h^\star$	Method	$n = 10$	$n = 30$
0.8	Monte Carlo	0.642316	1.070787
	Lower bound	0.628298 (0.000642)	1.092107 (0.001067)
	Upper bound	0.628709 (0.000728)	1.095878 (0.002013)
	95% C.I.	[0.6270, 0.6301]	[1.0900, 1.0998]
	Markov Chain	0.63	1.09
1.0	Monte Carlo	0.696359	1.119633
	Lower bound	0.680187 (0.000695)	1.142400 (0.001112)
	Upper bound	0.679323 (0.000770)	1.147375 (0.002043)
	95% C.I.	[0.6788, 0.6808]	[1.1402, 1.1514]
	Markov Chain	0.68	1.14
1.2	Monte Carlo	0.746471	1.166276
	Lower bound	0.729034 (0.000743)	1.189325 (0.001156)
	Upper bound	0.729605 (0.000816)	1.192303 (0.002188)
	95% C.I.	[0.7276, 0.7312]	[1.1871, 1.1966]
	Markov Chain	0.73	1.19



(a) (b)

FIGURE 6. Estimated exercice region $\widehat{\mathcal{E}}$ at period 0 for the American put option based on a N-GARCH model, for 10 exercise periods (panel a) and for 30 exercise periods (panel b).

5. CONCLUSION

Monotonicity and convexity are important properties shared by many American option prices, especially when it comes to define the exercise region. In this paper, it was shown that these properties can be preserved by a Monte Carlo algorithm which is easy to implement and yields the whole Snell envelope. Also a modified approach inspired by the

work of Andersen and Broadie (2004) can be used to improve the precision of the estimation. Numerical results showed that the proposed methodology provides accurate results for American option prices when assets follows a Markov chain with continuous state space. In addition, the Monte Carlo algorithm can also be applied for finding optimal hedging strategies, as in Hocquard et al. (2007) and Papageorgiou et al. (2008).

References

Andersen, L. and Broadie, M. (2004). Primal-dual simulation algorithm for pricing multidimensional American options. *Management Science*, 50:1222–1234.

Bally, V., Pagès, G., and Printems, J. (2005). A quantization tree method for pricing and hedging multidimensional American options. *Math. Finance*, 15(1):119–168.

Boyle, P. (1977). Options: A Monte Carlo approach. *Journal of Financial Economics*, 4:323–338.

Brennan, M. and Schwartz, E. (1977). The valuation of American put options. *Journal of Finance*, XXXII:449–462.

Broadie, M. and Detemple, J. (1997). The valuation of American options on multiple assets. *Mathematical Finance*, 7:241–286.

Broadie, M., Detemple, J., Ghysels, E., and Torrès, O. (2000). Nonparametric estimation of American options' exercise boundaries and call prices. *Journal of Economic Dynamics and Control*, 24:1829–1857.

Broadie, M. and Glasserman, P. (1997). Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control*, 21:1323–1352.

Broadie, M. and Glasserman, P. (2004). A stochastic mesh method for pricing high-dimensional American options. *Journal of Computational Finance*, 7:35–72.

Carriere, J. (1996). Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, 19:19–30.

Clément, E., Lamberton, D., and Protter, P. (2002). An analysis of a least squares regression method for American option pricing. *Finance Stoch.*, 6:449 – 471.

Cox, J., Ross, S., and Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of Financial Economics*, 7:229–263.

Duan, J.-C., Gauthier, G., Sasseville, C., and Simonato, J.-G. (2003). Approximating American option prices in the GARCH framework. *The Journal of Futures Markets*, 23:915–929.

Duan, J.-C., Gauthier, G., and Simonato, J.-G. (2004). Numerical pricing of contingent claims on multiple assets and/or factor - A low-discrepancy Markov chain approach derivatives in the GARCH model. Technical Report 80, GERAD.

Duan, J.-C. and Simonato, J.-G. (2001). American option pricing under GARCH by a Markov chain approximation. *J. Econom. Dynam. Control*, 25(11):1689–1718.

Fu, M., Laprise, S., Madan, D., Su, Y., and Wu, R. (2001). Pricing American options: A comparison of Monte Carlo simulation approaches. *Journal of Computational Finance*, 4(3):39–88.

Haugh, M. B. and Kogan, L. (2004). Pricing american options: A duality approach. *Operations Research*, 52(2):pp. 258–270.

Hocquard, A., Papageorgiou, N., and Rémillard, B. (2007). Optimal hedging strategies with an application to hedge fund replication. *Wilmott Magazine*, (Jan-Feb):62–66.

Kargin, V. (2005). Lattice option pricing by multidimensional interpolation. *Math. Finance*, 15(4):635–647.

Laprise, S., Fu, M., Marcus, S., Lim, A., and Zhang, H. (2006). Pricing American-style derivatives with European call options. *Management Science*, (1):95–110.

Longstaff, F. and Schwartz, E. (2001). Valuing American options by simulation: A simple least-square approach. *The Review of Financial Studies*, 14:113–147.

Neveu, J. (1975). *Discrete-parameter Martingales*. North Holland, Amsterdam.

Papageorgiou, N., Rémillard, B., and Hocquard, A. (2008). Replicating the properties of hedge fund returns. *Journal of Alternative Invesments*, 11:8–38.

Rogers, C. (2002). Monte Carlo valuation of American options. *Mathematical Finance*, 12:271–286.

Tilley, J. (1993). Valuing American options in a path simulation model. *Transactions of the Society of Actuaries*, 45:83–104.

Appendix A. Auxiliary results

Suppose that $g \in \mathcal{B}_+(\mathfrak{X})$, that is $g : \mathfrak{X} \mapsto [0, \infty)$ is measurable. Suppose also that $\phi : \mathfrak{X} \times \mathcal{Y} \mapsto \mathfrak{X}$ is measurable and that for any $y \in Y$, $\phi(\cdot, y)$ is continuous. Moreover, assume that for any $x \in \mathfrak{X}$, $g(\phi(x, \cdot))$ is integrable with respect to a measure ν on $\mathcal{Y}$. Then the function Mg, defined on $\mathfrak{X}$ by $x \mapsto Mg(x) = \int_{\mathcal{Y}} g(\phi(x, y))\nu(dy)$, belongs to $\mathcal{B}_+(\mathfrak{X})$.

Proposition A.1. *Assume $g \in \mathcal{B}_+(\mathfrak{X})$ and suppose that for any $y \in \mathcal{Y}$, $\phi(\cdot, y)$ is non decreasing.*

(i) If g is monotone then Mg has the same monotonicity.

(ii) If g is monotone and continuous, then Mg is monotone continuous.

(iii) If g is convex and non decreasing, and if for any $y \in \mathcal{Y}$, $\phi(\cdot, y)$ is convex over $\mathfrak{X}$, then Mg is also convex and non decreasing.

(iv) If g is convex and non increasing, and if, for any fixed $y \in \mathcal{Y}$, $\phi(\cdot, y)$, is concave, then Mg is also convex and non increasing.

Proof. Suppose g is non decreasing for any fixed y. Therefore, for any fixed y, $x \mapsto g(\phi(x, y))$ is non decreasing. If g is non increasing, then for any fixed y, $x \mapsto g(\phi(x, y))$ is non increasing. Integrating with respect to ν yields (i). If, for any fixed y, g is non decreasing and continuous, then $x \mapsto g(\phi(x, y))$ is continuous and non decreasing. Hence (ii) follows from the Dominated Convergence Theorem. If g is non increasing and continuous then $x \mapsto$

$g(\phi(x,y))$ is continuous and non decreasing, and (ii) follows from the Bounded Convergence Theorem.

To prove (iii), assume that g is convex and non decreasing, and that $\phi(\cdot, y)$ is convex, for any fixed $y \in E$. Then, for any $\lambda \in [0,1]$, for any $y \in E$, and for any $x, x' \in \mathfrak{X}$,

$$\phi(\lambda x + (1-\lambda)x', y) \leq \lambda\phi(x,y) + (1-\lambda)\phi(x',y),$$

so

$$\begin{aligned} g(\phi(\lambda x + (1-\lambda)x', y)) &\leq g(\lambda\phi(x,y) + (1-\lambda)\phi(x',y)) \\ &\leq \lambda g(\phi(x,y)) + (1-\lambda)g(\phi(x',y)). \end{aligned}$$

Integrating with respect to y, one obtains $Mg(\lambda x + (1-\lambda)x') \leq \lambda Mg(x) + (1-\lambda)Mg(x')$, proving that Mg is convex. Finally, if g is convex and non increasing, and $\phi(\cdot, y)$ is concave for any fixed y, then, for any $\lambda \in [0,1]$, and for any y, $\phi(\lambda x + (1-\lambda)x', y) \geq \lambda\phi(x,y) + (1-\lambda)\phi(x',y)$, so

$$\begin{aligned} g(\phi(\lambda x + (1-\lambda)x', y)) &\leq g(\lambda\phi(x,y) + (1-\lambda)\phi(x',y)) \\ &\leq \lambda g(\phi(x,y)) + (1-\lambda)g(\phi(x',y)). \end{aligned}$$

Integrating with respect to y, one obtains $Mg(\lambda x + (1-\lambda)x') \leq \lambda Mg(x) + (1-\lambda)Mg(x')$. This completes the proof. □

The next result is easy to prove.

Proposition A.2. *For any real numbers x, y, z, w,*

$$|x \vee z - y \vee w| \leq |x-y| \vee |z-w|.$$

In particular, $|x \vee z - x \vee w| \leq |z-w|$.

Appendix B. Proofs of the main results

B.1. **Proof of Proposition 3.2.** First, for any $k \in \{0, \ldots, n\}$,

$$U_k(x) = \sup_{\tau \in \mathcal{T}_{k,n}} E\left\{f_\tau(X_\tau) | X_k = x\right\},$$

so it follows that $V_k(x) \leq U_k(x) \leq H_k(x) = \sum_{j=k}^n E\left\{f_j(X_j)|X_k = x\right\}$. Next, by hypothesis, V_n and $H_k(\pi_k(\cdot, y))$ are continuous for any $y \in \mathcal{Y}$ and any $k \in \{0, \ldots, n\}$. To show that V_k is continuous for any $k \in \{0, \ldots, n\}$, it suffices to show that if V_k is continuous, then V_{k-1} is also continuous.

So suppose that V_k is continuous. Then so is U_k. Suppose that $(x_j)_{j\geq 1}$ is a sequence converging to $x \in \mathcal{X}$. Then, $U_k(\pi_k(x_j, y)) \leq H_k(\pi_k(x_j, y))$, and for all $y \in \mathcal{Y}$,

$$U_k(\pi_k(x_j, y)) \to U_k(\pi_k(x, y)), \quad H_k(\pi_k(x_j, y)) \to H_k(\pi_k(x, y))$$

as $j \to \infty$. Moreover, by hypothesis,

$$\begin{aligned} \int H_k(\pi_k(x_j, y))\mu_k(dy) &= H_{k-1}(x_j) - f_{k-1}(x_j) \\ &\overset{j\to\infty}{\longrightarrow} H_{k-1}(x) - f_{k-1}(x) \\ &= \int H_k(\pi_k(x, y))\mu_k(dy). \end{aligned}$$

Thus, by Fatou's Lemma, $V_{k-1}(x_j) \to V_{k-1}(x)$ as $j \to \infty$. Therefore V_{k-1} and U_{k-1} are continuous. Hence the result. □

B.2. **Proof of Theorem 3.5.**

Proof. Recall that for all $k \in \{0, \ldots, n\}$, the f_k, V_k and U_k are continuous. Set

$$M_k g(x) = \int_{\mathcal{Y}} g(\pi_k(x, y))\mu_k(dy)$$

and

$$\hat{V}_{k-1}(x) = \hat{M}_k \tilde{U}_k(x) = \frac{1}{N_k} \sum_{i=1}^{N_k} \tilde{U}_k\left\{\pi_k(x, \xi_{i,k})\right\}.$$

The approximation algorithm can be summarized as follows:

$$\tilde{U}_n = f_n \text{ and } \left\{ \begin{array}{lcll} \hat{V}_{k-1} & = & \hat{M}_k \tilde{U}_k \text{ on } \mathcal{V}_{k-1} = \mathcal{V}(\mathcal{P}_{k-1}), & \\ \tilde{V}_{k-1} & = & \mathcal{I}_{\mathcal{P}_{k-1}} \hat{V}_{k-1} & k = n, \ldots, 1, \\ \tilde{U}_{k-1} & = & f_{k-1} \vee \tilde{V}_{k-1}. & \end{array} \right.$$

where $\mathcal{I}_{\mathcal{P}}g$ denotes the linear interpolation of g over the compact set R with associated partition $\mathcal{P}$ of R as in Definition C.1. For simplicity set $\mathcal{I}_{k-1} = \mathcal{I}_{\mathcal{P}_{k-1}}$ and $\delta_{k-1} = \text{mesh}(\mathcal{P}_{k-1})$. First, for any compact subset R of $\mathfrak{X}$, $\|\tilde{U}_k - U_k\|_R \leq \|\tilde{V}_k - V_k\|_R$, $0 \leq k \leq n$, by Proposition A.2. Therefore, to prove the result, it suffices to show that for any $k = 1, \ldots, n$, given a compact set R_{k-1} and $\epsilon > 0$, one can find $\delta_{k-1} > 0$, a partition $\mathcal{P}_{k-1}$ with $\text{mesh}(\mathcal{P}_{k-1}) < \delta_{k-1}$, another compact set R_k and an integer N_{k0} so that $\|\tilde{V}_{k-1} - V_{k-1}\|_{R_{k-1}} \leq \|\tilde{V}_k - V_k\|_{R_k} + \epsilon$.

To begin, one has, for any $k = 1, \ldots, n$,

$$V_{k-1} - \tilde{V}_{k-1} = (V_{k-1} - \mathcal{I}_{k-1}V_{k-1}) + \left(\mathcal{I}_{k-1}V_{k-1} - \mathcal{I}_{k-1}\hat{V}_{k-1}\right).$$

Since

$$\begin{aligned}
\left\|\mathcal{I}_{k-1}V_{k-1} - \mathcal{I}_{k-1}\hat{V}_{k-1}\right\|_{R_{k-1}} &= \sup_{x \in \mathcal{V}_{k-1}} \left|V_{k-1}(x) - \hat{V}_{k-1}(x)\right| \\
&= \left\|V_{k-1} - \hat{V}_{k-1}\right\|_{\mathcal{V}_{k-1}} \\
&= \left\|M_k U_k - \hat{M}_k \tilde{U}_k\right\|_{\mathcal{V}_{k-1}}
\end{aligned}$$

and

$$\begin{aligned}
\left\|M_k U_k - \hat{M}_k \tilde{U}_k\right\|_{\mathcal{V}_{k-1}} &\leq \left\|M_k U_k - \hat{M}_k U_k\right\|_{\mathcal{V}_{k-1}} + \left\|\hat{M}_k U_k - \hat{M}_k \tilde{U}_k\right\|_{\mathcal{V}_{k-1}} \\
&\leq \left\|M_k U_k - \hat{M}_k U_k\right\|_{\mathcal{V}_{k-1}} + \left\|V_k - \tilde{V}_k\right\|_{R_k} \\
&\quad + \left\|\hat{M}_k (U_k 1_{R_k^c})\right\|_{\mathcal{V}_{k-1}} \\
&\quad + \left\|\tilde{U}_k\right\|_{R_k} \left\|\hat{M}_k 1_{R_k^c}\right\|_{\mathcal{V}_{k-1}},
\end{aligned}$$

it follows that

$$\begin{aligned}
\left\|\tilde{V}_{k-1} - V_{k-1}\right\|_{R_{k-1}} &\leq \omega(V_{k-1}, R_{k-1}, \delta_{k-1}) + \left\|M_k U_k - \hat{M}_k U_k\right\|_{\mathcal{V}_{k-1}} \\
&\quad + \left\|V_k - \tilde{V}_k\right\|_{R_k} + \left\|\hat{M}_k (U_k 1_{R_k^c})\right\|_{\mathcal{V}_{k-1}} \\
&\quad + \left\|\tilde{U}_k\right\|_{R_k} \left\|\hat{M}_k 1_{R_k^c}\right\|_{\mathcal{V}_{k-1}},
\end{aligned}$$

using the above inequalities together with Remark C.1.

The aim is to show that if $N_1, \dots, N_n$ are large enough, and if $R_1, \dots, R_n$ are large enough, and if δ_0, δ_{n-1} are small enough, then the terms $\omega(V_{k-1}, R_{k-1}, \delta_{k-1})$, $\left\| M_k U_k - \hat{M}_k U_k \right\|_{\mathcal{V}_{k-1}}$, $\left\| \hat{M}_k (U_k 1_{R_k^c}) \right\|_{\mathcal{V}_{k-1}}$ and $\left\| \tilde{U}_k \right\|_{R_k} \left\| \hat{M}_k 1_{R_k^c} \right\|_{\mathcal{V}_{k-1}}$ can be arbitrarily small, for all $1 \leq k \leq n$.

The first term is easy to handle since V_{k-1} is continuous and R_{k-1} is compact. To handle the second term, just use the strong law of large numbers, since $\mathcal{V}_{k-1}$ is finite. The last two terms can be made arbitrarily small since, by the strong law of large numbers, $\left\| \hat{M}_k (U_k 1_{R_k^c}) \right\|_{\mathcal{V}_{k-1}}$ and $\left\| \hat{M}_k 1_{R_k^c} \right\|_{\mathcal{V}_{k-1}}$ converge almost surely to $\left\| M_k (U_k 1_{R_k^c}) \right\|_{\mathcal{V}_{k-1}}$ and to $\left\| M_k 1_{R_k^c} \right\|_{\mathcal{V}_{k-1}}$, which both can be as small as one wants, choosing R_k large enough. □

B.3. **Proof of Corollary 3.6.** It follows from that $\hat{V}_{k-1} = \hat{M}_k \tilde{U}_k$ is continuous, monotone and convex, if $\tilde{U}_k$ is, by Proposition A.1. As shown in Proposition C.4 and Lemma C.3, the quick interpolation and the convex interpolation on a grid both preserve all monotonicity. In addition, the convex interpolation preserves convexity, by Lemma C.3. Finally, Theorem 3.5 yields the almost sure uniform convergence. □

Appendix C. Linear interpolations

Definition C.1. Given a function g and a partition $\mathcal{P}$ of R, the linear interpolation of g over $\mathcal{P}$ is the (unique) function $\tilde{g}$ defined in the following way:

- If $x \in S \subset R$, where $S \in \mathcal{P}$ is a simplex with vertices $x_1, \dots, x_{d+1}$,
$$\tilde{g}(x) = \sum_{i=1}^{d+1} \lambda_i g(x_i),$$
where the barycenters $\{\lambda_1, \dots, \lambda_{d+1}\}$ are the unique solution of
$$x = \sum_{i=1}^{d+1} \lambda_i x_i, \quad \sum_{i=1}^{d+1} \lambda_i = 1, \quad \lambda_i \in [0,1], i = 1, \dots d+1.$$
- If $x \notin R$, let x_R be the (unique) closest point to x that belongs to R, and set $\tilde{g}(x) = \tilde{g}(x_R)$.

Uniqueness follows from the convexity of R and the strict convexity of the Euclidean norm.

Remark C.1. Note that since each x_i is extreme in S, the unique solution of

$$x_i = \sum_{j=1}^{d+1} \lambda_j x_j, \quad \sum_{j=1}^{d+1} \lambda_j = 1, \quad \lambda_j \in [0,1], j = 1, \dots d+1,$$

is $\lambda_i = 1$ and $\lambda_j = 0$ for all $j \neq i$, yielding $\tilde{g}(x_i) = g(x_i)$ for all $1 \leq i \leq m$. Moreover, $\tilde{g}$ is affine on each simplex, justifying the term "linear interpolation".

Finally, $\tilde{g}$ is continuous and bounded on $\mathfrak{X}$ and

$$\sup_{x \in R} |g(x) - \tilde{g}(x)| \leq \omega(g, R, \mathrm{mesh}(\mathcal{P})),$$

where $\mathrm{mesh}(\mathcal{P}) = \max_{\mathrm{S} \in \mathcal{P}} \sup_{\mathrm{x,z} \in \mathrm{S}} \|\mathrm{x} - \mathrm{z}\|$ and $\omega(g, R, \delta)$ is the modulus of continuity of g over R, i.e., $\omega(g, R, \delta) = \sup_{x,z \in R,\, \|x-z\| \leq \delta} |g(x) - g(z)|$.

Example 1. *Suppose $d = 1$. Then the linear interpolation $\tilde{g}$ of a monotone (respectively convex) function g on $R = [a, b]$ is monotone (respectively convex). To see that, let $\mathcal{P}$ be the partition given by $\mathcal{P} = \{a = x_0 < x_1 < \cdots < x_m = b\}$, i.e the simplexes are $[x_{i-1}, x_i]$, $i = 1, \dots, m$. Set $\Delta_i = \frac{g(x_i) - g(x_{i-1})}{x_i - x_{i-1}}$, $1 \leq i \leq m$. Then the linear interpolation of g over R is given by*

$$\tilde{g}(x) = \begin{cases} g(a), & x \leq a, \\ g(x_i) + (x - x_i)\Delta_{i+1}, & x \in [x_i, x_{i+1}],\ i = 0, \dots, m-1, \\ g(b) & x \geq b. \end{cases}$$

If g is monotone, the slopes Δ_i all have the same sign, so $\tilde{g}$ has the same monotonicity. If g is convex, the slopes Δ_i are non decreasing, so $\tilde{g}$ is also convex.

Definition C.2. Given a convex function g and a partition $\mathcal{P}$ of R, a convex linear interpolation of g over $\mathcal{P}$ is a function $\tilde{g}$ defined on $\mathfrak{X}$ such that

(CLI1) $\tilde{g}(x)$ is convex on R;

(CLI2) $g(x) = \tilde{g}(x)$, for all $x \in \mathcal{V}(\mathcal{P})$;

(CLI3) For any $x \in R$, there exists non negative numbers λ_z, $z \in \mathcal{V}(\mathcal{P})$, such that

(a) $\displaystyle\sum_{z \in \mathcal{V}(\mathcal{P})} \lambda_z = 1,$

(b) $\displaystyle x = \sum_{z \in \mathcal{V}(\mathcal{P})} z\lambda_z,$

(c) $\tilde{g}(x) = \sum_{z \in \mathcal{V}(\mathcal{P})} \lambda_z g(z)$;

(CLI4) For any $x \notin R$, $\tilde{g}(x) = \tilde{g}(y)$, where y is the unique point satisfying $d(x, R) = d(x, y)$.

The following result proves that convex linear interpolation exists and it can be obtained via the simplex algorithm.

Lemma C.3. *Suppose C is the convex hull of $\{x_1, \ldots, x_n\}$. Suppose also that $g_1, \ldots, g_n$ are real numbers. For any $x \in C$, set*

$$\Lambda_x = \left\{ \lambda \in [0,1]^n; \sum_{i=1}^n \lambda_i x_i = x, \quad \sum_{i=1}^n \lambda_i = 1 \right\}.$$

Define

$$\tilde{g}(x) = \inf_{\lambda \in \Lambda_x} \sum_{i=1}^n \lambda_i g_i.$$

Then $\tilde{g}$ is convex over C and $\tilde{g}(x_i) \leq g_i$. If in addition, g is a convex function over C such that $g(x_i) = g_i$, then $\tilde{g}$ is a linear convex interpolation of g. In particular, $g(x) \leq \tilde{g}(x)$ for all $x \in C$, and $\tilde{g}(x_i) = g_i$, for all $1 \leq i \leq n$.

Suppose that the points $x_1, \ldots, x_n$ form a grid. Then if g is monotone, then $\tilde{g}$ has the same monotonicity, i.e., if g is non-decreasing in x_j the so is $\tilde{g}$, for any $1 \leq j \leq d$.

Proof. First, it is obvious that $\tilde{f}(x_i) \leq g_i$ for any $1 \leq i \leq n$. Next, suppose that $a, b \in C$. Take $\lambda \in \Lambda_a$ and $\lambda' \in \Lambda_b$. Let $t \in [0,1]$ be given and set $x = ta + (1-t)b$. Then $t\lambda + (1-t)\lambda' \in \Lambda_x$, so

$$\tilde{g}(x) \leq \sum_{i=1}^n \left(t\lambda_i + (1-t)\lambda_i'\right) g_i = t\left(\sum_{i=1}^n \lambda_i g_i\right) + (1-t)\left(\sum_{i=1}^n \lambda_i' g_i\right).$$

Taking the infimum over all $\lambda \in \Lambda_a$ and all $\lambda' \in \Lambda_b$ in the last inequality yields $\tilde{g}(x) \leq t\tilde{g}(a) + (1-t)\tilde{g}(b)$. Hence $\tilde{g}$ is convex over C.

Next, suppose that g is convex over C, and $g(x_i) = g_i$ for all $i = 1, \ldots, n$. Let $x \in C$ be given. Since $\lambda : \Lambda_x \mapsto \sum_{i=1}^n \lambda_i g_i$ is continuous and Λ_x is compact, there exists $\lambda \in \Lambda_x$ so that $\tilde{g}(x) = \sum_{i=1}^n \lambda_i g_i$. Therefore $g(x) \leq \sum_{i=1}^n \lambda_i g(x_i) = \sum_{i=1}^n \lambda_i g_i = \tilde{g}(x)$. Hence $g(x) \leq \tilde{g}(x)$. In particular, $g_i = g(x_i) \leq \tilde{g}(x_i) \leq g_i$, showing that $\tilde{g}(x_i) = g_i$, for all $1 \leq i \leq n$.

Finally, to prove the monotonicity property of $\tilde{g}$, note that it suffices to prove it for the interior points of C. For simplicity, the proof is given for a two-dimensional grid but it holds true for any dimension. Let the grid be given by $G = (y_i, z_j); 1 \leq i \leq M_1, 1 \leq j \leq M_2)$ and set $g_{ij} = g(y_j, z_j)$. From linear programming, it is known that for any $x = (y, z)$, then

$$\tilde{g}(y, z) = \sup_{a,b,c \in A} a + by + cz,$$

where $A = \{a, b, c \in \mathbb{R}; a + by_i + cz_j \leq g_{ij}, \text{ for all } (y_i, z_j) \in G\}$.

It is known that there are 3 points on the grid where the constraints are in fact an equality. One of these points, say (y_i, z_j) must an interior point of the grid, for otherwise (y, z) could not be an interior point of C. Therefore $a + by_i + cz_j = g_{ij}$. One also has $g_{ij} + b(y_{i+1} - y_i) = a + by_{i+1} + cz_j \leq g_{i+1,j}$, so $b \leq \frac{g_{i+1,j} - g_{ij}}{y_{i+1} - y_i}$. Similarly, $g_{ij} - b(y_i - y_{i-1}) = a + by_{i-1} + cz_j \leq g_{i-1,j}$, so $b \geq \frac{g_{ij} - g_{i-1,j}}{y_i - y_{i-1}}$. It follows that if g is non-decreasing in y, $b \geq 0$. Hence, for $\epsilon > 0$ small enough, $\tilde{g}(y + \epsilon, z) - g(y, z) = b\epsilon \geq 0$. The same argument applies in g is non-increasing in y. Hence the result. □

C.1. **Quick linear interpolation on rectangles.** Suppose that g is defined on $R = [0, 1]^d$, and g is known at all extreme points of R. If one wants to linearly interpolate g, as in Definition C.1, a convenient choice for the partition $\mathcal{P}$ is the set of all $d!$ simplexes defined, for any permutation π of $\{1, \ldots, d\}$, by $S_\pi = \{y \in [0, 1]^d; y_{\pi_1} \leq \cdots, y_{\pi_d}\}$. Note that the vertices of $S = \{y \in [0, 1]^d; y_1 \leq \cdots, y_d\}$ are 0 and the vectors $u_1, \ldots, u_d$ are such that the $A = (u_1| \cdots |u_d)$ is defined by $A_{ij} = 1$ if $1 \leq j \leq i \leq d$, and $A_{ij} = 0$ otherwise. Any $y \in S$ can be uniquely written as $y = \sum_{j=1}^d \lambda_j u_j$, with $\lambda_1 = y_1$, $\lambda_k = y_k - y_{k-1}$, $2 \leq k \leq d$, and $\lambda_{d+1} = 1 - y_d$. To define $\tilde{g}$ at $x = (x_1, \ldots, x_d) \in R$, proceed in the following way:

- Let π be a permutation such that $x_{\pi_1} \leq x_{\pi_2} \leq \cdots \leq x_{\pi_d}$ and let P be the associated permutation matrix. Set $y = Px$;
- Set $\lambda_1 = y_1$, $\lambda_k = y_k - y_{k-1}$, $2 \leq k \leq d$, and $\lambda_{d+1} = 1 - y_d$;

- Set $\hat{g}(x) = \sum_{j=1}^{d} \lambda_j g\left(P^\top u_j\right) + \lambda_{d+1} g(0)$.

Remark C.2. If $R = \prod_{j=1}^{d}[a_j, b_j]$, then set $x'_j = \frac{x_j - a_j}{b_j - a_j}$, $1 \leq j \leq d$ and apply the last procedure. The extreme points necessary for the interpolation are then given by $a + DP^\top u_j$, $1 \leq j \leq d$ and a, where D is the diagonal matrix with diagonal $b - a$. The quick interpolation on a grid is just the repeated application of the quick interpolation on the partition of rectangles generating the grid.

Example 2. *Suppose that one wants to interpolate linearly a convex function f over $[0,1]^2$, knowing only the values at extremes, that is $f(0,0)$, $f(1,0)$, $f(0,1)$ and $f(1,1)$. One could take the partition $\mathcal{P} = \{C_1, C_2\}$, where $C_1 = \{(x,y) \in [0,1]^2; y \leq x\}$ and $C_2 = \{(x,y) \in [0,1]^2; x \leq y\}$. This would lead to the quick interpolation described above, namely $g(x,y) = \begin{cases} f_1(x,y), & (x,y) \in C_1 \\ f_2(x,y), & (x,y) \in C_2 \end{cases}$, where $f_1(x,y) = x\{f(1,0) - f(0,0)\} + y\{f(1,1) - f(1,0)\} + f(0,0)$ and $f_2(x,y) = x\{f(1,1) - f(0,1)\} + y\{f(0,1) - f(0,0)\} + f(0,0)$.*

On the other hand, one could also consider the partition $\mathcal{P}' = \{C_3, C_4\}$, where $C_3 = \{(x,y) \in [0,1]^2; x + y \leq 1\}$ and $C_4 = \{(x,y) \in [0,1]^2; x + y \geq 1\}$. That would lead to the following interpolation g' of f: $g'(x,y) = \begin{cases} f_3(x,y), & (x,y) \in C_3 \\ f_4(x,y), & (x,y) \in C_4 \end{cases}$, where $f_3(x,y) = x\{f(1,0) - f(0,0)\} + y\{f(0,1) - f(0,0)\} + f(0,0)$, $f_4(x,y) = x\{f(1,1) - f(0,1)\} + y\{f(1,1) - f(1,0)\} - \Delta + f(0,0)$, and $\Delta = f(1,1) - f(0,1) - f(1,0) + f(0,0)$.

So which partition should be chosen? If convexity is at stake, the answer relies only on the sign of Δ. In fact, it it easy to check that g is convex on $[0,1]^2$ if and only if $\Delta \leq 0$, while g' is convex on $[0,1]^2$ if and only if $\Delta \geq 0$. It is also easy to check that Lemma C.3 applied with the four extreme points $(0,0)$, $(1,0)$, $(0,1)$ and $(1,1)$ of $[0,1]^2$ yields g when $\Delta \leq 0$ (resp. g' when $\Delta \geq 0$).

Proposition C.4. *Suppose that g is monotone and $\tilde{g}$ represents the quick interpolation of g over a grid. Then $\tilde{g}$ has the same monotonicity as g, i.e., if g is non-decreasing in x_j the so is $\tilde{g}$, for any $1 \leq j \leq d$.*

Proof. It suffices to prove the result on any rectangle forming the partition of the grid. By translation and rescaling, one can consider that $R = [0,1]^d$. In that case, note that one can write $\tilde{g}(x) = \sum_{j=1}^{d} y_j \left\{ g(P^\top u_j) - g(P^\top u_{j+1}) \right\}$, where $u_{d+1} = 0$. Next, $P^\top u_j - P^\top u_{j+1} = P^\top e_j$, where e_j is the unit vector with $(e_j)_j = 1$ and $(e_j)_i = 0$, $i \neq j$. Hence $P^\top u_j \geq P^\top u_{j+1}$, so if g is non-decreasing in x_j, then $b_j = g(P^\top u_j) - g(P^\top u_{j+1}) \geq 0$. It follows that if x in an interior point of one of the simplexes, then $x + \epsilon e_j$ belongs to the same simplex if $\epsilon > 0$ is small enough. Suppose that the rank of x_j is k. Then, the rank of $x_j + \epsilon$ is also k and $\tilde{g}(x + \epsilon e_j) - \tilde{g}(x) = \epsilon g(P^\top u_k) - g(P^\top u_{k+1}) \geq 0$, since $P^\top e_k = e_j$. Hence the monotonicity in x_j is preserved. □

INRIA Bordeaux-Sud Ouest and Université Bordeaux I, HEC Montréal, Université Nice-Sophia-Antipolis

E-mail address: bruno.remillard@hec.ca