Pricing American options under multi-factor models with recursive adaptive sparse expectations*

Simon Scheidegger[†‡] Adrien Treccani[§]

May 29, 2016

Abstract

We introduce a fast numerical framework for pricing American options under multi-factor models and in the presence of discrete dividends. The algorithm scales well to dimensions higher than two and is massively parallelized to further speed up the solution process. We make use of adaptive sparse grids to numerically compute conditional expectations and recursively approximate the value function with a low number of points. We benchmark our scheme under 1-factor Gaussian settings and 2-factor stochastic volatility settings including discrete dividends, and illustrate the complexity-scaling with a basket option example.

*We thank Olivier Scaillet, Felix Kubler, Paola Pederzoli, Antonio Cosma and Ken Judd for their extremely valuable comments. We thank seminar participants at the University of Zürich for valuable comments. We gratefully acknowledge financial support from PASC. This work was supported by a grant from the Swiss National Supercomputing Centre (CSCS) under project ID s555. Additionally, we acknowledge CPU time granted on the University of Zurich's Piz Dora HPC cluster.

[†]Hoover Institution, Stanford University, Galvez Mall 434, 94305–603 Stanford, USA. Phone: +1 650 725 3416, Email: sischei@stanford.edu

[‡]University of Zürich, Plattenstrasse 32, 8032 Zürich, Switzerland. Voice: +41 22 379 88 16. Fax: +41 22 379 81 04. Email: simon.scheidegger@uzh.ch.

[§]University of Zürich, Plattenstrasse 32, 8032 Zürich, Switzerland. Voice: +41 22 379 88 16. Fax: +41 22 379 81 04. Email: adrien.treccani@uzh.ch.

1 Introduction

Options are derivative contracts giving the holder the right to buy or sell a security at a predetermined price. They are said to be European when the exercise only happens at a fixed expiration date, and American if they can be exercised during the whole lifetime of the contract. Black and Scholes [8] derived in their seminal paper a closed-form solution for the value of European options under the hypothesis of normal asset returns with constant volatility. The formula allows to analyze such derivatives efficiently even with limited computational resource. The case of American options on the other hand is extremely challenging. Even in the canonical Black and Scholes settings, there exists no closed-form pricing solution (see, e.g., [36] for a review). Several semi-analytical approximation have been developed (see, e.g., [4, 13, 17, 52]) but they cannot easily be extended to more complex asset dynamics. The reason for this complexity is that the analysis of American options involves a dynamic optimization problem: Claim-holders can decide at every point in time whether to exercise optionalities, or to wait in order to maximize the value of the claim.

It has empirically been established that financial returns exhibit heteroskedasticity and volatility clustering, as well as occasional discontinuities (see, e.g., [1, 21, 26, 56, 2]). Modeling these empirical properties adequately is crucial in order to obtain consistent derivative prices. Gaussian models fail to do so: They cannot reproduce the implied volatility smile observed on the markets without state-dependent volatility, also called local volatility [34]. In particular, the Black and Scholes model undervalues out-of-the-money options and short-term options because of higher-order moments (see, e.g, [49] for a review). The pricing of American options therefore has to rely on higher-dimensional models jointly with numerical techniques including lattice [30], finite differences [12, 27, 50], Monte Carlo methods [14, 59, 71, 41, 44, 35] and the resolution of the associated integral equations [54, 48, 75, 37]. At the same time, it is typical in previous literature (e.g., [8, 46, 6]) to model dividends with a continuous payment stream rather than with discrete cashflows. The reason is that discrete (in particular non-proportional) payments massively impact algorithmic performance of some pricing algorithms. Lattice methods for example can lose their recombining property or require ad-hoc methods for re-aligning the grid. Though, modeling discrete dividends adequately is crucial to avoid incorrect conclusions on the magnitude of the exercise boundary as well as biased prices [28].

The main contribution of the work presented here is to introduce a fast, scalable and flexible pricing framework for American options in multi-factor settings that can even adequately model the presence of discrete dividend payments. Our scheme—below denoted as RASE (Recursive Adaptive Sparse Expectations)—exploits the smoothness of the value function and the transition probability distribution to reduce the complexity of the multi-dimensional pricing problem. It offers a solution for pricing Amer-

ican options under models that capture many sources of risk, such as stochastic volatility, interest rates and exchange rates. Moreover, it also applies to options on multiple underlying assets. We use adaptive sparse grids [69, 68, 18, 61, 15, 16] in combination with piecewise linear basis functions to recursively approximate the value function with a very low number of points. We believe to be the first to apply adaptive sparse grid theory in the context of American options pricing with a probabilistic approach. The idea of approximating the value function is not new in the literature though (see, e.g., [63, 55, 24, 32, 25]). Notably, Cosma et al. [28] project the value function on a set of orthogonal basis functions and transmit coefficients of the representation from one date to the previous one by an explicit recursion formula. They demonstrate the computational advantage of their method when the number of exercise periods is relatively low. As they rely on a dense approximation technique, a strict application of their algorithm to higher dimensional problems faces the *curse of dimensionality*: The complexity grows exponentially and is prohibitive beyond two or three dimensions. The key focus of RASE is to alleviate this specific issue.

Most straightforward pricing techniques are based on equally-spaced grid structures: Starting with a one-dimensional discretization scheme that employs N grid points, a naive extension to d dimensions using tensor products leads to N^d grid points. This approach has two major drawbacks. The exponential dependency on the dimensionality imposes severe restrictions on the number of dimensions that can be dealt with, usually only up to four. Moreover, they invest the same amount of resources throughout the computational domain, even though the function of interest might show very distinct local features in some areas while displaying little variation in others. Sparse grids however can drastically improve on those issues. First, they reduce the number of grid points needed from the order $\mathcal{O}\left(N^d\right)$ to $\mathcal{O}\left(N \cdot (\log N)^{d-1}\right)$, while the accuracy of the interpolation only slightly deteriorates in the case of sufficiently smooth functions [19]. Additionally, as they allow for arbitrary adaptive refinement, the general fast convergence rates can even be improved further by adapting to the special characteristics of the function of interest (see, e.g., [69, 68, 18, 61]). By doing so, resources are only invested where needed. In the case of vanilla options for example, grid points are mostly concentrated around the exercise price and in-the-money in order to approximate the high local curvature, whereas few points are spent on the rest of the domain. Sparse grids date back to Smolyak [74] and have in the meantime been applied to a whole range of different research fields such as physics, visualization, data mining, Hamilton-Jacobi Bellman equations, mathematical finance, insurance, economics and econometrics [9, 19, 39, 22, 45, 47, 64, 15, 16, 77, 67].

RASE applies to models whose transition risk-neutral probability density function (pdf) is known or can be uncovered. In particular, this is the case for Gaussian settings such as Black and Scholes model or local volatility extensions (see, e.g., [34]) as well as the broad class of models with analytical

characteristic function. The stochastic volatility process of Heston [46], and the extension of Bates [6] which introduces jumps in the asset dynamics, are notable examples.[1] RASE has algorithmic complexity of $\mathcal{O}(NnM)$, where N is the number of points of the value function grid, n is the number of points used for numerical integration, and M is the number of time iterations. Because N and n typically grow relatively slowly with dimension thanks to the properties of adaptive sparse grids, the recursion—which corresponds to the repeated computation of conditional expectations—is extremely fast. RASE saves additional resources by accommodating irregularly-spaced time grids: M is kept small by only computing value functions at dates of interest. This strategy avoids massive waste of resources in the case of options for which information about optimal exercise policy is known ex-ante, such as Bermudan options [48] which have discretely spaced exercise periods, or vanilla American call options which are only exercised at dividend dates optimally [70]. Additionally, the high degree of intrinsic parallelism of RASE allows to dispatch our framework to parallel computing units in order to further speed up the time to solution. This feature, as well as the favorable scaling of adaptive sparse grids, are key to alleviate the *curse of dimensionality* in option pricing.

The article is organized as follows. Sec. 2 formally defines the problem and its primitives. It briefly reviews the theory behind adaptive sparse grids and describes the RASE pricing algorithm. Sec. 3 presents numerical results including benchmarks under respectively Black and Scholes model and Heston stochastic volatility model with discrete dividends. Complexity scaling is illustrated with the application of RASE to a basket option. Sec. 4 concludes our work.

2 Methodology

In this section, we first formally define the pricing problem at hand and its primitives. We then review the theory behind adaptive sparse grids. Finally, we present the RASE algorithm and explain how it also applies to models for which the pdf is computationally costly to evaluate.

2.1 Pricing problem

We consider a continuous-time economy with underlying probability space $(\Phi, \mathbb{F}, \mathscr{F}, \mathbb{P})$, where Φ is the sample space, $\mathscr{F}$ is a σ-algebra defining the set of possible scenarios, and $\mathbb{P}$ is the physical probability measure. The filtration $\mathbb{F} = (\mathscr{F}_t)_{t\geq 0}$ represents the information flow over time. Let $(\vec{x}_t)_{t\geq 0}$ denote a d-dimensional, $\mathbb{F}$-adapted Markovian process modeling the relevant factors of the pricing problem and taking values in $\mathscr{D} \subseteq \mathbb{R}^d$. We say that $(\vec{x}_t)_{t\geq 0}$ is the state process and that $\vec{x}_t = (x_{0,t}, \ldots, x_{d,t})$ is the state

[1] Fourier transforms of transition probabilities also often describe price evolution in affine models [38], quadratic models [57, 23], and variance gamma and Levy models [62, 20].

(or state vector) at time t. In the case of vanilla options under Black and Scholes assumptions, the state process is only defined by the value of the underlying asset; under stochastic volatility frameworks such as in Heston [46] and Bates [6], it additionally includes the variance process. More generally, the state can include any number of relevant drivers of the option price. However, the more factors, the more computationally complex the pricing problem becomes. RASE aims at providing an efficient scheme for pricing options when d is larger than two or three.

We assume the market to be arbitrage-free. The fundamental theorem of asset pricing reveals that absence of arbitrage on markets is essentially equivalent to the existence of an equivalent martingale measure [42, 43].[2] The martingale measure $\mathbb{Q}$, or risk-neutral measure, is obtained by adjusting the physical probability measure for preferences using the Radon-Nykodim derivative, and allows to price assets using riskless discounting (see, e.g., [72, 10, 11, 29]). Consider an option contract whose exercise value at time t is given by $H(t,\vec{x}_t)$, where $H : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$, and whose set of exercise dates larger or equal to t is defined by $\mathscr{E}(t)$. The value $P(t,\vec{x}_t)$ of the contract is then given by

$$P(t,\vec{x}) = \sup_{\tau} \mathbb{E}_{\mathbb{Q}}[e^{-r_f(\tau-t)} H(\tau,\vec{x}_\tau)|\vec{x}_t = \vec{x}], \tag{1}$$

where the supremum is taken over stopping times taking values in $\mathscr{E}(t)$, $\mathbb{E}_{\mathbb{Q}}$ is the expectation operator under the martingale measure, and r_f is the risk free rate.[3]

Based on Eq. 1, we formulate the following Bellman principle in the case of options with early-exercise features:

$$P(t,\vec{x}) = \begin{cases} 0 & \text{if } \mathscr{E}(t) = \varnothing, \\ \lim\limits_{\Delta t \to 0} \max(H(t,\vec{x}), \mathbb{E}_{\mathbb{Q}}[e^{-r_f \Delta t} P(t+\Delta t, \vec{x}_{t+\Delta t})|\vec{x}_t = \vec{x}]) & \text{if } t \in \mathscr{E}(t), \\ \mathbb{E}_{\mathbb{Q}}[e^{-r_f(\inf \mathscr{E}(t) - t)} P(\inf \mathscr{E}(t), \vec{x}_{\inf \mathscr{E}(t)})|\vec{x}_t = \vec{x}] & \text{otherwise.} \end{cases} \tag{2}$$

The first condition in Eq. 2 covers the case where the option contract has matured and can no longer be exercised, in which case it has value zero. The second condition accounts for the early-exercise option: The claim-holder decides whether it is more profitable to exercise the option immediately or to wait. The final case treats periods during which early-exercise is contractually not possible.

Numerically, we set the pricing problem up as follows: We discretize the time dimension on a grid $\mathscr{E}(0) \supset \mathscr{T} = (t_i)_{i=0}^M$ satisfying $0 = t_0 < t_1 < \cdots < t_M = T$, where $T > 0$ is the expiration date of the option given by the largest element of $\mathscr{E}(0)$. We also define the time interval $\Delta_i = t_{i+1} - t_i$ for $i = 0, \ldots, M-1$.

[2] This result holds for a broad but restricted set of so called admissible trading strategies. Also, a finer notion of arbitrage referred as *no free lunch with vanishing risk* [33] may be considered in continuous-time.

[3] We limit our analysis to constant risk free rate models for simplicity. Our theory can be extended in a straightforward manner to time-varying interest rates models.

Note that we do not require times $(t_i)_i$ to be equidistant: The specific discretization scheme typically depends on the individual properties of the option contract and its underlying asset. In the case of American call options where the underlying asset pays no dividend, it is never optimal to exercise early; American call options therefore have the same value as European call options and we can set $M = 1$. In the presence of discrete dividends, American call options are only exercised at dividend dates optimally.[4] M may be kept small in this case by aligning the grid with the few dividend payments. In general settings however, optimal exercise could happen at any instant in $\mathscr{E}(0)$: The grid must be dense enough to mitigate the bias inherent to suboptimal exercise. The discrete-time version of Eq. 2 takes the recursive form

$$P_i(\vec{x}) = \begin{cases} H(T, \vec{x}_T) & \text{if } i = M, \\ \max(H(t_i, \vec{x}), E_i(\vec{x})) & \text{otherwise,} \end{cases} \tag{3}$$

for $i = 1, \ldots, M$, where

$$E_i(\vec{x}) = \mathbb{E}_{\mathbb{Q}}[e^{-r_f \Delta_i} P_{i+1}(\vec{x}_{t_{i+1}}) | \vec{x}_{t_i} = \vec{x}] \tag{4}$$

is the continuation value of the option. Our aim now is to provide a computationally efficient algorithm to approximate the contemporaneous value function $P(0, \cdot)$.

In order to solve the dynamic programming problem stated in Eq. 3 iteratively (see, e.g. [58, 53]), we need to repeatedly evaluate value functions at arbitrary coordinates within the domain of interest. As a single function evaluation can be very expensive — it is only known algorithmically — we need an efficient interpolation method. Moreover, standard grid-based approaches of interpolating a function exhibit the *curse of dimensionality* [7]: Any naive discretization scheme that employs N grid points in one dimension leads to N^d grid points in d dimensions, a prohibitive obstacle if one aims for an algorithm that delivers a fast time-to-solution in a multidimensional setting.

2.2 Sparse grid interpolation

Our method of choice to tackle the issues raised above is the adaptive sparse grid interpolation, which we now explain briefly below. We first describe the interpolation on Cartesian grids using hierarchical basis functions, from which we then derive the interpolation on an (adaptive) sparse grid by omitting the basis functions that contribute only little to the overall interpolant. For details regarding derivations, we refer to [19, 61, 69], and to [15, 16] for its highly parallelized implementation in the context of dynamic

[4] We refer to [70] for an elaborate explanation of theoretically optimal exercise policies. Note also that the presence of frictions can make the analysis of early-exercise strategies more complex, as argued in [51].

programming and time iteration problems.

We consider the piecewise $d-$linear interpolant u of a function $f : \Omega \to \mathbb{R}$ which is only algorithmically given (c.f., Sec. 2.1):

$$f(\vec{x}) \approx u(\vec{x}) := \sum_j \alpha_j \phi_j(\vec{x}). \tag{5}$$

As we aim to discretize Ω, we restrict our domain of interest to the compact subvolume $\Omega = [0,1]^d$, where d is the dimensionality of the pricing problem. This situation can be achieved for most other domains by rescaling and possibly carefully truncating the original domain. Moreover, let $\vec{l} = (l_1, ..., l_d) \in \mathbb{N}^d$ and $\vec{j} = (j_1, ..., j_d) \in \mathbb{N}^d$ denote multi-indices, and define $|\vec{l}|_1 := \sum_{t=1}^d l_t$ and $|\vec{l}|_\infty := \max_{1 \le t \le d} l_t$. The equidistant one-dimensional interpolant consists then of a combination of nested one-dimensional grids of different refinement levels $l \in \mathbb{N}$. For a given level l, the grid points are distributed as

$$x_{l,j} = \begin{cases} 0.5, & l = j = 1, \\ j \cdot 2^{1-l}, & j = 0, ..., 2^{l-1},\ l > 1, \end{cases} \tag{6}$$

as depicted in Fig. 1. The corresponding piecewise linear basis functions for $x \in [0,1]$ are given by

$$\phi_{l,j}(x) \quad = \begin{cases} 1, & l = i = 1, \\ \max(1 - 2^{l-1} \cdot \left| x - x_{l,j} \right|, 0) & j = 0, ..., 2^{l-1},\ l > 1 \end{cases} \tag{7}$$

and allow for non-vanishing boundaries.[5] The one-dimensional basis functions can easily be extended to d-dimensional ones by a tensor product construction. For each grid point $\vec{x}_{\vec{l},\vec{j}} = (x_{l_1,j_1}, ..., x_{l_t,j_t})$, an associated piecewise d-linear basis function $\phi_{\vec{l},\vec{j}}(\vec{x})$ is then namely defined as the product of the one-dimensional basis functions

$$\phi_{\vec{l},\vec{j}}(\vec{x}) := \prod_{t=1}^d \phi_{l_t,j_t}(x_t). \tag{8}$$

The so-called hierarchical increment spaces $W_{\vec{l}}$ are defined by

$$W_{\vec{l}} := \operatorname{span}\{\phi_{\vec{l},\vec{j}} : \vec{j} \in I_{\vec{l}}\} \tag{9}$$

[5]Note that there exist also formulations of sparse grids that are based e.g. on local polynomial basis functions (see, e.g.,[18]). However, the piecewise linear basis functions provide features that are more favourable for option pricing problems that show very distinct local features.

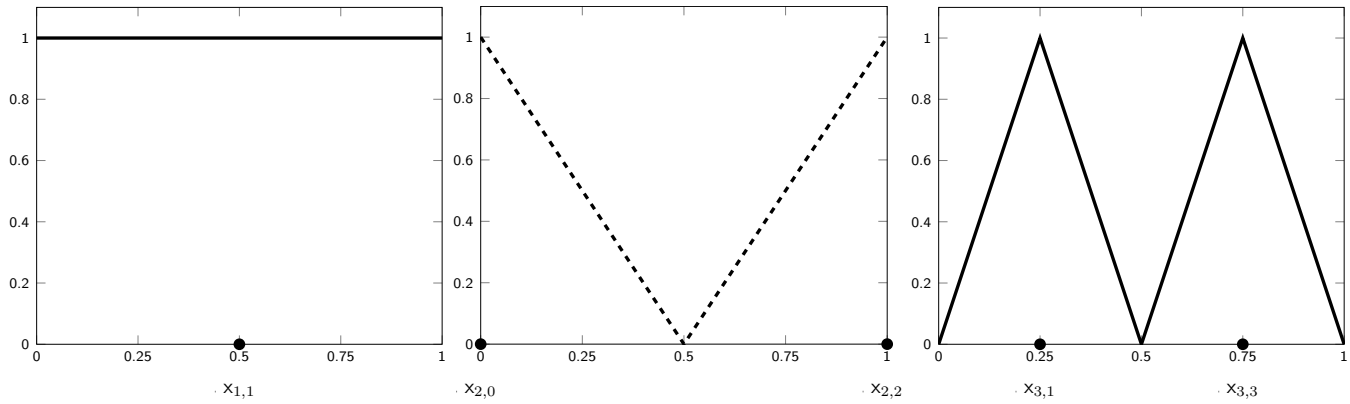


Figure 1: Panels from left to the right: one-dimensional hierarchical basis functions of levels 1 (left), 2 (middle) and 3 (right). $V_3 = W_1 \bigoplus W_2 \bigoplus W_3$.

with the index set $I_{\vec{l}}$ given as

$$I_{\vec{l}} := \begin{cases} \{\vec{j} : j_t = 1, 1 \leq t \leq d\} & \text{if } l = 1, \\ \{\vec{j} : 0 \leq j_t \leq 2,\ j_t \text{ even}, 1 \leq t \leq d\} & \text{if } l = 2, \\ \{\vec{j} : 0 \leq j_t \leq 2^{l_t - 1},\ j_t \text{ odd}, 1 \leq t \leq d\} & \text{else.} \end{cases} \tag{10}$$

Fig. 1 depicts the first three levels of the associated $1d$ hierarchical, piecewise linear basis functions. The hierarchical increment spaces $W_{\vec{l}}$ are related to the space V_n of $d-$*linear* functions by

$$V_n := \bigoplus_{|l|_\infty \leq n} W_{\vec{l}}. \tag{11}$$

The actual interpolant of f, namely, $u(\vec{x}) \in V_n$, can uniquely be represented by

$$f(\vec{x}) \approx u(\vec{x}) = \sum_{|l|_\infty \leq n} \sum_{\vec{j} \in I_{\vec{l}}} \alpha_{\vec{l},\vec{j}} \cdot \phi_{\vec{l},\vec{j}}(\vec{x}) \tag{12}$$

with $\alpha_{\vec{l},\vec{j}} \in \mathbb{R}$. Note that the coefficients $\alpha_{\vec{l},\vec{j}} \in \mathbb{R}$ are commonly termed the *hierarchical surpluses* in the sparse grid literature (see, e.g., [78, 19]). These coefficients can easily be determined due to the nested structure of the hierarchical grid. The set of points X^{l-1} at level $l-1$ is contained in X^l (i.e. $X^{l-1} \subset X^l$). The hierarchical coefficients along one dimension are given by:

$$\begin{aligned} \alpha_{l,j} &= f\left(x_{l,j}\right) - \frac{f\left(x_{l,j} - h_l\right) + f\left(x_{l,j} + h_l\right)}{2} \\ &= f\left(x_{l,j}\right) - \frac{f\left(x_{l-1,(j-1)/2}\right) + f\left(x_{l-1,(j+1)/2}\right)}{2}. \end{aligned} \tag{13}$$

In operator form, Eq. 13 can be rewritten as

$$\alpha_{l,i} = \left[-\frac{1}{2} \;\; 1 \;\; -\frac{1}{2} \right]_{l,j} f. \tag{14}$$

We denote the mesh-size of the Cartesian grid of refinement level n by $h_n = 2^{1-n}$. For the the interpolant $u \in V_n$ [19], an asymptotic error decay of

$$\| f(\vec{x}) - u(\vec{x}) \|_{L_2} \in \mathscr{O}\left(h_n^2\right) \tag{15}$$

at the cost of

$$\mathscr{O}\left(h_n^{-d}\right) = \mathscr{O}\left(2^{nd}\right) \tag{16}$$

function evaluations is achieved, encountering the *curse of dimensionality*. Luckily, it can be shown (see, e.g., [19, 78]) that if the function f which we want to approximate meets certain smoothness conditions, a so-called sparse grid can be constructed. The catch hereby is that if the second mixed derivatives of f are bounded, the coefficients of the interpolant decay rapidly, namely as $|\alpha_{\vec{l},\vec{j}}| = \mathscr{O}\left(2^{-2|\vec{l}|_1}\right)$. The strategy for constructing a sparse grid thus is to leave out those subspaces among the full grid space V_n that only contribute little to the interpolant [19]. This can be done a priori as we can derive bounds for the contributions of the the different subspaces. The sparse grid space V_n^S of level n is then defined by

$$V_n^S := \bigoplus_{|\vec{l}|_1 \leq n+d-1} W_{\vec{l}}. \tag{17}$$

Fig. 2 shows the selection of the subspaces and the resulting sparse grid for n =3, i.e. V_3^S.[6] The number of grid points required by the space V_n^S is now given by (see, e.g., [19, 39]).

$$|V_n^S| = \mathscr{O}\left(h_n^{-1} \cdot \left(\log(h_n^{-1})\right)^{d-1}\right). \tag{18}$$

This is of order $\mathscr{O}\left(2^n \cdot n^{d-1}\right)$, which is a significant reduction of the number of grid points, and thus of the computational and storage requirements compared to $\mathscr{O}\left(2^{nd}\right)$ of the full grid space $|V_n|$. In analogy to Eq. 12, a function $f \in V_n^S \subset V_n$ can now be approximated by

$$f \approx u(\vec{x}) = \sum_{|l|_1 \leq n+d-1} \sum_{\vec{j} \in I_{\vec{l}}} \alpha_{\vec{l},\vec{j}} \cdot \phi_{\vec{l},\vec{j}}(\vec{x}). \tag{19}$$

The asymptotic accuracy of the interpolant deteriorates only slightly from $\mathscr{O}\left(h_n^2\right)$ in the case of the

[6] Note that the concrete choice of subspaces depends on the norm in which we measure the error. The result derived in Eq. 17 is optimal for the L_2-norm and the L_∞-norm [19].

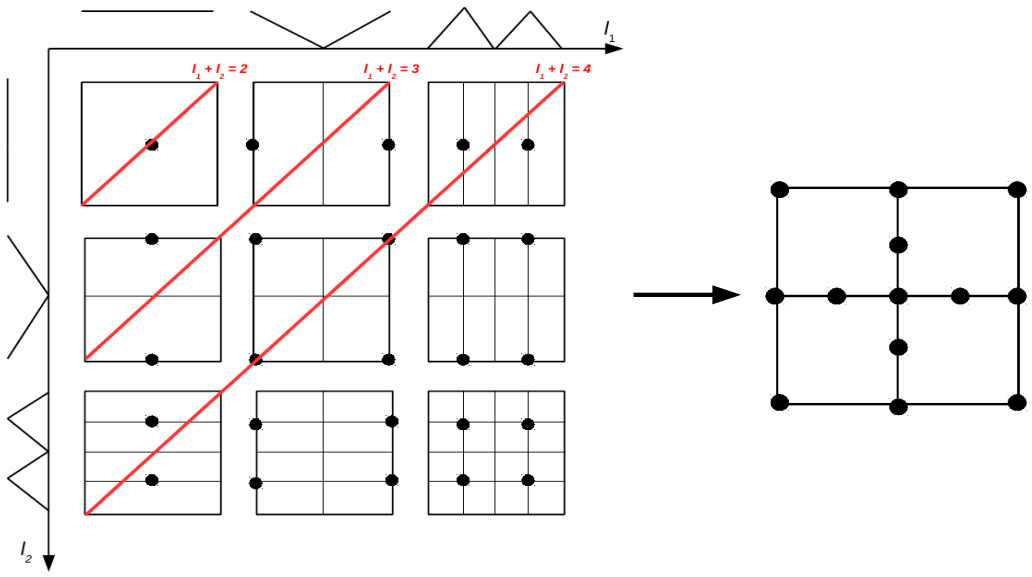


Figure 2: **Sparse grid subspaces.** Left panel: the two-dimensional subspces $W_{\vec{l}}$ up to l=3 in each dimension as well as the corresponding one-dimensional basis functions. The optimal selection of subspaces is indicated by the red lines of constant $l_1 + l_2$. Right panel: The resulting sparse grid of level 3 (see, Eq. 17). Note that an ordinary Cartesian grid with this setting would consist of 25 points, whereas the corresponding sparse grid only contains 13 points.

full grid (cf. Eq. 15) down to $\mathscr{O}\left(h_n^2 \cdot \log(h_n^{-1})^{d-1}\right)$, as shown in [19, 39]. Taken together, this clearly demonstrate why sparse grids favourable to use in our application. In contrast to full grids, they require significantly fewer grid points than full grids, while preserving then asymptotic error decay of the full grid interpolant with increasing grid resolution up to a logarithmic factor, allowing for a significant faster time-to-solution.

Functions which do not meet the smoothness requirements or that show significantly varying characteristics across the domain of interest such as steep gradients in some regions as well as flat ones in another ones (e.g., the terminal payoff of an option or economic models with occasionally binding constraints [16, 15]) can still be tackled as well if adaptivity is used. The sparse-grid structure introduced in Eq. 18 defines an a priori selection of grid points that is optimal for function with bounded second-order mixed derivatives. An adaptive (a posteriori) refinement can additionally select which grid points in the sparse grid structure should be refined due to some local criterion (see, e.g., [68, 18, 61]). The most common way of doing so is the following heuristics: When approximating a function as a sum of piecewise linear basis functions, the main contributions to the interpolant (most likely) stem from comparatively few terms with big surpluses (cf., Eq. 19). The logic of the refinement strategies is therefore to monitor the size of the hierarchical surpluses. The magnitude of the hierarchical surplus reflects the local irregularity of the function and thus serves as a natural error indicator. $2d$ *children* in the hierar-

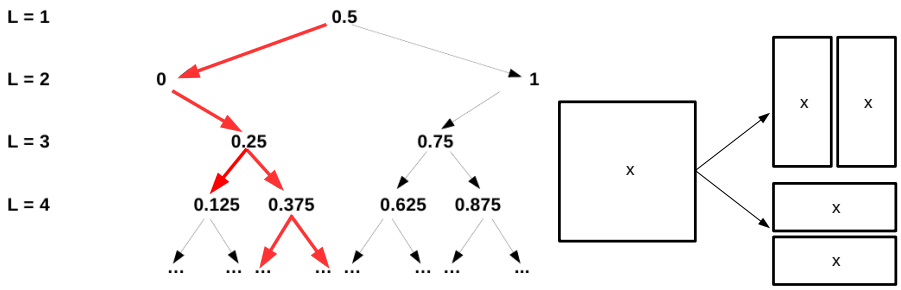


Figure 3: **Adaptive sparse grid refinement.** Left panel: one-dimensional tree-like structure of a classical sparse grid for the first four levels. The red arrows symbolically depict where the hierarchical surpluses satisfy $g(\alpha_{\vec{l},\vec{j}}) \geq \varepsilon$ and thus the gris is refined. Black arrows represent the part of the sparse grid data structure that is left out due to adaption. Right panel: example of nodes and supports of a locally refined sparse grid in two dimensions. $2d = 4$ *children*, which are neighboring points of the *parent*, are subsequently added to the grid.

chical structure (except for the first refinement level) are added to the current grid for those hierarchical basis functions $\phi_{\vec{l},\vec{j}}$ that have a hierarchical surplus, $\alpha_{\vec{l},\vec{j}}$, that satisfy $|\alpha_{\vec{l},\vec{j}}| \geq \varepsilon$ for a so-called refinement threshold $\varepsilon \geq 0$. Whenever this criterion is satisfied, the *children* of the current point are added to the sparse grid. Let us consider a one-dimensional example. The left panel of Fig. 3 schematically displays how grid points are adaptively added: At every node, the adaptive sparse grid algorithm checks whether the refinement threshold is satisfied. If so, points are added. Else, it leaves the grid it unaltered. The right panel of Fig. 3 shows the workings for adaptive sparse grid refinement for an arbitrary point in $2d$. Note that if additional knowledge about the problem at hand is available, it can be used in the criterion for adaptive refinement, allowing to better adapt the problem specific characteristics (see, e.g., [16]).[7] For more details regarding adaptive sparse grids, we refer e.g. to [18, 61, 16, 68].

One feature that makes adaptive sparse grids particular useful in the context of option pricing is that they can also be used for high-dimensional numerical integration, such as the computation of expectations (see [40, 61, 65]). Starting from Eq. 19, the expected value of the interpolant can be evaluated as follows:

$$\mathbb{E}[u(\vec{x})] = \sum_{|l|_1 \leq n+d-1} \sum_{\vec{i} \in I_{\vec{l}}} \alpha_{\vec{l},\vec{i}} \int_\Omega \phi_{\vec{l},\vec{i}}(\vec{x}) d\vec{x}, \tag{20}$$

where we assume here for simplicity that the probability density is 1 on $\Omega = [0,1]^d$. Using the basis functions described by Eq. 7, the one dimensional integral can now easily be computed analytically by

$$\int_0^1 \phi_{l,i}(x)\, dx = \begin{cases} 1, & \text{if } l = 1 \\ \frac{1}{4} & \text{if } l = 2 \\ 2^{1-l} & \text{else.} \end{cases} \tag{21}$$

[7] In the context of option pricing, we typically impose a stricter refinement threshold close to the exercise price and a more permissive threshold deep in- and out-of-the-money.

Note that this result is independent of the location of the interpolation point due to dilation and translation properties of the hierarchical basis functions. The multi-dimensional integrals directly follow the product of one dimensional integrals. Denoting $\int_\Omega \phi_{l,i}(\vec{x})\, d\vec{x} = J_{\vec{l},\vec{i}}$, we can rewrite Eq. 20 as

$$\mathbb{E}[u(\vec{x})] = \sum_{|l|_1 \leq n+d-1} \sum_{\vec{i} \in I_{\vec{l}}} \alpha_{\vec{l},\vec{i}} \cdot J_{\vec{l},\vec{i}}. \tag{22}$$

Eq. 22 states that the expectation is given by the arithmetic sum over all grid points of the product between the hierarchical surpluses and the integral weights.

Another key advantage of the (adpative) sparse grid data structure is its inherent parallelism that we exploit in order achieving a fast time-to-solution process. At each point of the (adaptive) sparse grid, we have to evaluate a function that is only known algorithmically. In option pricing, the value at a point typically depends on the computation of a conditional expectation. Fortunately, within each iteration step, these tasks are independent from each other and can thus be solved in parallel. Our implementation follows [16, 15, 16, 15], who introduced a so-called hybrid parallelization scheme that is based on distributed and shared memory parallel programming paradigms.

2.3 RASE algorithm

We now make use of adaptive sparse grids in order to solve the dynamic pricing problem stated in Eq. 3. In typical applications, value functions are smooth with rare kinks, depending on the features of the option. Adaptive sparse grids are powerful tools in this context as they can reach high density at non-smooth regions and yet maintain an overall low number of points globally. Our strategy is to recursively build adaptive sparse grids for value functions $P_i(\vec{x})$ from index $i = M-1$ to $i = 0$ through the repeated application of a fast numerical integration scheme. We first present the time iteration algorithm below, which returns the contemporaneous value of the option on a compact subdomain of the state space. Moreover, we explain that hedging coefficients can be determined with simple finite differences. We then introduce a precomputation method allowing fast evaluation of the pdf even when the function is not known in closed-form. This is key to the efficient computation of conditional expectations when evaluation of pdf else would be slow—for instance in the case of pdfs which require a Fourier inversion of the characteristic function. Finally, we explain how RASE copes with discrete dividends—proportional, constant-size or even state-contingent.

2.3.1 Time iteration

We show in Eq. 3 how the value function $P_i(\vec{x})$ depends recursively on the conditional expectation in $E_i(\vec{x})$, which characterizes the continuation value and the exercise policy. The integral form of Eq. 4 is given by

$$E_i(\vec{x}) = e^{-r_f \Delta_i} \int_{\mathscr{D}} P_{i+1}(\vec{y}) f_{\vec{x}}(\vec{y}; \Delta_i) d\vec{y}, \tag{23}$$

where $f_{\vec{x}}(\vec{y}; \Delta) := f_{\vec{x}}(\vec{y}; \Delta, \Theta)$ is the pdf of the transition from state $\vec{x}$ to state $\vec{y}$ on a time interval Δ, and with model parameters $\Theta = (\theta_1, \ldots, \theta_q)$. Models of interest typically have no closed-form solution for the multidimensional integral in Eq. 23 and require numerical computation. The process of evaluating the integral is repeated at every time iteration i and for every relevant state $\vec{x}$: The number of evaluation points $\vec{x}$, as well as the efficiency of numerical integration, are therefore dominant contributors to the overall complexity of RASE. Naive numerical integration algorithms face the *curse of dimensionality* in the sense that the complexity increases exponentially with the dimension of the problem.[8] Even in low dimension, limiting the number of grid points in the numerical integration process is critical as each evaluation of the pdf can be computationally costly depending on the model.

Our algorithm presented hereafter relies on the theory of Sec. 2.2 to alleviate the curse of dimensionality: We use adaptive sparse grids to store the value function of an individual timestep with a low number of points. Moreover, we apply sparse integration (cf., Eq. 22) to solve the numerous instances of Eq. 23 efficiently. Alg. 1 formalizes the full time iteration algorithm.

We consider a d-dimensional compact subdomain $\mathscr{D}^x$ of state space $\mathscr{D}$. We say that $\mathscr{D}^x$ is the truncated domain of the state process, whose boundaries are defined such that it only excludes arbitrarily low probability states.[9] We also define the strictly increasing function $u : \mathscr{D}^x \to [0,1]^d$ mapping elements of the truncated domain to the unit hyper-cube. Similarly, we consider the family of compact subdomains $\mathscr{D}^y_{\vec{x},\Delta} \subset \mathscr{D}^x$ for $\vec{x} \in \mathscr{D}^x$ and $\Delta > 0$, as well as the family of strictly increasing functions $v_{\vec{x},\Delta}(\cdot) : \mathscr{D}^y \to [0,1]^d$ mapping elements of $\mathscr{D}^y_{\vec{x},\Delta}$ to the unit hyper-cube. We say that $\mathscr{D}^y_{\vec{x},\Delta}$ is the truncated image of $f_{\vec{x}}(\cdot;\Delta)$, which only excludes states that are conditionally negligible in probability. In typical applications, it is sufficient to consider d-orthotope $\mathscr{D}^x = [\bar{d}_1^L, \bar{d}_1^R] \times \cdots \times [\bar{d}_d^L, \bar{d}_d^R]$, $\bar{d}_i^L < \bar{d}_i^R \ \forall i = 1, \ldots, d$, paired with the linear mapping $u(\vec{x}) = (\frac{x_1 - \bar{d}_1^L}{\bar{d}_1^R - \bar{d}_1^L}, \ldots, \frac{x_d - \bar{d}_d^L}{\bar{d}_d^R - \bar{d}_d^L})$; the same remark applies to the definition of the image domains and mappings.

We initiate the algorithm by defining $\tilde{P}_M(\cdot) := H(T, \cdot)$ and setting $i := M - 1$. The iteration begins

[8] Additionally, theoretical complexity investigations reveal that lower bounds for the computing cost grow exponentially for many integration problems [66]. See, e.g., [18] for integration on sparse grids.

[9] The full, possibly unbounded domain $\mathscr{D}$ can be used without truncation if exercise value function $H(\cdot,\cdot)$ is uniformly bounded. We favor truncation even in this context to avoid waste of computational resources on negligible probability states.

Data: Time grid $\mathscr{T}$. Exercise function $H(t,\vec{x})$. Pdf $f_{\vec{x}}(\vec{y};\Delta,\Theta)$. Dividend functions $\delta_i(\vec{x})$. Mappings $u(\vec{x})$ and $v_{\vec{x},\Delta}(\vec{y})$. Minimal and maximal refinement levels $L_0 \leq L_{max}$. Refinement threshold ε.
Result: Approximate value function $\tilde{P}_0(\cdot)$.

Define $p_+(\vec{u}) = H(T, u^{-1}(\vec{u}))$.
Set $i = M-1$.
while $i \geq 0$ **do**
    Set $k = 1$, set $G \subset [0,1]^d$ to be the level 1 grid on $[0,1]^d$, and set $G_{old} = \emptyset, G_{new} = \emptyset$.
    while $G \neq G_{old}$ **do**
        for $\vec{u} \in G \setminus G_{old}$ **do**
            Set $\vec{x} = u^{-1}(\vec{u})$ and $\vec{x}' = \delta_i(\vec{x})$.
            Set $p(\vec{u}) = \max(H(t_i,\vec{x}), \tilde{E}_i(\vec{x}'))$. ; `// $\tilde{E}_i(\vec{x}')$ as given in Eq. 28`
            Define $\tilde{p}(\vec{u})$ by interpolating $\{p(u)\}_{u \in G_{old}}$.
            if *($k < L_{max}$ and $|\alpha_{\vec{k},\vec{j}}| \geq \varepsilon$) or $k < L_0$* **then**
                Add the neighboring points (*children*) of u to G_{new}.
            end
        end
        Set $G_{old} = G$, set $G = G_{old} \cup G_{new}$, set $G_{new} = \emptyset$, and set $k = k+1$.
    end
    Define $p_+(u) = \tilde{p}(u)$.
    Set $i = i-1$.
end
Define the approximate value function $\tilde{P}_0$ as the sparse grid interpolation of $\{p(u)\}_{u \in G}$.

Algorithm 1: Time iteration of RASE.

the construction of an adaptive sparse grid approximation of value function $P_i(\vec{x})$ on $\vec{x} \in \mathscr{D}^x$. For each point $\vec{p} \in [0,1]^d$, we compute the approximate continuation value $\tilde{E}_i(u^{-1}(\vec{p}))$ and set the value of the grid at point $\vec{p}$ to $\max(H(t_i, u^{-1}(\vec{p})), \tilde{E}_i(u^{-1}(\vec{p})))$. The approximate value function $\tilde{P}_i(\vec{x})$ at time t_i is then given by

$$P_i(\vec{x}) \approx \tilde{P}_i(\vec{x}) := \sum_{\vec{k},\vec{j}} \alpha_{i,\vec{k},\vec{j}} \phi_{\vec{k},\vec{j}}(u(\vec{x})), \tag{24}$$

where the multi-indices $\vec{k}, \vec{j}$ and as well as the bases $\phi_{\vec{k},\vec{j}}$ are given as in Sec. 2.2. $\alpha_{i,\vec{k},\vec{j}}$ denote surpluses at time i.

We set $i := i-1$ and repeat the process until time index $i = 0$, building successive approximations of the value function $\tilde{P}_i(\vec{x})$ on the truncated domain $\mathscr{D}^x$. The ultimate grid defines the output of RASE:

$$\tilde{P}_0(\vec{x}) \approx P(0,\vec{x}) \tag{25}$$

for any $\vec{x} \in \mathscr{D}^x$. The major difficulty lies in the computation of the approximate continuation value $\tilde{E}_i(\cdot)$. We consider the integrand of Eq. 23

$$G_{i,\vec{x}}(\vec{y}) = e^{-r_f \Delta_i} \tilde{P}_{i+1}(\vec{y}) f_{\vec{x}}(\vec{y}; \Delta_i) \tag{26}$$

for $\vec{x} \in \mathscr{D}^x$ and $\vec{y} \in \mathscr{D}^y_{\vec{x},\Delta_i}$ such that $E_i(\vec{x}) \approx \int_{\mathscr{D}^y_{\vec{x},\Delta_i}} G_{i,\vec{x}}(\vec{y}) d\vec{y}$, which we approximate on an adaptive sparse grid:

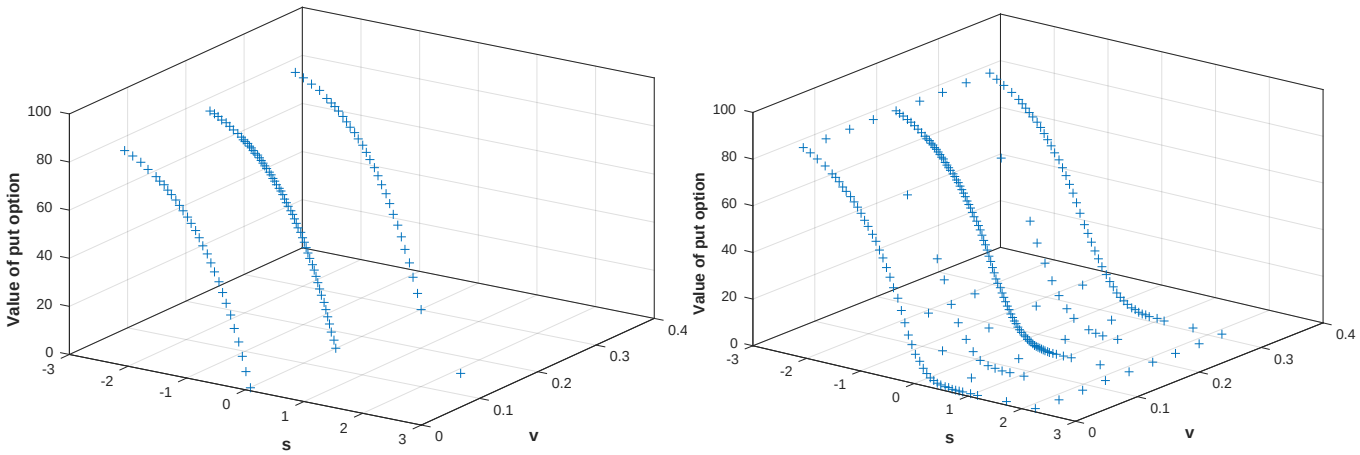
$$G_{i,\vec{x}}(\vec{y}) \approx \tilde{G}_{i,\vec{x}}(\vec{y}) := \sum_{\vec{k},\vec{j}} \beta_{i,\vec{x},\vec{k},\vec{j}} \phi_{\vec{k},\vec{j}}(v_{\vec{x},\Delta_i}(\vec{y})) \tag{27}$$

where $(\beta_{i,\vec{x},\vec{k},\vec{j}})_{\vec{k},\vec{j}}$ are the respective surpluses. We define the approximate continuation value as the integral of $\tilde{G}_{i,\vec{x}}(\cdot)$ given by

$$E_i(\vec{x}) \approx \tilde{E}_i(\vec{x}) := \sum_{\vec{k},\vec{j}} \beta_{i,\vec{x},\vec{k},\vec{j}} \int_{\mathscr{D}^y_{\vec{x},\Delta_i}} \phi_{\vec{k},\vec{j}}(v_{\vec{x},\Delta_i}(\vec{y})) d\vec{y}. \tag{28}$$

As explained in 2.2, we emphasize that for appropriate mapping functions $v_{\vec{x},\Delta}$ (e.g. linear mappings) the integrals in Eq. 28 are known in closed-form: The multidimensional integral of Eq. 23 is transformed into a series of trivial one-dimensional integrals. Assuming constant evaluation time of the value function and the pdf, the computational complexity of the continuation value at a given state $\vec{x}$ is therefore of $\mathscr{O}(n)$, where n is the number of points of the adaptive sparse grid approximating $G_{i,\vec{x}}(\cdot)$.

Fig. 4 illustrates the value function of a put option under two-factor Heston stochastic volatility



(a) At maturity

(b) One year before maturity

Figure 4: **Put option value approximation.** Approximation of a European put option value with exercise price 100 on an adaptive sparse grid (a) at maturity, and (b) one year before maturity. The pricing model follows two-factor stochastic volatility dynamics of Heston [46] with risk-free rate of 3%, mean-reversion of variance of 3, long-term volatility of 30%, volatility of variance of 20% and initial volatility of 15%. The grid has refinement criterion $\varepsilon = 10^{-3}$ and maximum level $L_{\max} = 12$. s-axis stands for log-moneyness, v axis stands for variance level. Grid (a) has 116 points, only one of which is out-of-the-money. The sparsity in the variance dimension is notable due to the absence of impact of variance at maturity. Grid (b) has 220 points and illustrates the impact of volatility on the value function. Most new points are close to the exercise price and out-of-the-money.

model, as approximated on an adaptive sparse grid. We observe the very low number of points required for a good fit of the value function: The respective grids do not exceed 250 points with a refinement threshold of $\varepsilon = 10^{-3}$. At maturity, the value function is flat out-of-the-money as well as in the variance dimension: Points are concentrated in-the-money. In general, we obtain a concentration of points at-the-money due to the high local curvature. The grid is otherwise sparse out-of-the-money where the slope is minimal, and relatively sparse in-the-money.[10] More notably, we can see the extreme sparsity in the variance dimension due to smooth, and relatively minor, impact of variance on the option price. An interesting aspect here is that the value function gets smoother at every iteration because of the application of the expectation operator, therefore reducing the number of necessary basis functions for a given precision.

Because RASE returns the value function $\tilde{P}_0(\cdot)$ on the entire domain $\mathscr{D}^x$, one execution of Alg. 1 provides option prices for any state of interest. Not only is it very practical for pricing an entire cross-section of options in a single execution, but it also provides an immediate way to estimate the hedging coefficients—or so-called *Greeks*. In theory, *Greeks* are obtained by no-arbitrage and define a perfect dynamic hedging strategy in frictionless complete markets (see, e.g., [31]).[11] The assessment of the

[10] We can even decrease further the number of points in-the-money by approximating the log value function rather than the value function itself—or by working on the refinement criterion to make it less aggressive when the option is deep in-the-money.

[11] Actual markets are subject to frictions (e.g., trading fees) and are generally incomplete. For instance, volatility is not a tradeable asset, preventing the implementation of a perfect hedge in the case of Heston [46] dynamics (see [49] for an extensive review). In this context, it is typical to minimize the risk of non-traded assets through an optimization problem (see, e.g., [5]).

derivative sensitivity to relevant underlying factors, such as the underlying asset price and volatility, is therefore crucial for risk management and hedging. In the context of RASE, we evaluate the *Greeks* via direct application of finite difference operator on $\tilde{P}_0(\cdot)$.

2.3.2 Precomputation of probability density function

The efficiency of the iterative algorithm presented above strongly depends on our ability to evaluate the transition probability density function $f_{\vec{x}}(\vec{y};\Delta,\Theta)$ as fast as possible for any tuple $(\vec{x},\vec{y},\Delta,\Theta)$. In specific contexts such as with Gaussian dynamics, the pdf is known in closed-formed or can be evaluated efficiently. Many important models studied in asset pricing literature have no closed-form pdf and can be computationally intensive to evaluate. This is particularly true for models with known characteristic function whose pdf is only known up to a Fourier transform [46, 6, 38, 57, 23, 62, 20]. Numerical Fourier inversion is computationally involved and would slow RASE execution down dramatically.[12]

We now introduce the novel idea of mapping the high-dimensional function $f(\vec{x},\vec{y},\Delta,\Theta) := f_{\vec{x}}(\vec{y};\Delta,\Theta)$ onto an adaptive sparse grid through an initial precomputation process. This computation, performed on our massively parallel framework, has to be done only once (see, [15, 16]). The grid entirely defines the pdf of a given model for any set of parameters on the truncated domain: We can use it to value any sort of derivative contract satisfying the same modeling assumptions without additional precomputation requirement. In the course of pricing an option with RASE, we hence only have to call the generated interpolant, which creates a negligible overhead to the overall computational burden.

Fig. 5 illustrates an adaptive sparse grid approximating the Gaussian pdf for different levels of volatility, which can be used for inference under Black and Scholes settings. As expected, we observe a concentration of points around the mode of the distribution and at the beginning of the tails due to high local curvature. The function requires otherwise very few points, summing up to 1369 coordinates in total with a refinement threshold of $\varepsilon = 10^{-3}$.[13]

With d state variables, q parameters and the time interval Δ, the dimension of a naive grid for the pdf is capped by $2d+q+1$. However, it is often easy to reduce the size of the problem ex-ante. First, an immediate way to save one dimension is to set a fixed $\Delta = \Delta_i = T/N$ for all $i = 1,\ldots,M$ resulting in an equaly-spaced time grid $\mathscr{T}$. We can then build an adaptive sparse grid for the simpler function $f(\vec{x},\vec{y},\Theta) := f(\vec{x},\vec{y},\Delta,\Theta)$ where Δ is set to some positive constant. The flexibility brought by irregular time grids for matching specific events (e.g., dividend dates or exercise periods) being a key advantage

[12] We refer to [76] for an extensive review of Fourier calculus and numerical implementations of the Fourier transform.

[13] In addition, we can rescale the pdf before mapping it to the grid in order to decrease the number of points even more. One simple way in the Gaussian case would be to scale the pdf by the standard deviation parameter. We note also that due to the presence of slight oscillations in the tails of the interpolated function, we set to zero any value below a small, arbitrary positive threshold.

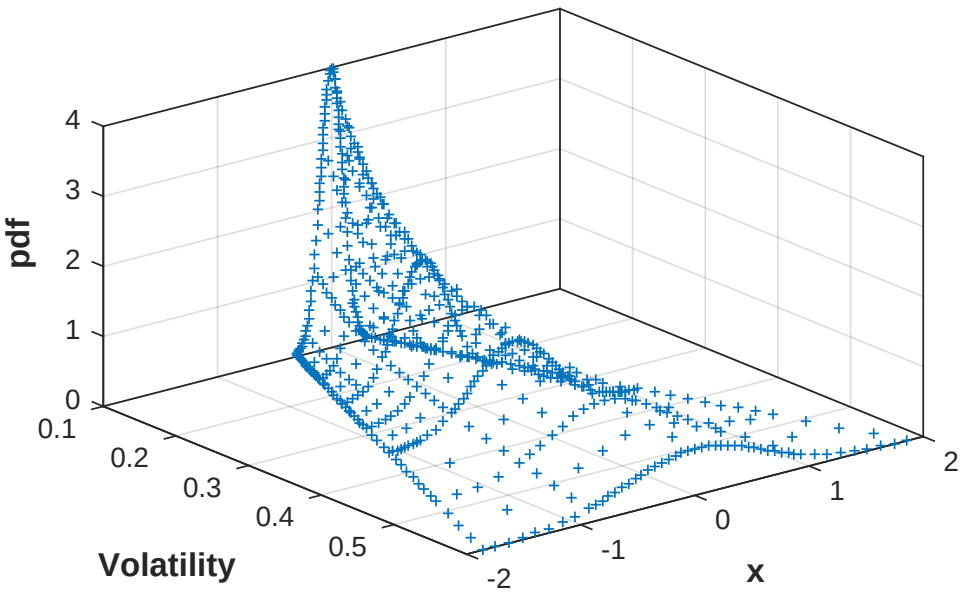


Figure 5: **Gaussian pdf approximation.** The figure shows the adaptive sparse grid approximation of the Gaussian pdf with mean parameter set to 0 and volatility parameter kept free on a domain from 0.1 to 0.6. The refinement threshold is set to $\varepsilon = 10^{-3}$ and the refinement level is capped at 16, resulting in a grid of 1369 points. We observe a concentration of points in high curvature regions and a relative smoothness in the volatility dimension.

of RASE, we do not develop this direction further.

Second, we can exploit translation invariance (or other related properties) of certain state variables as a dimension-reduction technique. Translation invariance allows us to only model the distance between origin and future states rather than their absolute position. Consider the illustrative example of a basket option under Black and Scholes framework: The state space is uniquely defined by d underlying log asset prices. Translation invariance holds for each state variable here as the probability of a change in the log price does not depend on the price level: We have $f_{\vec{x}}(\vec{y};\Delta,\Theta) = f_{\vec{0}}(\vec{y}-\vec{x};\Delta,\Theta)$ for any $\vec{x},\vec{y} \in \mathscr{D}$. We therefore build a grid for function $f(\vec{y},\Delta,\Theta) := f(\vec{0},\vec{y},\Delta,\Theta)$ setting the origin state $\vec{x}$ uniformly to zero: The size of the problem reduces to $d+q+1$, saving a total of d dimensions.

Third, in contexts of degenerate dependence structure, we decompose the problem into simpler sub-problems: Rather than building a large grid for the pdf on the entire state space, we precompute multiple simpler grids corresponding to independent partitions of the state space. Consider the same basket option example as above under the assumption that log prices are cross-sectionally uncorrelated, that is, each asset has independent Gaussian log returns. Informally, we have $f_{\vec{0}}(\vec{y};\Delta) = f_0(y_1;\Delta) \times \cdots \times f_0(y_d;\Delta)$, where the right-hand-side is the product of univariate marginal probability density functions. We can use the decomposition to focus on the construction of the grid of a unique marginal pdf.

We illustrate the precomputation method with a concrete application to Heston stochastic volatility model below in Sec. 3.

2.3.3 Modeling of discrete dividend payments

It is typical in previous literature (e.g., [8, 46, 6]) to model dividends with a continuous payment stream rather than with discrete cashflows. The reason is that discrete (in particular non-proportional) payments massively impact algorithmic performance of some pricing algorithms – e.g., lattice methods can lose their recombining property – or require ad-hoc methods for realigning the grid (see, e.g., [49]). In practice though, stocks do not distribute a continuous dividend yield, but discrete dividends: It can be optimal to exercise the options on the days just before the ex-dividend dates [70]. Modeling discrete dividends adequately is therefore crucial to avoid incorrect conclusions on the exercise boundary as well as biased prices [28].

RASE has two key advantages in the presence of dividends. First, the flexibility of time discretization scheme allows to align time grid $\mathscr{T}$ with dividend payment dates with no additional computational burden—even when dividend dates are not exactly regular. This is particularly useful when pricing vanilla American call options, which are known to be only exercised optimally at dividend dates [70]: Our algorithm can set the time grid to only consider dividend dates in the recursion, therefore saving massive computational resources.

Second, RASE keeps track of the value function on the whole domain at every time iteration using interpolation on an adaptive sparse grid. This functional approximation allows us to account for discrete dividends through a simple transformation of the state: Conditionally on cum-dividend state $\vec{x}$, we evaluate ex-dividend continuation value $\tilde{E}_i(\delta_i(\vec{x}))$, where $\delta_i(\cdot)$, $i = 0,\dots,M-1$, is a family of functions defining the impact of dividends on the state. Eq. 29 takes the following form:

$$P_i(\vec{x}) = \begin{cases} H(T,\vec{x}_T) & \text{if } i = M, \\ \max(H(t_i,\vec{x}), E_i(\delta_i(\vec{x}))) & \text{otherwise.} \end{cases} \tag{29}$$

Note that $(\delta_i(\cdot))_i$ do not need to be uniform on the whole domain and can even be state-contingent. Consider the illustrative case of a state $\vec{x}_t := s_t$ governed by the underlying asset log price s_t. The assumption of constant cash-denominated dividend payments—which amount to shifting current log price s to $s' = \log(e^s - d)$, where $d > 0$ is the cash value of the payment—is implemented with $\delta_i(s) := \log(e^s - d)$ for all i such that t_i is a dividend date, and $\delta_i(s) := s$ otherwise. Similarly, the case of proportional dividends can also be easily modeled as $\delta_i(s) = s + \log(1-d)$ where $d > 0$ is defined by the proportion of the price paid as dividend.

3 Numerical application

In this section, we benchmark the performance of RASE algorithm. We value a series of vanilla put options under one-dimensional Black and Scholes settings and two-dimensional Heston stochastic volatility settings in the presence of discrete dividends, and compare our results with well-studied numerical schemes. Second, we illustrate the scalability of our algorithm in the context of a basket option contract.

3.1 One-factor Gaussian benchmark

We first consider the valuation of American put options under the 1-factor model of Black and Scholes [8]. Although relatively straightforward with traditional numerical schemes, the framework is still an important benchmark due to its wide use and relative tractability. The dynamics of the log price of the underlying assets are given by

$$ds_t = (r_f - \frac{1}{2}\sigma^2)dt + \sigma dW_t \tag{30}$$

where $\sigma > 0$ defines the volatility of returns and $(W)_{t\geq 0}$ is a standard Brownian motion. Under those assumptions, stochastic integration of the equation reveals the normality of log returns, such that the pdf takes the form

$$f_s^{\mathrm{BS}}(s';\Delta,r_f,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2\Delta}} e^{-\frac{(s'-s-(r_f-\sigma^2/2)\Delta)^2}{2\sigma^2\Delta}}. \tag{31}$$

The payoff of the option is analytically given by

$$H(t,s) = \max(0, K - e^s) \tag{32}$$

for any $0 \leq t \leq T$, where $K > 0$ is the exercise price of the put option and s is the log price of the underlying asset.

We implement an implicit finite difference scheme (FDS) as benchmark: The approach is well-studied and is often used in previous literature as the method-of-choice in option pricing (see, e.g., [12, 27, 50]). Fig 6 illustrates the error rate of RASE when compared to our FDS implementation. We set the risk-free rate r_f to 3% and the volatility σ to 30%. Value function grids are refined up to a maximum level of $L_{\max} = 16$ and with a threshold of $\varepsilon = 10^{-3}$, and we consider an equally-spaced grid $\mathscr{T}$ with 100 time steps. The boundaries of truncated domains are set to the 10^{-4} and $1 - 10^{-4}$ quantiles of the distribution, respectively. The FDS grid uses the same time discretization as well as 200 equally-spaced points in the price dimension.

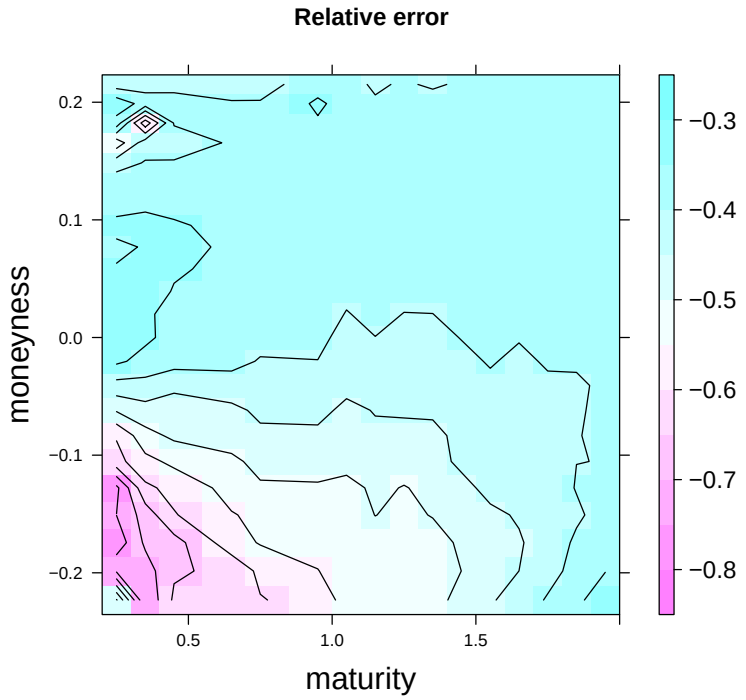


Figure 6: **American put option pricing under Black and Scholes framework.** The figure illustrates the relative error of put option prices obtain with RASE in comparison to an implicit finite difference scheme (FDS) implementation, defined as $100 \times \frac{\text{RASE price} - \text{FDS price}}{\text{FDS price}}$. Horizontal axis shows the results for maturities ranging from 0.2 year to 2 years and vertical axis shows the result for different moneyness levels given by $\log(\frac{\text{Underlying price}}{\text{Exercise price}})$. We set the risk-free rate r_f to 3% and the volatility σ to 30%. The value function grids are refined up to a maximum level of $L_{\max} = 16$ and with a criterion of $\varepsilon = 10^{-3}$, and we consider an equally-spaced time grid with 100 steps. The boundaries of truncated domains are set to the 10^{-4} and $1 - 10^{-4}$ quantiles of the distribution, respectively. The FDS grid uses the same time discretization, and 200 equally-spaced points in the price dimension.

The results confirm the good behavior of RASE: The relative error, defined as $100 \times \frac{\text{RASE price} - \text{FDS price}}{\text{FDS price}}$, reaches an (absolute) maximum of 0.9% for deep in-the-money options with short maturities and is otherwise within a 0.5% band. RASE is particularly precise for maturities larger than half a year, where the fit is excellent on the whole set of moneyness levels considered. The error for shorter maturities is mostly explained by the shape of the pdf when the time interval shrinks: The pdf becomes more and more localized and peaky as time interval gets shorter. For this reason, small errors in the computation of the continuation value have a high relative contribution and can accumulate through the repeated application of expectations during the time iteration. A simple approach for reducing the error rate further when the level is considered excessive is to allow more refinement in the construction of the grid in Eq. 27.

3.2 Two-factor stochastic volatility benchmark with discrete dividends

We consider the stochastic volatility process of Heston [46] defined by

$$ds_t = (r_f - \frac{1}{2}v_t)dt + \sqrt{v_t}dW_t \tag{33}$$

$$dv_t = \kappa(\theta - v_t)dt + \nu\sqrt{v_t}dZ_t, \tag{34}$$

where $(s_t)_{t\geq 0}$ is the log price of an asset at time t, $(v_t)_{t\geq 0}$ is its stochastic variance process, and $(W)_{t\geq 0}$ and $(Z)_{t\geq 0}$ are two correlated standard Brownian motions with $dW_t dZ_t = \rho dt$. Typically, $\theta > 0$ models the long-term mean of the variance process, $\kappa > 0$ the mean-reversion speed, $\nu > 0$ the volatility of variance, and $\rho \in [-1,1]$ the leverage effect. The pdf is not know analytically but can be uncovered by direct inversion of the characteristic function

$$\begin{aligned}
\hat{f}^{\text{heston}}_{v_0}(\xi,\psi;\Delta,\kappa,\theta,\nu,\rho) &= e^{p(\xi,\psi)+q(\xi,\psi)v_0} \qquad (35)\\
p(\xi,\psi) &= -i(\kappa\theta\rho/\nu)\Delta t\xi + (\kappa\theta/\nu^2)(\Delta t\kappa + \Delta t\sqrt{A} + 2\log(\frac{B+1}{Be^{\Delta t\sqrt{A}}+1}))\\
q(\xi,\psi) &= (1/\nu^2)(\kappa - i\rho\nu\xi - \sqrt{A}\frac{Be^{\Delta t\sqrt{A}}-1}{Be^{\Delta t\sqrt{A}}+1})\\
A(\xi,\psi) &= \nu^2(1-\rho^2)\xi^2 + \kappa^2 + i(\nu^2 - 2\rho\nu\kappa)\xi\\
B(\xi,\psi) &= -\frac{i\rho\nu\xi + i\nu^2\psi - \kappa - \sqrt{A}}{i\rho\nu\xi + i\nu^2\psi - \kappa + \sqrt{A}},
\end{aligned}$$

so that

$$f^{\text{heston}}_{v_0}(r,v;\Delta,\kappa,\theta,\nu,\rho) := \frac{1}{(2\pi)^2}\int\int e^{-i(\xi r+\psi v)}\hat{f}^{\text{heston}}_{v_0}(\xi,\psi;\Delta,\kappa,\theta,\nu,\rho)d\xi d\psi. \tag{36}$$

We use translation invariance to account for the impact of risk-free interest rate on the mean value of the pdf: For the initial state process $\vec{x}_t = (s_t, v_t)$ and the final state $\vec{x}_{t+\Delta} = (s_{t+\Delta}, v_{t+\Delta})$, the pdf takes value $f^{\text{heston}}_{v_t}(s_{t+\Delta} - (s_t + r_f\Delta), v_{t+\Delta};\Delta,\kappa,\theta,\nu,\rho)$ for any $t \geq 0$. The Heston pdf is therefore an 8-dimensional object.

In order to determine adequate boundaries for truncated domains $\mathscr{D}^x$ and $\mathscr{D}^y_{\vec{x},\Delta}$ for a given state $\vec{x} = (s,v)$, we need approximate quantiles of $f^{\text{heston}}_{v_0}(\cdot,\cdot;\Delta,\kappa,\theta,\nu,\rho)$. The distribution of the scaled variance process $v_{t+\Delta}/2c$ given spot variance v_t, where $c = \frac{2\kappa}{(1-e^{-\kappa\Delta})\nu^2}$, follows a non-central chi-squared distribution with $4\kappa\theta/\nu^2$ degrees of freedom and a non-centrality parameter $2cv_te^{-\kappa\Delta}$. As the cumulative probability function is only known semi-analytically, we invert the closed-form approximation of [73] to proxy the quantile function $q^\alpha_{v_0} := q^\alpha_{v_0}(\Delta,\kappa,\theta,\nu)$ of the variance margin, where $\alpha \in [0,1]$ is the probability level. In the following, we use $\mathscr{D}^y_{\vec{x},\Delta} := [s - 3.89\sqrt{q_v^{1-10^{-4}}}, s + 3.89\sqrt{q_v^{1-10^{-4}}}] \times [q_v^{10^{-4}}, q_v^{1-10^{-4}}]$

and $\mathscr{D}^x := \mathscr{D}^y_{\vec{x}_0,\Delta}$, where 3.89 is the rounded 10^{-4}-quantile of the Normal distribution.

The most common algorithm for inverting a characteristic function is the fast Fourier transform, which efficiently computes a full section of discrete Fourier transforms on a regular grid. The coordinate system of RASE being irregular, we rather apply a traditional discrete transform for each point dynamically added to the pdf grid. The theoretical loss in algorithmic efficiency is compensated by the low number of points of the adaptive sparse grid.[14] We precompute the entire functional on an adaptive sparse grid as presented in Sec. 2.3.2. The grid has a maximum level of $L_{\max} = 16$. We use the following custom criterion: $\frac{\alpha_{\vec{l},\vec{j}}}{\max f_{v_0}(\cdot,\cdot;\Delta,\kappa,\theta,\nu,\rho)} \geq \varepsilon$, where $\alpha_{\vec{l},\vec{j}}$ is a surplus, and the maximum is taken over known points of the grid conditionally on variance level and model parameters. We automatically refine if the grid is empty or maximum value is zero. This criterion provides a simple way to control the relative error irrespective of current state and parameters. An absolute refinement criterion as in Sec. 2.2 would refine excessively when the pdf takes large values (e.g., when variance level is low or time interval is short) and insufficiently when the pdf takes small values. The final pdf grid contains 815398 points.

We consider the parameters values $r_f = 0.03$, $\kappa = 2$, $\theta = 0.09$, $\nu = 0.3$, and $\rho = -0.7$, which are in line with typical estimates of Heston model on US equity indices (see, e.g., [3]). We assume quarterly discrete dividend payments of 1.25—with the first payment after exactly three months, and the dividend size capped by the value of the stock (to avoid negative stock price). Tab. 1 reports the value of a cross-section of Bermudan put option with monthly exercise, with respective maturities of 1 year and 6 months and initial volatility regimes 25%, 30% and 40%. We benchmark the results of RASE algorithm with $100,000$ Monte Carlo simulations (including half antithetic variates) of the Euler discretization of Eq. 34 with 600 time increments.[15] We apply the algorithm of Longstaff and Schwartz [59] to account for early exercise: We use scaled Laguerre polynomials for the regression estimate of the continuation value, with polynomials up to order 6 in the price dimension, order 4 in the variance dimension, and cross products up to order 5. As their simulation scheme is well-known to select a suboptimal exercise policy, we expect the results of the simulation to be subject to a slight negative bias.

Tab. 1 reports the prices obtain with the two numerical methods as well as the the relative difference, computed as $\frac{\text{RASE price}-\text{MC-LS price}}{\text{MC-LS price}}$. For in- and at-the-money options, RASE has maximum absolute relative difference of -0.7% with respect to Monte Carlo simulations on the grid considered. RASE prices lie withing a two standard deviations interval of Monte Carlo estimates. Out-of-the-money, we observe a larger difference: Monte Carlo simulations undervalue the option price in most scenario with respect to RASE. The relative difference reaches a maximum of 1.6% in the deep out-of-the-money case and the difference is statistically significant in multiple occasions. As a robustness check, we increase

[14] We refer to [76] for an extensive review of Fourier calculus and numerical implementations of the Fourier transform.

[15] Simulation of the discretized version of Eq. 34 can incorrectly generate paths with negative variances—and therefore complex values. We use the full truncation scheme as described in [60] to mitigate this issue.

Volatility		Exercise price				
		85	95	100	105	115
25%	RASE	4.95	8.80	11.17	13.86	20.25
	MC-LS	4.89	8.67	11.10	13.85	20.36
	MC std	0.03	0.04	0.05	0.05	0.06
	Rel. diff.	1.2%	1.5%	0.6%	$< 0.1\%$	-0.5%
30%	RASE	5.62	9.54	11.96	14.65	21.00
	MC-LS	5.53	9.43	11.90	14.67	21.09
	MC std	0.03	0.05	0.05	0.05	0.06
	Rel. diff.	1.6%	1.1%	0.5%	-0.1%	-0.4%
40%	RASE	7.00	11.10	13.70	16.40	22.73
	MC-LS	6.98	11.16	13.67	16.50	22.89
	MC std	0.04	0.05	0.05	0.06	0.07
	Rel. diff.	0.3%	-0.5%	0.2%	-0.6%	-0.7%

Table 1: **Bermudan put option pricing under stochastic volatility model with discrete dividends.** The table shows the prices of Bermudan put options obtained with RASE under Heston stochastic volatility model parametrized with $r_f = 0.03$, $\kappa = 2$, $\theta = 0.09$, $\nu = 0.3$, and $\rho = -0.7$. The maturity is set to $T = 1$ year and the option is given monthly exercise rights at a price 100. The underlying asset pays quarterly dividends of 1.25. The maximum refinement level of the value function is $L_{\max} = 15$ and the refinement threshold is $\varepsilon = 10^{-3}$. *MC-LS prices* are given by $100,000$ Monte Carlo simulations with antithetic variates and the algorithm of Longstaff and Schwartz [59]. *MC std* indicates the standard deviation of the statistic. *Rel. diff.* is the relative difference computed as $\frac{\text{RASE price} - \text{MC-LS price}}{\text{MC-LS price}}$.

the maximum level of the grid to $L_{\max} = 16$, slightly decrease the refinement threshold to $\varepsilon = 5 \times 10^{-4}$ and use second-order basis functions (see, e.g., [18]) with qualitatively similar results. The identified difference is most probably explained by the well-known bias of Longstaff and Schwartz algorithm, which selects imperfect exercise strategies.

3.3 Scalability in higher dimensions

Finally, we illustrate the scalability of RASE beyond two dimensions with an application to the pricing of basket options. We extend the Gaussian framework of Sec. 3.1 to d-dimensional settings: We consider the canonical case of i.i.d. assets, each following Black and Scholes dynamics of Eq. 30:

$$ds_t^i = (r_f - \frac{1}{2}\sigma^2)dt + \sigma dW_t^i \tag{37}$$

for $i = 1, \ldots, d$. We focus on European basket options to isolate the complexity scaling of one specific time iteration. The final payoff is given by

$$H(T, \vec{s}) = \max(0, K - \frac{1}{d}\sum_{i=1}^{d} \exp(s^i)), \tag{38}$$

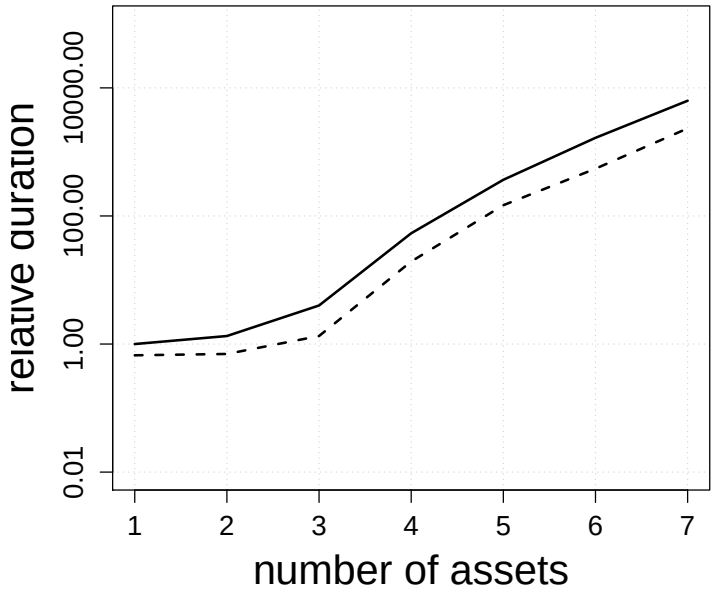


Figure 7: **Performance of RASE in the pricing of basket options.** The figure illustrates time-to-solution of RASE in the context of basket options pricing relative to a single-underlying option. Horizontal axis shows the number of underlying assets and vertical axis shows the relative time on a logarithmic scale. Solid line indicates the relative time using a single computation unit, dashed line indicates the relative time using four units.

where $K > 0$ is the exercise price and $\vec{s} = (s^1, \ldots, s^d)$ is a vector of underlying assets log prices. We set the risk-free rate r_f to 3% and the volatility σ to 30%. Value function grids are refined up to a maximum level of $L_{\max} = 16$ and with a threshold of $\varepsilon = 10^{-3}$. We truncate the domains of each respective dimension at 10^{-4} and $1 - 10^{-4}$ quantiles of the Gaussian distribution.

Fig. 7 shows the time-to-solution of RASE algorithm for increasingly large portfolios of $d = 1$ to $d = 7$ underlying assets, relative to the execution time of the single-asset case. Horizontal axis shows the number of underlying assets and vertical axis shows the relative time on a logarithmic scale. We additionally show the relative time-to-solution when allocating four processing units to the pricing problem in order to demonstrate the intrinsic degree of parallelism of our algorithm. As explained in previous sections, sparse grids can drastically improve on the complexity of multidimensional problems as compared to traditional, dense grids: They reduce the number of grid points needed from the order $\mathcal{O}\left(N^d\right)$ to $\mathcal{O}\left(N \cdot (\log N)^{d-1}\right)$. Adaptivity saves even more points by only refining regions of the domain where the function has significant local features. Fig. 7 validates the complexity reduction provided by adaptive sparse grids: The time-to-solution grows with an approximate factor of 4 per asset added to the portfolio, which is a massive improvement over dense grids. The basket option pricing problem can be run on a traditional desktop computer up to six or seven dimensions. The figure also shows the high degree of parallelism inherent to the algorithm: The construction of the grid can be split across many different

computation units with almost no overhead. The pricing of basket options is sped up by a factor of 2.5 when run on a four-core desktop computer. The execution of RASE on a computer cluster therefore allows to trivially scale up beyond seven dimensions.

4 Conclusion

We develop a novel scheme for pricing American derivatives under multi-factor models. It scales well to higher dimensions and has massive degree of parallelism thanks to the use of adaptive sparse grids. The algorithm can deal with discrete dividends and irregular time grids. We illustrate its performance in the context of option pricing under Black and Schole settings and Heston stochastic volatility settings. We consider the case of basket options to illustrate the scaling and parallelism. RASE brings a promising alternative to Monte Carlo simulations when low dimension models are not satisfying, and opens doors for further improving research in derivatives pricing.

References

[1] Yacine Aït-Sahalia and Jean Jacod. Telling from discrete data whether the underlying continuous-time model is a diffusion. *The Journal of Finance*, 57(5):2075–2112, oct 2002.

[2] Pierre Bajgrowicz, Olivier Scaillet, and Adrien Treccani. Jumps in High-Frequency Data: Spurious Detections, Dynamics, and News. *Management Science*, 2015.

[3] Gurdip Bakshi, Charles Cao, and Zhiwu Chen. Empirical Performance of Alternative Option Pricing Models. *The Journal of Finance*, 52(5):2003–2049, 1997.

[4] Giovanni Barone-Adesi and Robert E Whaley. Efficient Analytic Approximation of American Option Values. *The Journal of Finance*, 42(2):301–320, jun 1987.

[5] Suleyman Basak and Georgy Chabakauri. Dynamic hedging in incomplete markets: A simple solution. *Review of Financial Studies*, 25(6):1845–1896, 2012.

[6] David S. Bates. Jumps and stochastic volatility: exchange rate processes implicit in deutsche mark options. *Review of Financial Studies*, 9(1):69–107, 1996.

[7] R Bellman and R E Bellman. *Adaptive Control Processes: A Guided Tour*. 'Rand Corporation. Research studies. Princeton University Press, 1961.

[8] Fischer Black and Myron Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637, 1973.

[9] Olivier Bokanowski, Jochen Garcke, Michael Griebel, and Irene Klompmaker. An Adaptive Sparse Grid Semi-Lagrangian Scheme for First Order Hamilton-Jacobi Bellman Equations. *Journal of Scientific Computing*, 55(3):575–605, 2013.

[10] Douglas T. Breeden and Robert H. Litzenberger. Prices of State-Contingent Claims Implicit in Option Prices, 1978.

[11] M J Brennan. The Pricing Of Contingent Claims In Discrete Time Models. *Journal of Finance*, 34(1):53–68, 1979.

[12] M J Brennan and E S Schwartz. Convertible Bonds: Valuation And Optimal Strategies For Call And Conversion. *Journal of Finance*, 32(5):1699–1715, 1977.

[13] Mark Broadie and Jerome Detemple. American Option Valuation: New Bounds, Approximations, and a Comparison of Existing Methods. *The Review of Financial Studies*, 9(4):1211–1250, 1996.

[14] Mark Broadie and Paul Glasserman. Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control*, 21:1323–1352, 1997.

[15] Johannes Brumm, Dmitry Mikushin, Simon Scheidegger, and Olaf Schenk. Scalable high-dimensional dynamic stochastic economic modeling. *Journal of Computational Science*, 11:12–25, 2015.

[16] Johannes Brumm and Simon Scheidegger. Using Adaptive Sparse Grids to Solve High-Dimensional Dynamic Models. *Available at SSRN 2349281*, 2013.

[17] David S Bunch and Herb Johnson. The American Put Option and Its Critical Stock Price. *The Journal of Finance*, 55(5):2333–2356, oct 2000.

[18] H.-J. Bungartz and S Dirnstorfer. Multivariate Quadrature on Adaptive Sparse Grids. *Computing*, 71(1):89–114, 2003.

[19] Hans-Joachim Bungartz and Michael Griebel. Sparse grids. *Acta Numerica*, 13:1–123, 2004.

[20] Peter Carr, Hélyette Geman, Dilip B Madan, and Marc Yor. Stochastic volatility for Lévy processes. *Mathematical Finance*, 13(3):345–382, 2003.

[21] Peter Carr and Liuren Wu. What Type of Process Underlies Options? A Simple Robust Test. *Journal of Finance*, 58:2581–2610, 2003.

[22] C.˜Hager, S.˜Hüeber, and B.˜Wohlmuth. Numerical techniques for the valuation of basket options and its Greeks. *J. Comput. Fin.*, 13(4):1–31, 2010.

[23] Peng Cheng and Olivier Scaillet. Linear-quadratic jump-diffusion modeling. *Mathematical Finance*, 17(4):575–598, 2007.

[24] Carl Chiarella, Nadima El-Hassan, and Adam Kucera. Evaluation of American option prices in a path integral framework using Fourier–Hermite series expansions. *Journal of Economic Dynamics and Control*, 23(9):1387–1424, 1999.

[25] Carl Chiarella and Andrew Ziogas. Pricing American options on jump-diffusion processes using Fourier Hermite series expansions. *Quantitative Finance Research Centre Research Paper*, (145), 2005.

[26] Kim Christensen, Roel Oomen, and Mark Podolskij. Fact or friction: jumps at ultra high frequency. *Journal of Financial Economics*, (forthcoming), 2014.

[27] Nigel Clarke and Kevin Parrott. Multigrid for American option pricing with stochastic volatility. *Applied Mathematical Finance*, 6(3):177–195, 1999.

[28] Antonio Cosma, Stefano Galluccio, Paola Pederzoli, and Olivier Scaillet. Valuing American options using fast recursive projections. 2015.

[29] John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. An Intertemporal General Equilibrium Model of Asset Prices. *Econometrica*, 53(2):363–384, 1985.

[30] John C. Cox, Stephen a. Ross, and Mark Rubinstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229–263, 1979.

[31] Jaksa Cvitanic and Fernando Zapatero. *Introduction to the Economics and Mathematics of Financial Markets*, volume 1. 2004.

[32] Serge Darolles and Jean-Paul Laurent. Approximating payoffs and pricing formulas. *Journal of Economic Dynamics and Control*, 24(11):1721–1746, 2000.

[33] F. Delbaen and W. Schachermayer. A General Version of the Fundamental Theorem of Asset Pricing. *Math. Annalen*, 300:463–520, 1994.

[34] Emanuel Derman, Iraj Kani, and Joseph Z. Zou. The Local Volatility Surface. *Goldman Sachs Quantitative Strategies Research Notes*, (December), 1995.

[35] V. V. Desai, V. F. Farias, and C. C. Moallemi. Pathwise Optimization for Optimal Stopping Problems. *Management Science*, 1909(1996):1–17, 2012.

[36] Jérôme Detemple. *American-style derivatives: Valuation and computation*. Financial mathematics series. Chapman & Hall/CRC Press, Boca Raton, Florida, USA, 2005.

[37] Jérôme B. Detemple and Weidong Tian. The Valuation of American Options for a Class of Diffusion Processes. *Management Science*, 48(7):917–937, 2002.

[38] Darrell Duffie, Jun Pan, and Kenneth Singleton. Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica*, 68(6):1343–1376, nov 2000.

[39] J Garcke and M Griebel. *Sparse Grids and Applications*. Lecture Notes in Computational Science and Engineering Series. Springer-Verlag GmbH, 2012.

[40] Thomas Gerstner and Michael Griebel. Numerical integration using sparse grids. *Numerical Algorithms*, 18(3-4):209–232, 1998.

[41] Paul Glasserman. Monte Carlo methods in finanical engineering, 2004.

[42] J Michael Harrison and David M Kreps. Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations. *The Quarterly Journal of Economics*, 92(2):323–336, 1978.

[43] J.Michael Harrison and Stanley R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215–260, 1981.

[44] Martin B. Haugh and Leonid Kogan. Pricing American Options: A Duality Approach. *Operations Research*, 52(2):258–270, 2004.

[45] Alexander Heinecke and Dirk Pflüger. Emerging Architectures Enable to Boost Massively Parallel Data Mining Using Adaptive Sparse Grids. *International Journal of Parallel Programming*, 41(3):357–399, 2013.

[46] 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.

[47] Markus Holtz. *Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance*. Lecture Notes in Computational Science and Engineering. Springer, Dordrecht, 2011.

[48] Jing-zhi Huang, Marti G. Subrahmanyam, and G. George Yu. Pricing and Hedging American Options: A Recursive Integration Method. *Review of Financial Studies*, 9(1):277–300, 1996.

[49] John Hull. *Options, Futures and Other Derivatives*. Pearson Education, 8 edition, apr 2011.

[50] Samuli Ikonen and Jari Toivanen. Efficient numerical methods for pricing American options under stochastic volatility. *Numerical Methods for Partial Differential Equations*, 24(1):104–126, 2008.

[51] Mads Vestergaard Jensen and Lasse Heje Pedersen. Early Option Exercise: Never Say Never. 2015.

[52] Nengjiu Ju. Pricing an American Option by Approximating Its Early Exercise Boundary as a Multipiece Exponential Function. *The Review of Financial Studies*, 11(3):627–646, 1998.

[53] Kenneth L Judd. *Numerical methods in economics*. The MIT press, 1998.

[54] In N. Kim. The analytic valuation of American options. *Review of Financial Studies*, 3(4):547–572, 1990.

[55] Vincent Lacoste. Wiener chaos: a new approach to option hedging. *Mathematical Finance*, 6(2):197–213, 1996.

[56] Suzanne Lee and Per Mykland. Jumps in equilibrium prices and market microstructure noise. *Journal of Econometrics*, 168(1):396–406, 2012.

[57] Markus Leippold and Liuren Wu. Asset Pricing under the Quadratic Class. *The Journal of Financial and Quantitative Analysis*, 37(2):271–295, 2002.

[58] L Ljungqvist and T J Sargent. *Recursive macroeconomic theory*. Mit Press, 2000.

[59] F. a. Longstaff and Schwartz E.S. Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies*, 14:113–147, 2001.

[60] Roger Lord, Remmert Koekkoek, and Dick Van Dijk. A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2):177–194, 2010.

[61] Xiang Ma and Nicholas Zabaras. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *J. Comput. Phys.*, 228(8):3084–3113, may 2009.

[62] Dilip B Madan, Peter P Carr, and Eric C Chang. The variance gamma process and option pricing. *European finance review*, 2(1):79–105, 1998.

[63] Dilip B Madan and Frank Milne. Contingent claims valued and hedged by pricing and investing in a basis. *Mathematical Finance*, 4(3):223–245, 1994.

[64] Alin Murarasu, Josef Weidendorfer, Gerrit Buse, Daniel Butnaru, and Dirk Pflüeger. Compact Data Structure and Parallel Alogrithms for the Sparse Grid Technique. In *16th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, 2011.

[65] Erich Novak and Klaus Ritter. High dimensional integration of smooth functions over cubes. *Numerische Mathematik*, 75(1):79–97, 1996.

[66] Edward W. Packel and J. F. Traub. Information-based complexity. *Nature*, 328(6125):29–33, 1987.

[67] Benjamin Peherstorfer, Pablo Gómez, and Hans-Joachim Bungartz. Reduced models for sparse grid discretizations of the multi-asset Black-Scholes equation. *Advances in Computational Mathematics*, 41(5):1365–1389, 2015.

[68] Dirk Pflüger. Spatially Adaptive Refinement. In Jochen Garcke and Michael Griebel, editors, *Sparse Grids and Applications*, Lecture Notes in Computational Science and Engineering, pages 243–262, Berlin Heidelberg, oct 2012. Springer.

[69] Dirk Pflüger, Benjamin Peherstorfer, and Hans-Joachim Bungartz. Spatially adaptive sparse grids for high-dimensional data-driven problems. *Journal of Complexity*, 26(5):508–522, oct 2010.

[70] Veronika Krepely Pool, Hans R Stoll, and Robert E Whaley. Failure to exercise call options: An anomaly and a trading game. *Journal of Financial Markets*, 11(1):1–35, 2008.

[71] L C G Rogers. Monte Carlo valuation of American options. *Mathematical Finance*, 12(3):271–286, 2002.

[72] Mark Rubinstein. The Valuation of Uncertain Income Streams and the Pricing of Options. *The Bell Journal of Economics*, 7(2):407–425, 1976.

[73] M. Sankaran. Approximations to the non-central chi-squared distribution. *Biometrika*, 50(1-2):199–204, 1963.

[74] S A Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Math. Dokl.*, 4:240–243, 1963.

[75] M A Sullivan. Valuing American put options using Gaussian quadrature. *Review of Financial Studies*, 13(1):75–94, 2000.

[76] Lance A Waller, Bruce W Turnbull, and J Michael Hardin. Obtaining Distribution Functions by Numerical Inversion of Characteristic Functions With Applications. *The American Statistician*, 49(4):346–350, 1995.

[77] Viktor Winschel and Markus Kraetzig. Solving, Estimating, and Selecting Nonlinear Dynamic Models Without the Curse of Dimensionality. *Econometrica*, 78(2):803–821, 2010.

[78] Christoph Zenger. Sparse Grids. In Wolfgang Hackbusch, editor, *Parallel Algorithms for Partial Differential Equations*, volume 31 of *Notes on Numerical Fluid Mechanics*, pages 241–251. Vieweg, 1991.