SIAM J. SCI. COMPUT.
Vol. 29, No. 1, pp. 440–458

EFFICIENT HIERARCHICAL APPROXIMATION OF HIGH-DIMENSIONAL OPTION PRICING PROBLEMS*

CHRISTOPH REISINGER† AND GABRIEL WITTUM‡

Abstract. A major challenge in computational finance is the pricing of options that depend on a large number of risk factors. Prominent examples are basket or index options where dozens or even hundreds of stocks constitute the underlying asset and determine the dimensionality of the corresponding degenerate parabolic equation. The objective of this article is to show how an efficient discretization can be achieved by hierarchical approximation as well as asymptotic expansions of the underlying continuous problem. The relation to a number of state-of-the-art methods is highlighted.

Key words. sparse grids, multigrid methods, option pricing, asymptotic expansions, dimension reduction

AMS subject classifications. 65N06, 65N12, 65N15, 65N40

DOI. 10.1137/060649616

1. Introduction.

Financial options are an essential element of today's risk management. They give the holder the right (but not the obligation) to perform a specific transaction on a risky asset X (an equity, for instance) or a number of assets $\mathbf{X} = (X_1, \dots, X_n)^T$ at some point(s) or time span in the future for a price that is agreed upon now, the *strike* price K.

This work focuses on the cases where either the asset can only be exercised at the expiry T (*European* option), which means the holder may then buy (*call*) or sell (*put*) the asset for the strike K, or the asset can be exercised at any time up to T (*American* option).

For a put on a claim g the price u of the option at expiry (the *pay-off*) is clearly

$$u(\mathbf{X}, T) = (K - g(\mathbf{X}))_+ = \max(K - g(\mathbf{X}), 0),$$

and analogously for the call. At any point up to then the value of the option depends decisively on the stochastic nature of the process.

1.1. Model framework.

The most common models for the underlying assets fall into the class of *Itô processes*

$$\mathrm{d}X_i = \beta_i^X(\mathbf{X}, \mathbf{Y})\,\mathrm{d}t + \alpha_i^X(\mathbf{X}, \mathbf{Y})\,\mathrm{d}W_i^X, \tag{1}$$

with correlated *Wiener processes* W_i^X, where $\mathbf{Y} = (Y_1, \dots, Y_m)^T$ is a set of parameters. Traditionally market parameters have been assumed constant, but time series data in many markets reveal that some characteristics can be captured appropriately only if they are themselves modeled as stochastic differential equations (SDEs):

$$\mathrm{d}Y_i = \beta_i^Y(\mathbf{X}, \mathbf{Y})\,\mathrm{d}t + \alpha_i^Y(\mathbf{X}, \mathbf{Y})\,\mathrm{d}W_i^Y.$$

*Received by the editors January 11, 2006; accepted for publication (in revised form) September 13, 2006; published electronically February 26, 2007.
http://www.siam.org/journals/sisc/29-1/64961.html

†Mathematical Institute, Oxford University, 24–29 St. Giles, Oxford, OX1 3LB, UK (christoph.reisinger@maths.ox.ac.uk).

‡Simulation in Technology Center, Heidelberg University, Im Neuenheimer Feld 368, 69120 Heidelberg, Germany (wittum@uni-hd.de).

Let us write $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}^X, \boldsymbol{\beta}^Y)$, $\mathbf{W} = (\mathbf{W}^X, \mathbf{W}^Y)$ and denote by ρ_{ij}, $1 \le i, j \le d := n + m$, the correlation of these Wiener processes.

In many cases assets and parameters are bounded by positivity constraints (equities, exchange rates, etc.), and this can be guaranteed by conditions of the form

$$\boldsymbol{\beta} \cdot \mathbf{n} \le 0, \qquad \boldsymbol{\alpha} \cdot \mathbf{n} = 0 \tag{2}$$

on the boundary Γ of the positive quadrant $\mathbb{R}^d_+$ with outward normal $\mathbf{n}$ and sufficiently smooth coefficients ($\boldsymbol{\alpha} \cdot \mathbf{n}$ must be of order at least $1/2$ at the boundary).

If $m = 0$, that is, parameters are deterministic, the pay-off can be replicated by a dynamic trading strategy with the assets X_i and a money market account with risk-free instantaneous interest rate r (assumed constant or more generally modeled by one of the parameter processes). This has the consequence that the price of an option can be seen as the expectation of the pay-off under a unique so-called *risk-neutral* measure, which turns the stock price process (1), discounted by the risk-free interest rate r, into a martingale. For details of this, see, for example, [7] or [16]. The corresponding Feynman–Kac PDE for the option price reads

$$\frac{\partial u}{\partial t} + \sum_{i,j=1}^{d} a_{ij}(\mathbf{Z}, t) \frac{\partial^2 u}{\partial Z_i \partial Z_j} + \sum_{i=1}^{d} b_i(\mathbf{Z}, t) \frac{\partial u}{\partial Z_i} + c(\mathbf{Z}, t) u = 0, \tag{3}$$

which is a linear advection-diffusion-reaction equation with coefficients

$$\begin{aligned} a_{ij}(\mathbf{Z}, t) &= \rho_{ij} \alpha_i(\mathbf{Z}, t) \alpha_j(\mathbf{Z}, t), \quad 1 \le i, j \le d, \\ c(\mathbf{Z}, t) &= -r. \end{aligned}$$

The advective term $\mathbf{b}(\mathbf{Z}, t)$ depends on the product-specific hedge that creates a risk-free investment (the change of measure). Examples will be given in the next section. Condition (2) on the boundary translates to

$$\mathbf{a}(\mathbf{Z}, t) \cdot \mathbf{n}(\mathbf{Z}) = 0 \quad \text{and} \quad \mathbf{b}(\mathbf{Z}, t) \cdot \mathbf{n}(\mathbf{Z}) \ge 0 \quad \text{for } \mathbf{Z} \in \Gamma \tag{4}$$

and guarantees uniqueness of the solution without boundary conditions [24]. Equation (3) is an antidiffusion equation that is solved backwards in time and is thus parabolic and well-posed.

Equations similar to (3) can be derived in the presence of stochastic parameters ($m \ne 0$), which are often not traded in the market themselves (like volatilities) and sometimes cannot be directly observed. To hedge the additional risk introduced by the stochastic components of $\mathbf{Y}$, options with different expiration dates have to be traded dynamically, and an ad hoc assumption on the market price of risk has to be made from practical considerations.

It has to be remarked that in practice it is often not feasible to trade options dynamically due to transaction costs (for pricing under transaction costs, see, e.g., [14] and the references therein). In the case of static positions in the option a perfect hedge is no longer possible, but instead the expected utility can be optimized and the option price is adjusted such that the buyer/seller is indifferent to such a deal. This gives rise to (nonlinear) Hamilton–Jacobi–Bellmann equations [6]. In the scope of this article we will restrict ourselves to the linear setting.

1.2. Examples.

1.2.1. FX option. We consider an option on the exchange rate X between a domestic and a foreign currency, which is driven by the short rates r_d (*domestic*) and r_f (*foreign*) in both markets according to

$$\mathrm{d}X = (r_d - r_f)X\,\mathrm{d}t + \sigma_X X\,\mathrm{d}W$$

with volatility σ_X. In practice typically a time- and state-dependent volatility surface is calibrated to quoted option prices with different maturities and strikes. For short maturities (typically under 3 years) interest rates may be assumed constant in good approximation and the following equation for u is obtained by standard arguments:

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sigma_X^2 X^2 \frac{\partial^2 u}{\partial X^2} + (r_d - r_f)X\frac{\partial u}{\partial X} - r_d u = 0. \tag{5}$$

In the more interesting case of longer time horizons the spot rate dynamics are relevant. Commonly used models are of the type

$$\mathrm{d}r_i = \mu_i(r_i, t)\,\mathrm{d}t + \nu_i(r_i, t)\,\mathrm{d}W_i,$$

where $(i = d, f)$

$$\mu_i(r_i, t) = \kappa_i(\theta_i(t) - r_i), \tag{6}$$

$$\nu_i(r_i, t) = r_i^{\beta_i}\sigma_i(t) \tag{7}$$

with positive parameters κ_i, β_i and a time-dependent reversion level θ_i, which can be calibrated to the term structure of interest rates. The corresponding PDE reads $(\mathbf{Z} = (X, r_d, r_f))$

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sum_{i,j=1}^{3}\nu_i\nu_j\rho_{ij}\frac{\partial^2 u}{\partial Z_i \partial Z_j} + \sum_{i=1}^{3}\mu_i\frac{\partial u}{\partial Z_i} + \rho_{31}\nu_3\nu_1\frac{\partial u}{\partial Z_3} - Z_2 u = 0, \tag{8}$$

where $\nu_1 = \sigma_X Z_1$ and ρ_{ij}, $1 \le i, j \le 3$, are again the correlations of the Wiener processes.

For $\beta_i \ge 1/2$ the interest rates stay nonnegative and the PDE is solved backwards on $\mathbb{R}^3_+$ from the terminal condition (put)

$$u(\mathbf{Z}, T) = (K - Z_1)_+ .$$

1.2.2. Equity basket. A classical high-dimensional application is a basket (linear combination) of d stocks, assumed to follow a geometric Brownian motion

$$\mathrm{d}S_i = \mu_i S_i\,\mathrm{d}t + \sigma_i S_i\,\mathrm{d}W_i,$$

as in [1], with Brownian motions W_i with correlation $\langle \mathrm{d}W_i, \mathrm{d}W_j\rangle = \rho_{ij}$. Denote by r again the constant risk-free interest rate. The *arbitrage-free* price (the price that does not allow any instantaneous risk-free return) of a European option follows the *Black–Scholes equation*

$$\frac{\partial u}{\partial t} = \mathcal{L}_{BS}u := -\frac{1}{2}\sum_{i,j=1}^{d}\sigma_i\sigma_j\rho_{ij}S_iS_j\frac{\partial^2 u}{\partial S_i \partial S_j} - r\sum_{i=1}^{d}S_i\frac{\partial u}{\partial S_i} + ru. \tag{9}$$

A put on the basket $\sum_i \mu_i S_i$ with positive weights μ_i has terminal value

$$u(\mathbf{S}, T) = g(\mathbf{S}) := \Big(K - \sum_{i=1}^{d} \mu_i S_i\Big)_+, \qquad \mathbf{S} \in \mathbb{R}^d_+.$$

American options, which can be exercised at any time up to T, solve the linear complementarity problem

$$\frac{\partial u}{\partial t} + \mathcal{L}_{BS} u \le 0, \tag{10}$$

$$u \ge g, \tag{11}$$

$$\left(\frac{\partial u}{\partial t} + \mathcal{L}_{BS} u\right) \cdot (u - g) = 0 \tag{12}$$

(see, e.g., [22]). A free boundary separates the exercise region and the continuation region of the option.

1.3. Curse of dimensionality, Monte Carlo, and sparse grids. The dimension of (3) can be very large, depending on the number d of stochastic components considered. The complexity of a standard grid-based approach to compute the option price with a desired accuracy ε is

$$W(\varepsilon, d) = \mathcal{O}(\varepsilon^{-d/p})$$

for a discretization of convergence order p and grows exponentially with the dimension d of the problem. This is referred to as the *curse of dimensionality*.

As a result of this the vast majority of problems in financial practice—and virtually all with dimension larger than three—are solved by Monte Carlo or related approaches [10]. The practical drawbacks of these methods, in addition to the somewhat disturbing fact that error bounds can only be specified in a probabilistic sense, are the following:

1. The convergence of the method, although it depends only very weakly on the dimension of the problem, is very slow. Typically if N is the number of simulated paths,
$$\varepsilon = \mathcal{O}(N^{-\frac{1}{2}}), \tag{13}$$
where ε has to be read as the standard deviation of the result.
2. Standard Monte Carlo methods have difficulties yielding accurate sensitivities of the solution with respect to the underlying stocks, which are important trading parameters (and directly given by the partial derivatives of the PDE solution). This is particularly the case if the option exhibits *knock-out* characteristics, that is, the option ceases to exist if the stock crosses a certain *barrier*. Note that such features fit very naturally in the PDE context where only the boundary conditions need to be changed.
3. Early exercise features like those in the American case do not fit into the "forward" simulation paradigm, and the determination of the free boundary by regression of the value function [13] is very costly, whereas the modification of the PDE to a linear complementarity problem is usually straightforward.

In contrast, quasi-Monte Carlo methods are based on (deterministic) low-discrepancy point sequences and thus restricted to integration problems. They exhibit the same disadvantages 2 and 3 above, but improve the asymptotic complexity (13) to

$$\varepsilon = \mathcal{O}\left((\log N)^{d-1} N^{-1}\right)$$

with N points of the sequence, requiring higher smoothness (bounded variation) of the integrand [15].

Lattice rules exploit even higher regularity and yield error bounds

$$\varepsilon = \mathcal{O}\left((\log N)^{d-1} N^{-r}\right)$$

for integrands with r derivatives [15].

A further progression in this direction is the Smolyak construction [20] of quadrature rules in spaces with bounded mixed derivatives, which was more recently studied numerically in the context of financial derivatives [8] and extended to a dimension adaptive procedure in [9].

The approximation property of sparse grids was first employed for the solution of PDEs by finite elements in [23] and was subsequently studied extensively for model problems in [3, 4], where convergence of the form

$$\varepsilon = \mathcal{O}(|\log h|^{d-1} h^p) \tag{14}$$

for grid size h is shown for elements of order p. Recently sparse wavelet bases have been analyzed in the context of parabolic problems with nonsmooth initial data in [21].

A very attractive variant from a practical viewpoint, the combination technique [12] requires only solution of the original equation on conventional subspaces defined on Cartesian grids and a subsequent extrapolation step, but still retains the convergence order (14). This approach, analyzed in [18], is the method of choice here, and details are given in the following section.

2. Discretization on sparse grids.

2.1. The combination technique. Consider now the d-dimensional unit cube and the family of (anisotropic) grids with grid sizes $h_k = 2^{-i_k}$ in direction k, $i_k \in \mathbb{N}_0$. We write the vector of grid sizes as $\mathbf{h} = 2^{-\mathbf{i}}$ with $\mathbf{i} = (i_1, \ldots, i_d) \in \mathbb{N}_0^d$ and denote numerical solutions on these grids by $u_{\mathbf{h}}$. In the present context we will consider piecewise linear interpolants of finite difference solutions. This defines a hierarchy of grid solutions $U = (u_{2^{-\mathbf{i}}})_{\mathbf{i}\in\mathbb{N}_0^d}$, depicted in Figure 1 for the first levels in the two-dimensional case.

The sparse grid solution at level n is then defined as

$$u_n = \sum_{l=n}^{n+d-1} a_{l-n} \sum_{i_1+\cdots+i_d=l} U(\mathbf{i}) \tag{15}$$

with

$$a_i := (-1)^{d-1-i} \begin{pmatrix} d-1 \\ i \end{pmatrix}, \quad 0 \le i \le d-1. \tag{16}$$

The grid solutions involved in the inner sum of (15) all have $i_1 + \cdots + i_d = l$ and correspond to columns of grids in Figure 1. The number of elements in each of these grids is 2^l, regardless of the dimension, and the number of grid solutions in this sum is $\binom{l+d-1}{d-1}$ and grows like l^{d-1}. Therefore the dimension of the sparse grid space on level n is

$$N = \mathcal{O}\left(2^n n^{d-1}\right) = \mathcal{O}\left(h^{-1}|\log h|^{d-1}\right) \tag{17}$$

if $h = 2^{-n}$ is the finest grid size. This compares to $N = \mathcal{O}(2^{dn}) = \mathcal{O}(h^{-d})$ for the full grid.

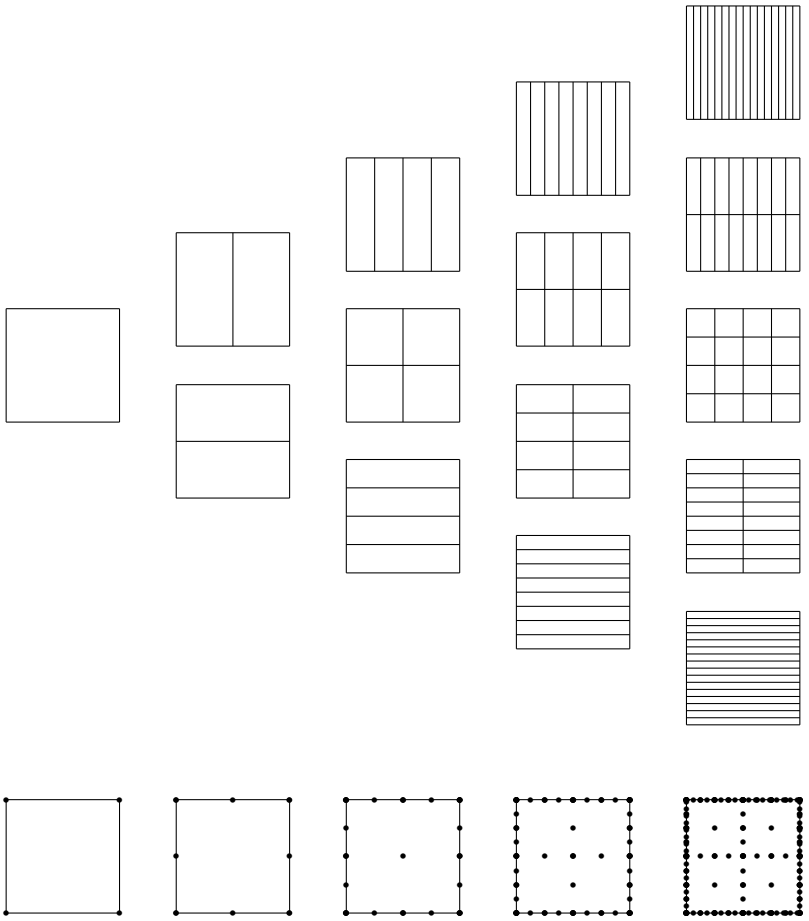

FIG. 1. *Two-dimensional sparse grid hierarchy for levels* $n = 0, 1, \ldots, 4$. *The grids on level* n *are obtained by bisection from level* $n-1$ *and all have* 2^n *elements. The number of grids on this level is* $n+1$. *In the bottom the degrees of freedom are shown for a vertex centered scheme.*

This approach has two crucial advantages over a conventional grid-based method: First, the asymptotic complexity of the discretized problem is reduced substantially to the complexity of a one-dimensional problem (up to polynomial factors). Thus the curse of dimensionality can be dampened to some extent. Second, the discrete system of equations decouples into much smaller conventional problems on tensor product grids, which can be solved efficiently in parallel and then superimposed using (15).

2.2. Discretization error. It remains to analyze how the reduction of the approximation space affects the convergence of the approximations. The extrapolation character of the combination technique necessitates a detailed analysis of the discretization error on the grids involved. It was already understood in [12] that for the numerical approximation $u_{\mathbf{h}}$ from a second order method on structured grid with mesh sizes $h_1, \ldots, h_d$, expansions of the form

$$u - u_{\mathbf{h}} = \sum_{m=1}^{d} \sum_{\substack{\{j_1, \ldots, j_m\} \\ \subset \{1, \ldots, d\}}} h_{j_1}^2 \cdot \ldots \cdot h_{j_m}^2 \cdot \gamma_{j_1, \ldots, j_m}(.; h_{j_1}, \ldots, h_{j_m}) \tag{18}$$

are crucial for second order convergence on the sparse grid. Such a representation with pointwise bounds $|\gamma_{j_1,\dots,j_m}| \leq K$ is found for the two-dimensional Poisson equation in [5] and more generally for elliptic and parabolic equations in higher dimensions in [18].

The essential principle of the extrapolation is that all lower order terms cancel out in the combination formula (15) and only the highest order terms $h_1^2 \cdot \ldots \cdot h_d^2 = 4^{-n}$ remain. Taking advantage of this cancellation mechanism, Griebel, Schneider, and Zenger derive in [12] from (18) a pointwise error estimate for the three-dimensional case

$$|u - u_n| \leq K 4^{-n} \left(1 + \frac{65}{32}n + \frac{25}{32}n^2\right).$$

In [18] the following result for the general case is shown:

$$|u - u_n| \leq \frac{2K}{(d-1)!}\left(\frac{5}{2}\right)^{d-1}(n + 2(d-1))^{d-1}4^{-n}; \tag{19}$$

in fact the asymptotic behavior is

$$u - u_n = \gamma_0 \left(\frac{3}{4}\right)^{d-1} \frac{n^{d-1}}{(d-1)!}4^{-n} + \mathcal{O}(n^{d-2}4^{-n}). \tag{20}$$

For parabolic problems, it is shown in [18] that $\gamma_0 := \gamma_{1,\dots,d}(.; 0, \dots, 0) = 0$ depends on certain mixed derivatives of the solution and the magnitude of the error therefore depends on their behavior for high dimensions.

In particular, the case $\gamma_0 = 0$ indicates a lower superposition dimension of the solution—we shall return to this point later.

2.3. Numerical examples. We now investigate in detail the performance of sparse grids for important practical examples that arose out of the project "Analytical Methods and Efficient Numerical Algorithms for Financial Derivatives" with the Dresdner Bank AG.[1]

The examples involve equity and exchange rate derivatives and appear to be diffusion dominant with grid Péclet numbers smaller than 2 on all Cartesian grids such that central finite differences can be used. Sufficiently many time steps of the fractional-step-θ scheme are performed such that the time discretization error is negligible. This scheme, which was first introduced in [11] and for each time step δt consists of three substeps of the standard θ-scheme with step sizes $\tau_1 = (\sqrt{2}-1)\delta t$, $\tau_2 = \tau_3 = (1-\sqrt{2}/2)\delta t$ and parameters $\theta_1 = \sqrt{2}-1$, $\theta_2 = \theta_3 = 2-\sqrt{2}$, is second order accurate like the Crank–Nicolson scheme, but has more favorable stability properties (is strongly A-stable). The linear systems in each time step are solved with sufficient accuracy by multigrid techniques as outlined in section 4.1.

2.3.1. FX option. We specify model (8) further by choosing the Hull–White model for the interest rate dynamics ($\beta_i = 0$ in (7)), and for simplicity we assume that the processes are uncorrelated. The equation becomes

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sigma_X^2 X^2 \frac{\partial^2 u}{\partial X^2} + \frac{1}{2}\sum_{i=d,f}\sigma_{r_i}^2 \frac{\partial^2 u}{\partial {r_i}^2} + (r_d - r_f)X\frac{\partial u}{\partial X} + \sum_{i=d,f}\kappa_i(\theta_i - r_i)\frac{\partial u}{\partial r_i} - r_d u = 0.$$

[1]The authors would like to thank Jürgen Linde for many discussions on appropriate models and for providing the data used here.

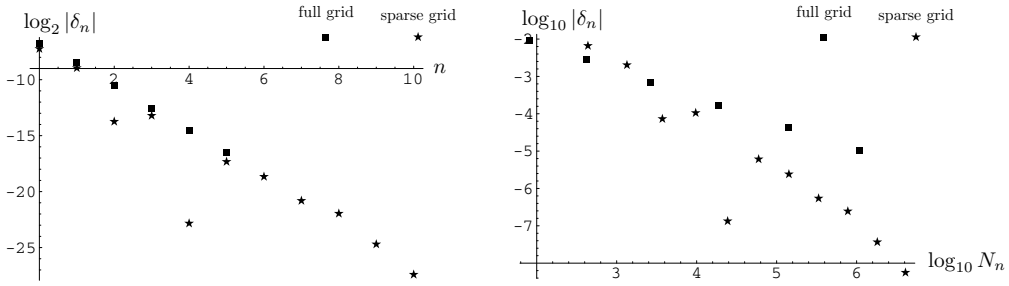

FIG. 2. *Left: Pointwise difference δ_n between levels n and $n+1$ for full and sparse grids. Regression of the curve $\log_2 |\delta_n| \approx -pn + q \log_2 n$ yields $p = 1$, $q = -3.8$, instead of the theoretically expected $p = 2$, $q = d - 1 = 2$. However, if the logarithmic term (which is difficult to estimate) is neglected, one obtains $p = 1.81$, and even $p = 1.99$ if only the last 6 values are used for the regression. Right: The same data, now shown versus the number of degrees of freedom N_n. Here the full grid data have slope $-2/3$ due to the second order convergence in three dimensions, whereas the sparse grid data still show slope -2. Hence the sparse grid method behaves like a sixth order method on the full grid.*

The other parameters are set to $\theta_i = 0.045$, $\kappa_i = 0.5$, $\sigma_X = \sigma_d = \sigma_f = 0.15$. The interest rates fluctuate around the constant level θ_i (Vasicec model); they are not bounded and become negative with (small, but) positive probability in this commonly used model. The domain is truncated to $[0, 4K] \times [0, 2\theta_d] \times [0, 2\theta_f]$, where $K = 0.25$ is the strike price. For $X = 0$ natural boundary conditions hold (note that the relevant coefficients vanish) for $X \to \infty$ $u \to 0$, and thus we set $u = 0$ for $X = 4K$, which is almost 20 standard deviations away from the strike. For the other boundaries homogeneous Neumann conditions are set.

We look at the pointwise convergence of the price of a European put with maturity $T = 1$ year, evaluated at the spot rates $X(0) = 0.9$, $r_d(0) = r_f(0) = 0.05$ (e.g., EUR/USD), which is the practically interesting measure of accuracy. Two refinement strategies, a full refinement with bisection in all directions and a sparse grid hierarchy, are compared. We start from a coarse grid that has four elements in the X direction, such that the piecewise linear initial condition with a kink at the strike is captured exactly on all grids. The difference δ_n between the solution on subsequent refinement levels of the sparse and the full grid is plotted in Figure 2. On the same levels n the errors almost coincide asymptotically, but the sparse grid can be refined much further in practice due to the reduced number of nodes. Hence if the error is plotted versus the number of degrees of freedom (right), the superior complexity of the sparse grid becomes apparent.

Figure 3 (left) shows the nodes of the sparse grid on an intermediate level. The density of nodes of a full grid on the same level can be seen from the nodes at the boundary edges.

We also solve the 1-factor model (5) where the interest rates are assumed constant at $r_d(0) = 0.05$, $r_f(0) = 0.05$ over the life span of the option, and compare in Figure 3 (right) the solution to the one on the full 3-factor model evaluated at the line $(x, r_d(0) = 0.05, r_f(0) = 0.05)$, $x \in [0, 1]$ (the coordinates have been scaled to the unit cube; the strike lies at 0.25). The deviation in the interesting range is significant and stresses the importance of including the interest rate dynamics.

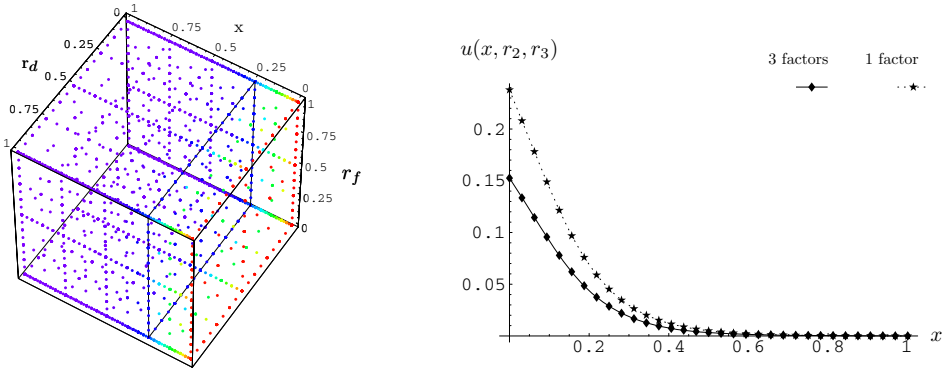

FIG. 3. *Left: Sparse grid for the* 3*-factor model with* 3726 *nodes (colored with the solution). Right: Comparison of the solution for fixed* r_i *with the* 1*-factor model.*

2.3.2. Basket option. In the rest of the paper we will be concerned with higher-dimensional applications, where no reference solutions on full grids exist. We focus on an option on a basket consisting of five assets, which is specified in Table 1.

TABLE 1
Basket of five equities with volatilities σ_i *and correlations* ρ_{ij}*, which have been estimated from historical time series. The weights* μ_i *are chosen as in the DAX.*

Equity	i	μ_i	σ_i	ρ_{ij}				
Deutsche Bank	1	38.1	0.518	1.00	0.79	0.82	0.91	0.84
Hypo-Vereinsbank	2	6.5	0.648	0.79	1.00	0.73	0.80	0.76
Commerzbank	3	5.7	0.623	0.82	0.73	1.00	0.77	0.72
Allianz	4	27.0	0.570	0.91	0.80	0.77	1.00	0.90
Münchner Rück	5	22.7	0.530	0.84	0.76	0.72	0.90	1.00

We next observe that the Black–Scholes equation for the put

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sum_{i,j=1}^{d} \sigma_i \sigma_j \rho_{ij} S_i S_j \frac{\partial^2 u}{\partial S_i \partial S_j} + r\sum_{i=1}^{d} S_i \frac{\partial u}{\partial S_i} - ru = 0, \quad (\mathbf{S}, t) \in \mathbb{R}^d_+ \times (0, T),$$

$$u(\mathbf{S}, T) = \Big(K - \sum_{i=1}^{d} \mu_i S_i\Big)_+, \quad \mathbf{S} \in \mathbb{R}^d_+,$$

admits a transformation to the heat equation, which is composed of a change to logarithmic coordinates $\log \mathbf{S}$, a rotation with the eigenvectors $\mathbf{Q} = (\mathbf{q}_1, \ldots, \mathbf{q}_d)$ of the covariance matrix, a translation to eliminate the drift,

$$\mathbf{x} = \mathbf{Q} \ln \mathbf{S} - \mathbf{b}t,$$

where $b_i = \sum_{j=1}^{d} q_{ij}(r - 1/2\sigma_j^2)$, and a reversion of the time axis $t \to T - t$. This results in

$$\frac{\partial u}{\partial t} - \frac{1}{2}\sum_{i=1}^{d} \lambda_i \frac{\partial^2 u}{\partial {x_i}^2} + ru = 0, \quad (\mathbf{x}, t) \in \mathbb{R}^d \times (0, T), \tag{21}$$

$$u(\mathbf{x}, 0) = \Big(K - \sum_{i=1}^{d} \mu_i \mathrm{e}^{\sum_{j=1}^{d} q_{ji} x_j}\Big)_+, \quad \mathbf{x} \in \mathbb{R}^d, \tag{22}$$

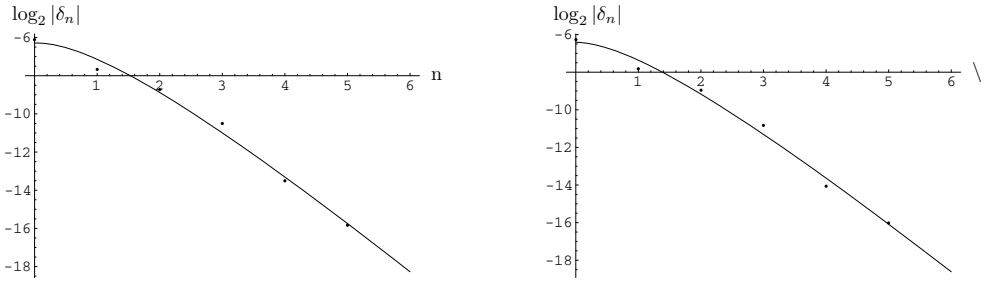

FIG. 4. *Left: Pointwise difference δ_n between the spot value on refinement levels n for the European 5 asset basket option. The asymptotic curve $-pn + q \log n$ was estimated by regression, where $p = 2.085$, $q = 3.115$ correspond well with the theoretical $p = 2$, $q = d - 1 = 4$. Right: The same analysis for the American put shows almost identical results with $p = 2.066$, $q = 2.915$. In both cases the option was evaluated at-the-money, that is, with the spot initially at the exercise boundary, with maturity $T = 1$ year.*

with the spectrum $\boldsymbol{\lambda}$ of the covariance matrix. The linear reaction term can be eliminated by introducing a discounting factor e^{-rt}. The domain in the new coordinates has to be localized, which is usually done by introducing upper bounds for the coordinates and setting asymptotic values. Instead we use a transformation

$$x_i \to \frac{1}{\pi} \arctan L x_i + \frac{1}{2}$$

of $\mathbb{R}^d$ to the unit cube ($L > 0$). Due to the nature of the transformation, the diffusion vanishes over the boundaries and no artificial conditions need to be set (see section 1.1), which would introduce additional errors that need to be controlled. Along the boundary faces the equation reduces naturally to a lower-dimensional one.

The convergence for the sparse grid combination technique, again with central differences, is shown in Figure 4. The number of unknowns on the finest grid is 372 909 780, which have been computed in parallel on 200 processors (see section 4.2).

The same numerical convergence analysis was performed for the American option, where the corresponding linear complementarity problem was solved by the techniques described in section 4.1. Despite the reduced regularity due to the presence of the free boundary, no deterioration of the convergence can be observed practically.

2.3.3. Considerations about higher dimensions. As seen in previous sections, the theoretically predicted convergence was recovered well even in cases lacking the required smoothness. In view of the asymptotic formula (17) it appears that given a certain refinement level n, the number of degrees of freedom only grows very weakly with the dimensionality. At the same time, in (19) the error only seems to deteriorate very weakly with increasing d, which is confirmed by the numerical results. Equation (20) even suggests that the prefactor goes down with increasing d. However, this asymptotic expansion is only a good approximation for $n \gg d$, and from Stirling's formula one sees that the prefactor actually grows like $(5e)^d/\sqrt{d}$.

Moreover, nothing has been said so far about the d-dependence of the "constant" K in (19), which depends crucially on higher mixed derivatives as derived in [18]. We return to this question to some extent in the context of dimension reduction techniques later.

Regarding the complexity of the method, although the total number of elements only increases asymptotically by a factor of two from one refinement level to the

next (regardless of the dimension!), this is again an asymptotic statement and the prefactors in (17) blow up with d as well.

To estimate the highest dimensions that can be handled, we consider the coarsest possible sparse grids at dimension d. In the combination technique d levels of grids are combined and it is only on level $d+1$ that the first inner point arises, yielding a grid of 3^d nodes. In terms of nodes this is in fact the smallest among the ν_d grids on that level, the largest grid having $m_d = 2^{d-1} \cdot \left(2^d + 1\right)$ nodes. A rough estimation shows that already in eight dimensions the coarsest reasonable sparse grid yields about 200 million unknowns, and in fact this was the highest dimension where this coarsest sparse grid fit into the main memory of the HELICS cluster, where the computations were performed (see section 4.2). It may be remarked that for the computations not all grids need be kept in memory simultaneously; however, it is clear that the computation time also increases proportionally. Further, the number of nonzero matrix entries in the discretized systems increases with d as well.

The above consideration also implies for Dirichlet problems that on levels lower than d the solution is interpolated only from the boundary data. Nevertheless in several cases good approximation may be achieved on much coarser levels and [21] reports relative errors below 5% in up to 20 dimensions for parabolic problems with sparse wavelets of only level $n = 5$. It has to be kept in mind that such results are not evidence for asymptotic convergence, but rather indicate an approximate lower-dimensional structure of the solution. In the following section we investigate such behavior theoretically and derive from a principal component analysis of typical financial data an approximation of the problem on a much smaller subspace of the sparse grid space.

3. Asymptotic expansions. To attribute the error components to the different dimensions we reconsider the error expansion

$$u - u_{\mathbf{h}} = \sum_{m=1}^{d} \sum_{\substack{\{j_1,\ldots,j_m\} \\ \subset \{1,\ldots,d\}}} h_{j_1}^2 \cdot \ldots \cdot h_{j_m}^2 \cdot \gamma_{j_1,\ldots,j_m}(.; h_{j_1}, \ldots, h_{j_m}). \tag{23}$$

Not only does (23) express the smoothness of the solution in terms of a multivariate Taylor expansion, but at the same time it decomposes the error into terms corresponding to lower-dimensional subspaces.

An obvious consequence is that if the solution is captured exactly in direction k, all terms containing h_k drop out and the expression reduces to a lower-dimensional one. In particular it is straightforward to reduce the problem if u has a representation of the form

$$u(\mathbf{x}, t) = u^{(m)}(x_{i_1}, \ldots, x_{i_m}, t), \tag{24}$$

and in that case u is said to have *truncation dimension* m. Likewise, if u has a representation of the form

$$u(\mathbf{x}, t) = \sum_{\substack{\{i_1,\ldots,i_m\} \\ \subset \{1,\ldots,d\}}} u^{(i_1,\ldots,i_m)}(x_{i_1}, \ldots, x_{i_m}, t), \tag{25}$$

i.e., u has *superposition dimension* m, again all terms containing more than m factors drop out of (23) and only m-dimensional problems need to be computed. We will now show how a representation of this sort can be obtained approximately for diffusion problems.

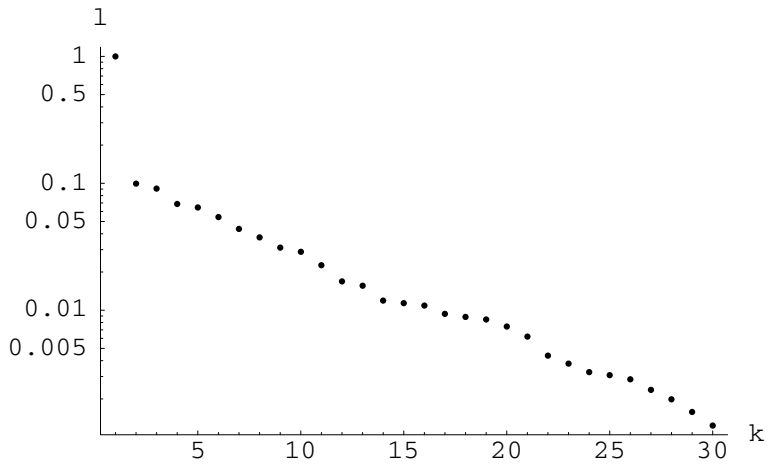

FIG. 5. *Ordered spectrum* $\{\lambda_k\}$ *of the covariance matrix of the* 30 *DAX assets from January* 16, 2003. *The volatilities* $\sigma_t = \sqrt{\frac{1}{T}\sum_{k=0}^{T-1} \varepsilon_{t-k}^2}$ *and correlations are estimated from the daily log-returns* ε_t *over a period of* $T = 250$ *days. The largest eigenvalue was scaled to* 1.

3.1. Data analysis. The geometric Brownian motion for the stocks

$$\mathrm{d}S_i/S_i = \mu_i \, \mathrm{d}t + \sigma_i \, \mathrm{d}W_i$$

is characterized by the covariance matrix of the Wiener process. The following discussion, illustrated by data from the German stock index DAX, is representative for many index and basket products and can be adapted for cases involving LIBOR rates or other correlated assets. In practice the high-dimensional driving process is often approximated by a small number of factors, for instance, in the case of equity indices by a single geometric Brownian motion, describing the movement of the index as a whole with an effective *basket volatility*. Often, particularly if the number of factors is moderate and they have nonnegligible idiosyncratic components, the deviation in the price is significant and such a procedure is inconsistent with the pricing of derivatives on single assets.

To allow for corrections (and to estimate the error) we study the spectrum $\boldsymbol{\lambda}$ of the covariance matrix and express the dependence of the solution on the vector $\boldsymbol{\lambda}$ of eigenvalues explicitly by $u(\boldsymbol{\lambda}, \mathbf{x}, t)$. For the DAX the spectrum is shown in Figure 5. A gap is observed after the first dominating eigenvalue that corresponds to the movement of the "market" itself. Note that in this example all correlations are positive so from Perron and Frobenius all entries of the first eigenvector are also positive.

The numerical approximations to the DAX options will be compared with results from Monte Carlo simulations.[2] The price found from sufficiently many ($N = 40000000$) Monte Carlo runs (see Table 2) for a European put option on the DAX with $T = 1$ (year), interest rate $r = 0.05$, and $K = 1$ *at-the-money* is

$$u(\mathbf{S}_0, 0) \approx 0.135323$$

with four significant digits.

[2]Thanks to Jürgen Schumacher from the Institute for Computer Science, Bonn University, for performing the simulations.

TABLE 2
Convergence of the Monte Carlo result u_N for a put on the DAX with N simulations, requiring CPU time t. The standard deviation of the result is $\sigma(u_N) \sim 1/\sqrt{N}$.

N	10^5	10^6	10^7	$2 \cdot 10^7$	$4 \cdot 10^7$
u_N	0.135083	0.135251	0.135343	0.135372	0.135323
$\sigma(u_N)$	5.36 e-4	1.69 e-4	5.36 e-5	3.79 e-5	2.68 e-5
t [sec]	7.7	76.9	768	1537	—

To assess the approximation quality in more detail we also study the five-factor basket already introduced in Table 1. This example was chosen because it is computationally fully tractable, and very accurate sparse grid solutions can be given. At the same time it shows the same characteristic features as higher-dimensional examples. There the spectrum is

$$\{1.409, 0.113, 0.101, 0.0388, 0.0213\},$$

and the eigenvector corresponding to the largest eigenvalue is

$$(0.41185, 0.49201, 0.46551, 0.45490, 0.40580)$$

and has (again positive) entries of similar magnitude. The price of a put on this subbasket (with the same other data) is

$$u(\mathbf{S}_0, 0) \approx 0.175866.$$

3.2. Principal component analysis. In (21) the Black–Scholes problem was expressed in terms of the eigensystem of the covariance matrix. This equation is now approximated by lower-dimensional ones by truncating the parameter vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$ to $\boldsymbol{\lambda}^{(n)}$ with

$$\lambda_i^{(n)} := \begin{cases} \lambda_i, & 1 \le i \le n, \\ 0 & \text{else.} \end{cases} \tag{26}$$

The corresponding solution

$$u^{(n)}(\mathbf{x}, t) := u(\mathbf{x}, t, \boldsymbol{\lambda}^{(n)}) \tag{27}$$

solves

$$\frac{\partial u}{\partial t} = \frac{1}{2} \sum_{i=1}^{n} \lambda_i \frac{\partial^2 u}{\partial {x_i}^2} - ru, \quad (\mathbf{x}, t) \in \mathbb{R}^n \times (0, T), \tag{28}$$

$$u(x_1, \dots, x_n, 0) = \left(K - \sum_{i=1}^{d} \tilde{\mu}_i \mathrm{e}^{\sum_{j \le n} q_{ji} x_j} \right)_+, \quad \mathbf{x} \in \mathbb{R}^n. \tag{29}$$

Here $\tilde{\mu}_i = \mu_i \exp(\sum_{n<j\le d} q_{ji} x_j)$.

Note that if the solution is required only at a specific point $\mathbf{x}_0$ (in our context the option price is meaningful only at the current stock price $\mathbf{S}_0$), $x_j = x_j(\mathbf{S}_0, t)$ are kept fixed for $j > n$. It is sufficient to solve the PDE at an n-dimensional plane through $\mathbf{x}_0$. Table 3 shows the corresponding results for the five-factor basket which have been obtained with sufficient accuracy on sparse grids.

Note that approximations (27) do not indicate any truncation dimension in the sense of (24), because the solution is parameterized by the point $\mathbf{z}_0$. However, if we are interested in the solution at $\mathbf{z}_0$ alone, the consequences for the computation are the same.

TABLE 3
Approximations of dimension $n = 1, \dots, 4$ to the five-factor basket as in (27).

n	1	2	3	4	5
$u(\mathbf{S}_0, T)$	0.1806	0.1796	0.1777	0.1764	0.1758
rel. error	2.7 %	2.2 %	1.1 %	0.38 %	—

3.3. Asymptotic expansion. The results of the previous section show that additional components do not significantly improve the result, because the decay in the spectrum is rather slow. For problems of higher dimension this will not result in a reasonable approach, for many components will be required for reasonable accuracy and a sparse grid approximation would be practically impossible. We conclude that the smaller contributions cannot be entirely neglected, but possibly can be approximated linearly as

$$u(\mathbf{x}, t, \boldsymbol{\lambda}) = u^{(1)}(\mathbf{x}, t) + \sum_{j=2}^{d} \lambda_j \left.\frac{\partial u}{\partial \lambda_j}(\mathbf{x}, t, \boldsymbol{\lambda})\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(1)}} + \mathcal{O}\left(\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^{(1)}\|^2\right) \tag{30}$$

with $\boldsymbol{\lambda}^{(1)}$ as in (26). This requires differentiability of the solution with respect to λ_j for $j > 1$ in $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(1)}$, which can be shown for sufficiently regular pay-offs. In the case of the basket option this means that the kink in the pay-off is not aligned with the principal component [17]. In fact in the Black–Scholes case the directional derivatives $\frac{\partial u}{\partial \lambda_j}$ can be explicitly expressed via Green's function of the heat equation.

Here we follow a more general approach and substitute a finite difference

$$\left.\frac{\partial u}{\partial \lambda_j}(\mathbf{x}, t, \boldsymbol{\lambda})\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(j)}} = \frac{u^{(1,j)}(\mathbf{x}, t) - u^{(1)}(\mathbf{x}, t)}{\lambda_j} + \mathcal{O}(\lambda_j^2),$$

where

$$u^{(1,j)}(\mathbf{x}, t) = u(\mathbf{x}, t, \boldsymbol{\lambda}^{(1,j)}) \quad \text{with}$$
$$\lambda_i^{(1,j)} = \begin{cases} \lambda_i, & i = 1 \vee i = j, \\ 0 & \text{else} \end{cases}$$

can be computed by solving a two-dimensional problem that results from a perturbation of the one-dimensional equation for $u^{(1)}$ by λ_j. Substitution in (30) yields

$$u(\mathbf{x}, t) = u^{(1)}(\mathbf{x}, t) + \sum_{j=2}^{d} \left(u^{(1,j)}(\mathbf{x}, t) - u^{(1)}(\mathbf{x}, t)\right) + \mathcal{O}\left(\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^{(1)}\|^2\right). \tag{31}$$

The crucial point is that the value $u(\mathbf{x}_0, t)$ that is the solution of the full d-dimensional equation can be expressed up to second order errors in λ_j $(j > 1)$ by the solution of one- and two-dimensional equations.

Application to the five-dimensional basket gives the results in Table 4, where the corrections to the one-dimensional approximation are compiled. The resulting approximation for $u(\mathbf{x}_0, t)$ is 0.175972 with an absolute error of 0.000106, i.e., 0.06%. For the full DAX the result is 0.135393 with comparable error 0.000073, i.e., 0.06%.

Note that even for the thirty-dimensional DAX only one one-dimensional and twenty-nine two-dimensional equations need to be solved. This number increases

TABLE 4
Corrections to the one-dimensional approximation $u^{(1)}$ according to (31).

k	2	3	4	5
$u^{(1,k)}$	0.179264	0.180095	0.178896	0.179344
$u^{(1,k)} - u^{(1)}$	-1.3810^{-3}	-5.5210^{-4}	-1.7510^{-3}	-1.0110^{-3}

linearly in d. The computational cost, however, increases superlinearly due to the evaluation of the transformation which is hidden in the initial condition and from the coordinate rotation yields an order d^2. Hence the total complexity is asymptotically $\mathcal{O}(d^3)$, which is not visible for low dimensions where the cost is dominated by the actual solution algorithm (solution of the linear systems in each time step) and not the transformation. Consequently even problems with hundreds of dimensions can be solved very quickly and accurately.

Finally it can be observed that the representation (31) is a special case of a hierarchical grid-based approach where the solution is assumed constant in all but two directions and is effectively represented by a single element in these directions. This can be the basis for a *dimension adaptive* procedure, which automatically detects lower-dimensional structures of this type by estimation of the hierarchical surplus.

4. Implementation issues. In connection with the reduction of the computational complexity it seems worthwhile pointing out a few key ingredients for the fast and robust solution of the implicit equations and an efficient implementation of the presented schemes. To fully benefit from the optimal approximation results and inherent parallelism, it is essential that the discrete systems are solved robustly in linear complexity and that a reasonable load balancing algorithm can be devised.

4.1. Multigrid. The main characteristics that arise in the discrete systems are related to the anisotropy of the grids and equations. In addition to the notion of grid size independent convergence rates over a specific refinement path, it is crucial here that these rates are robust with respect to the anisotropic refinements. In the transformed system (21) additional anisotropies of similar type enter through the diffusion coefficients and both effects can be dealt with by semicoarsening. In most practical cases where no transformation to constant coefficients exists, and especially if the diffusion operator comes from a bounded process or stretched grids and transformations to finite domains are employed, local degeneracy and anisotropies are inherently present (see, e.g., (9)). In [19] smoothers have been developed that generalize the concepts of line and plane smoothing to higher dimensions and are robust with respect to the effects just outlined. For the linear complementarity problems of types (10) to (12) (American options), projected smoothers and transfer operators that are adapted at the free boundary give results comparable to the European case.

As an example we consider the three-dimensional Black–Scholes equation for uncorrelated assets with $\sigma_i = 0.4$, $r = 0.05$, $T = 1$ (one time-step). Plane smoothing in the $V(1,1)$ cycle, where the plane systems are solved approximately by a single multigrid step with line smoothing, was used. The solution time is recorded in Table 5 for various grids with $N = \left(2^{l_1}+1\right) \cdot \left(2^{l_2}+1\right) \cdot \left(2^{l_3}+1\right)$ points, where $l_1 + l_2 + l_3 = 15$. These are the type of grids required for the combination technique. Ideally T/N would be constant over all grids, which is achieved within a factor of about 3. The consequences for load balancing are explained in section 4.2.

TABLE 5

CPU time T (on one node of the HELICS cluster; see below) for a relative error reduction by 10^{-12} *for multigrid with the smoother explained above. On refinement levels* (l_1, l_2, l_3) *the* $2^{l_i} + 1$ *points in direction i give different totals of N, although the number of elements is equal for all grids. The sum of levels* l_i *is constant at* 15 *as in a sparse grid of that level.*

l_1	l_2	l_3	N	T	T/N
0	0	15	131076	106.88	0.000815405
0	3	12	73746	29.71	0.000402869
0	6	9	66690	21.46	0.000321787
1	2	12	61455	27.89	0.000453828
1	5	9	50787	21.7	0.000427275
2	2	11	51225	25.54	0.000498585
2	5	8	42405	20.12	0.000474472
3	4	8	39321	19.91	0.000506345
4	4	7	37281	33.61	0.000901532
5	5	5	35937	17.58	0.000489189

TABLE 6

Total degrees of freedom M_n*, maximum number of points* m_n *on a single grid, and the number* ν_n *of grids in the inner sum of* (15) *in five dimensions.*

n	M_n	m_n	ν_n	n	M_n	m_n	ν_n	n	N_n	m_n	ν_n
1	32	32	1	6	42363	528	126	11	6042330	16400	1001
2	240	48	5	7	122125	1040	210	12	15185610	32784	1365
3	1120	80	15	8	337755	2064	330	13	37600980	65552	1820
4	4200	144	35	9	904745	4112	495	14	91913985	131088	2380
5	13890	272	70	10	2362620	8208	715	15	222166875	262160	3060

4.2. Parallelization. The combination technique has an inherent coarse-grain parallelism: the PDE is solved independently on a family of grids, which need to be combined linearly in the end only. In the present case, where we are interested in the evaluation of the solution at a single point, the numerical solution on each grid is interpolated at this point and then added up with the appropriate weights. The communication between processors is restricted to a single concentration step for a scalar value.

Table 6 collects data for the (largest) five-dimensional sparse grid used in the simulation of Figure 4. Evidently the large number of nodes is distributed over a large number of grids that can be solved fully in parallel, and the remaining problem is efficient load balancing. In the implementation that produced the data of this section the grids were numbered level by level in their canonical order and then distributed in as many loops as required over the processors. That way each processor obtained roughly an equal number of unknowns, where the deviation was of the order of magnitude of the largest single grid. A more sophisticated a priori distribution which minimized the memory imbalance did not necessarily lead to shorter CPU times.

The parallel computations were performed on the Heidelberg Linux Cluster System (HELICS, http://helics.iwr.uni-heidelberg.de).

From the distribution of solution times of the single processors in a 200 processor simulation (Figure 6) we see that the difference to the ideal case (in which all processors would finish at the same time) is about one-third. This comes from two reasons. First, a perfectly even distribution of the unknowns is not possible, but bounded by the number of unknowns on the finest grids in the combination technique. In this case

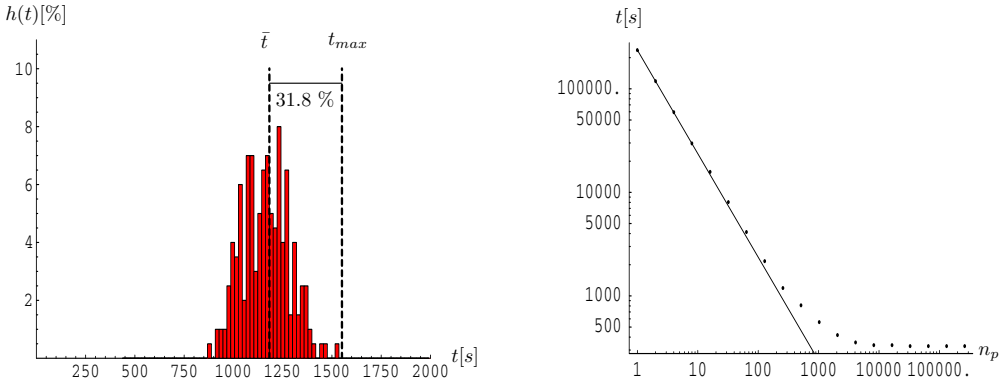

FIG. 6. *Histogram $h(t)$ of CPU time t of each of* 200 *processors and mean $\bar{t}$ (left) and scaling with increasing number n_p of processors (right; see also Table* 7*).*

TABLE 7
Scaling of CPU time t with number of processors n_p.

$\log_2 n_p$	1	2	3	4	5	6	7	
	8	9	10	11	12	13	14,15	≥ 16
$t[s]$	2.37 e+5	11.8 e+5	5.94 e+4	2.98 e+4	1.58 e+4	8.04 e+3	4.14 e+3	
	2.17 e+3	1.19 e+3	8.11 e+2	5.59 e+2	4.18 e+2	3.55 e+2	3.34 e+2	3.27 e+2

the number of grids per processor, which corresponded to the finest level, was either six or seven. This can be expected to account for about half of the observed difference. Second, and more importantly, due to the anisotropic structure of the discrete problems, even the presented robust and asymptotically linear solver is of course not exactly linear for finitely many unknowns.

Since the communication between processors is negligible, it is possible to predict the efficiency of a parallelization strategy for a certain number of processors, even if such a large number is actually not available, given the solution time on all the grids involved in the combination technique. Table 7 was compiled through this procedure; the data are also plotted in Figure 6. It shows that the speedup is almost optimal as long as the number of grids to distribute is large compared to the number of processors. In the given example very good results were observed for up to 100 processors. Clearly if there are more processors than grids, there is no improvement and the curve flattens out. In this case the problem is just too small for the computer. From a more reasonable viewpoint one would then increase the problem size linearly with the number of nodes and then the speedup remains constant at the given level.

Finally an a posteriori analysis which identified a CPU-time optimal distribution after the computation showed that no significant improvement to the simple original scheme can be achieved in most cases.

5. Discussion and outlook. We have seen that there are two seemingly complementary aspects to the sparse grid combination technique. On the one hand, if the problem can be well approximated by a sum of lower-dimensional functions, the hierarchical representation is well suited to decompose the solution in this way and the complexity is reduced accordingly. On the other hand, for a fully high-dimensional structure, sparse grids provide an asymptotically optimal approximation (see [4]), for

which the complexity grows only very weakly with the dimension of the problem. This can be combined in an adaptive algorithm which identifies refinement directions from already computed hierarchical surpluses. Such a strategy has been applied to quadrature in [9]; a robust strategy that automatically reduces to the cases presented in this article (in the respective setting) is currently under investigation by the authors.

In financial practice a central question is the calibration of the models, and often time-dependent and local parameters are fitted such that the model hits quoted option prices for different strikes and maturities. The same ideas illustrated here can be carried forward to the solution of the resulting PDEs with nonconstant coefficients, and in fact the FX-option model was such an example. Principal component analysis has to be performed locally in such a setting. A further open question is then the interplay between calibration and hierarchical approximation.

As an example of products with a more complicated cash-flow structure, interest rate derivatives have been studied in [2]. Swaptions are options on interest rate swaps between fixed and floating interest rates at predefined dates, often with Bermudan features such that the option may also be exercised at these tenor dates. This results in update conditions at these dates, between which a parabolic equation is satisfied. Dimension reduction for such a correlation structure between forward rates with similar time horizons is currently being investigated.

Along similar lines, extensions of the results for basket options to basket credit derivatives are the subject of ongoing research.

REFERENCES

[1] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, J. Pol. Econ., 81 (1973), pp. 637–659.

[2] J. Blackham, *Sparse Grid Solutions to the LIBOR Market Model*, Master's thesis, Oxford University, Oxford, UK, 2004.

[3] H.-J. Bungartz, *Dünne Gitter und deren Anwendung bei der adaptiven Lösung der dreidimensionalen Poisson–Gleichung*, Ph.D. thesis, Technische Universität München, Munich, Germany, 1992.

[4] H.-J. Bungartz, *Finite Elements of Higher Order on Sparse Grids*, Habilitationsschrift, Technische Universität München, Munich, Germany, 1998.

[5] H.-J. Bungartz, M. Griebel, D. Röschke, and C. Zenger, *Pointwise convergence of the combination technique for the Laplace equation*, East-West J. Numer. Math., 2 (1994), pp. 21–45.

[6] R. Carmona, ed., *Utility Indifference Pricing*, Princeton University Press, Princeton, NJ, 2006.

[7] A. Etheridge, *A Course in Financial Calculus*, Cambridge University Press, Cambridge, UK, 2002.

[8] T. Gerstner and M. Griebel, *Numerical integration using sparse grids*, Numer. Algorithms, 18 (1998), pp. 209–232.

[9] T. Gerstner and M. Griebel, *Dimension-adaptive tensor-product quadrature*, Computing, 71 (2003), pp. 65–87.

[10] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer-Verlag, New York, 2004.

[11] R. Glowinski, *Numerical methods for the Navier–Stokes equations*, Comput. Phys. Rep., 6 (1987), pp. 73–187.

[12] M. Griebel, M. Schneider, and C. Zenger, *A combination technique for the solution of sparse grid problems*, in Iterative Methods in Linear Algebra, P. de Groen and R. Beauwens, eds., IMACS, Elsevier, North–Holland, Amsterdam, 1992, pp. 263–281.

[13] F. A. Longstaff and E. S. Schwartz, *Valuing American options by simulation: A simple least-squares approach*, Rev. Financial Stud., 14 (2001), pp. 113–148.

[14] M. Monoyios, *Efficient option pricing with transaction costs*, J. Comput. Fin., 7 (2003), pp. 107–126.

[15] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.
[16] B. Øksendal, *Stochastic Differential Equations*, 5th ed., Springer-Verlag, Berlin, 1998.
[17] C. Reisinger, *Numerische Methoden für hochdimensionale parabolische Gleichungen am Beispiel von Optionspreisaufgaben*, Ph.D. thesis, Universität Heidelberg, 2004.
[18] C. Reisinger, *Analysis of Linear Difference Schemes in the Sparse Grid Combination Technique*, preprint, http://eprints.maths.ox.ac.uk, 2006.
[19] C. Reisinger and G. Wittum, *On multigrid for anisotropic equations and variational inequalities*, Comput. Vis. Sci., 7 (2004), pp. 189–197.
[20] S. A. Smolyak, *Quadrature and interpolation formulas for tensor products of certain classes of functions.*, Dokl. Akad. Nauk SSSR, 148 (1963), pp. 1042–1043 (in Russian); English translation in Soviet Math. Dokl. 4 (1963), pp. 240–243.
[21] T. von Petersdorff and C. Schwab, *Numerical Solution of Parabolic Equations in High Dimensions*, Tech. report, Isaac Newton Institute, Cambridge, UK, 2002.
[22] P. Wilmott, S. D. Howison, and J. N. Dewynne, *The Mathematics of Financial Derivatives*, Cambridge University Press, Cambridge, UK, 1995.
[23] C. Zenger, *Sparse grids*, in Parallel Algorithms for Partial Differential Equations, Proceedings of the 6th GAMM-Seminar, W. Hackbusch, ed., Notes Numer. Fluid Mech. 31, Vieweg Verlag, 1990, pp. 241–251.
[24] Y.-L. Zhu and J. Li, *Multi-factor financial derivatives on finite domains*, Commun. Math. Sci., 1 (2003), pp. 343–359.