Optimization
Vol. 00, No. 00, January 2009, 1–15

RESEARCH ARTICLE

Galerkin methods in dynamic stochastic programming

Matti Koivu[a] and Teemu Pennanen[b*]

[a] *The Finnish Financial Supervisory Authority, P.O.Box 103, FIN-00101 Helsinki, Finland*; [b] *Department of Mathematics and Systems Analysis, Helsinki University of Technology, P.O.Box 1100, 02015 TKK, Finland*

(*v4.4 released October 2008*)

The Galerkin method is a classical technique for approximating infinite-dimensional optimization problems by finite-dimensional ones. When applied to convex multistage stochastic programs, it yields computationally attractive alternatives to scenario tree based discretizations. We describe its implementations for dynamic portfolio optimization problems and report some encouraging numerical results.

Keywords: Dynamic stochastic programming, Galerkin method, basis strategies, convexity

AMS Subject Classification: 90C15, 90C25, 49M15

1. Introduction

Let $(\Xi, \mathcal{F}, P)$ be a probability space and let $(\mathcal{F}_t)_{t=0}^T$ be the filtration generated by a stochastic process $\xi = (\xi_t)_{t=0}^T$, where ξ_0 is assumed fixed so that $\mathcal{F}_0 = \{\emptyset, \Xi\}$. Consider the problem

$$\begin{aligned} &\text{minimize} && Ef_0(x(\xi),\xi) \quad \text{over} \quad x \in \mathcal{N},\\ &\text{subject to} && Ef_j(x(\xi),\xi) \le 0 \quad j = 1,\dots,m, \\ &&& x(\xi) \in D(\xi) \quad P\text{-a.s.} \end{aligned} \tag{SP}$$

where $\mathcal{N} = \{(x_t)_{t=0}^T \,|\, x_t \in L^\infty(\Xi, \mathcal{F}_t, P; \mathbb{R}^{n_t})\}$ is the space of essentially bounded $(\mathcal{F}_t)_{t=0}^T$-adapted decision strategies with the time t decision being an n_t-dimensional real vector. The functions f_j are convex Carathéodory integrands on $\mathbb{R}^n \times \Xi$ (continuous in x and measurable in ξ) where $n = n_0 + \cdots + n_T$ and $D : \Xi \rightrightarrows \mathbb{R}^n$ is a measurable closed convex-valued mapping.

One could incorporate the pointwise constraint into the objective or the inequality constraints through infinite penalties but we will not do this since we will treat pointwise constraints differently from the functional inequality constraints. Whereas the pointwise constraints are regarded as "hard", the functional inequality constraints will be treated as "soft". This is quite natural since probability distributions and thus expectations are always more or less subjective. Even if the inequality constraints were taken as hard, there exists an equivalent "soft" formulation as long as certain mild regularity conditions are met; see e.g. Rockafellar [21].

*Corresponding author. Email: teemu.pennanen@tkk.fi

ISSN: 0233-1934 print/ISSN 1029-4945 online

DOI: 10.1080/0233193YYxxxxxxxx
http://www.informaworld.com

Accordingly, we take as our basic model the optimization problem

$$\underset{x\in\mathcal{N}\cap\mathcal{D}}{\text{minimize}} \quad \psi(x) := Ef_0(x(\xi),\xi) + \theta[Ef_1(x(\xi),\xi),\dots,Ef_m(x(\xi),\xi)] \tag{CSP}$$

where

$$\mathcal{D} = \{x \in L^\infty(\Xi,\mathcal{F},P;\mathbb{R}^n) \,|\, x(\xi) \in D(\xi) \quad P\text{-a.s.}\}$$

and θ is a continuous, convex, nondecreasing penalty function on $\mathbb{R}^m$, e.g.

$$\theta(u) = \lambda \sum_{j=1}^{m} \max\{u_j, 0\}.$$

As long as the functions $\xi \mapsto f_j(x(\xi),\xi)$ are integrable for all $x \in \mathcal{N}\cap\mathcal{D}$, the objective ψ of (CSP) will be finite and convex on $\mathcal{N}\cap\mathcal{D}$.

In many applications, optimization under uncertainty leads to *infinite-dimensional* instances of (CSP) which cannot be solved analytically. The problem is infinite-dimensional as soon as $n_t > 0$ for some $t > 0$ and the marginal distribution of ξ_t has infinite support. A classical technique for approximating infinite-dimensional optimization problems by finite-dimensional ones is the *Galerkin method*. It has been widely applied in other fields of applied mathematics, most notably in the finite element method for partial differential equations. The idea behind the Galerkin method is to take a finite collection of *basis functions* that are feasible (not necessarily optimal) solutions in the original problem and then to seek for better solutions among their linear combinations. As long as the original problem is convex, the resulting finite-dimensional optimization problem will be convex as well. In the case of (CSP), feasible basis functions would be elements of $\mathcal{N}\cap\mathcal{D}$, i.e. adapted strategies that satisfy the pointwise constraints almost surely. In the context of stochastic programming and stochastic control, such strategies are sometimes called "decision rules"; see e.g. [9].

The idea of seeking good solutions to dynamic stochastic programs among parameterized decision rules is an old one and it has been widely used in applications. However, many of the proposed parameterizations lead to nonconvex optimization problems which may be difficult to solve. An essential feature of the Galerkin method is that it optimizes over decision rules which are linear in parameters so that convexity and continuity properties of the original (infinite-dimensional) problem will be preserved. This paper studies Galerkin approximations of (CSP) that result in convex stochastic programming problems in one period, to which known techniques such as sample average approximations or stochastic approximation can be applied; see e.g. [14, 19, 23].

Modern applications of Galerkin methods in optimization under uncertainty can be found in the field of *robust optimization* where good results have been obtained on many practically interesting applications; see e.g. [3, 4, 5]. In the context of stochastic programming, Galerkin methods have been recently considered in [24, 26]. In [24], the decision rules were assumed linear in the uncertainties (as in the robust optimization approach) whereas in [26], the underlying sample space was assumed finite.

This paper describes implementations of the Galerkin method for convex multistage stochastic programs over general probability spaces. The implementations are less laborious than discretizations based on conditional sampling, especially in problems with large number of periods. In our numerical tests, the Galerkin method produces better solutions with considerably less computational effort than the quadrature based discretizations studied in [12, 16, 17]. The Galerkin method is also applied to a problem with 50 decision stages where tree-based discretizations are hardly applicable with present computational capabilities.

When using feasible basis strategies, the Galerkin method produces feasible solutions that are easily evaluated through simulation. This is an important advantage over scenario tree-based discretizations, whose evaluation is very cumbersome and time consuming. Without unbiased evaluations, it is often impossible to judge the quality of candidate solutions. Efficient simulations are important also for many purposes in practical risk management, pricing etc. Since our approach does not rely on scenario trees, it is easily applicable in various industrial applications where problem specific simulation tools may be available. The Galerkin method is easy to implement on top of such frameworks. Potential application area is financial institutions that apply dynamic financial analysis in the evaluation of their investment strategies. In such a context, the solutions obtained with Galerkin approximations can be interpreted (and implemented) as weighted combinations of "managed portfolios". The Galerkin method has been successfully implemented in the Finnish pension industry for an asset liability management problem with 29 decision stages; see [11].

The Galerkin method can efficiently exploit good guesses for good solutions which may be available in applications. Expert guesses are often based on good knowledge of the underlying problem and they may contain valuable insights that may be hard to construct with generic algorithmic techniques. In fact, simple decision rules can outperform solutions obtained with tree-based discretizations, especially in problems with a high number of periods; see Section 3. On the other hand, the performance of the Galerkin approach depends crucially on the availability of good basis strategies. The approach presented here is only a heuristic in the sense that we do not offer systematic constructions of basis strategies that would guarantee good performance of the corresponding Galerkin approximations. The value of the Galerkin method is in its capability to produce something strictly better than its ingredients (the basis functions). This can be observed in our numerical experiments.

The following section describes our implementation of the Galerkin method. The numerical experiments are reported in Section 3.

2. Galerkin approximations

Let $x^i \in \mathcal{N} \cap \mathcal{D}$, $i \in I$ be a finite collection of *basis strategies* (decision rules) and let $\Delta \subset \mathbb{R}^I$ be a convex set of *weights* such that

$$\mathcal{N}_\Delta := \left\{ \sum_{i \in I} \alpha^i x^i \,|\, \alpha \in \Delta \right\} \subset \mathcal{N} \cap \mathcal{D}.$$

Since $\mathcal{N}$ is a vector space, we have $\mathcal{N}_\Delta \subset \mathcal{N} \cap \mathcal{D}$ as long as $\mathcal{N}_\Delta \subset \mathcal{D}$. To this end, the set Δ has to be chosen in accordance with the basis strategies x^i and the set $\mathcal{D}$. Since the basis strategies are assumed feasible, we can take

- $\Delta = \mathbb{R}^I$ when $\mathcal{D}$ is linear,
- $\Delta = \mathbb{R}^I_+$ when $\mathcal{D}$ is conical,
- $\Delta = \{\alpha \in \mathbb{R}^I \mid \sum_{i \in I} \alpha^i = 1\}$ when $\mathcal{D}$ is affine,
- $\Delta = \{\alpha \in \mathbb{R}^I_+ \mid \sum_{i \in I} \alpha^i = 1\}$ when $\mathcal{D}$ is convex.

Note that $\mathcal{D}$ is linear/conical/affine/convex whenever the sets $D(\xi)$ have the corresponding property P-almost surely.

Given a set $\{x^i \mid i \in I\}$ of feasible basis strategies and a convex set Δ of feasible weights Δ, the corresponding *Galerkin approximation* of (CSP) is the optimization problem

$$\text{minimize} \quad Ef_0(x(\xi),\xi) + \theta(Ef_1(x(\xi),\xi),\dots,Ef_m(x(\xi),\xi)) \quad \text{over} \quad x \in \mathcal{N}_\Delta,$$

which can be written in the finite-dimensional form

$$\text{minimize} \quad E\varphi_0(\alpha,\xi) + \theta(E\varphi_1(\alpha,\xi),\dots,E\varphi_m(\alpha,\xi)) \quad \text{over} \quad \alpha \in \Delta, \tag{GP}$$

where the functions $\varphi_j : \mathbb{R}^I \times \Xi \to \mathbb{R}$ are given by

$$\varphi_j(\alpha,\xi) := f_j\left(\sum_{i\in I} \alpha^i x^i(\xi), \xi\right).$$

This is a *static* stochastic programming problem whose objective is convex in the decision variable α. This suggests using techniques of static stochastic programming that aim at optimizing such objectives by evaluating the integrand and its subgradients with respect to α along a finite number of scenarios ξ; see e.g. [13, 14, 19, 23]. An evaluation of $\varphi_j(\alpha,\xi)$ requires the evaluation of each basis strategy x^i along the given scenario ξ. However, once $x^i(\xi)$ have been evaluated for all $i \in I$, we can use them to evaluate $\varphi(\cdot,\xi)$ for different values of α. The same applies to the subgradients of φ which can be expressed as

$$\partial\varphi_j(\alpha,\xi) = \left\{ (x^i(\xi)\cdot v)_{i\in I} \;\middle|\; v \in \partial f_j\left(\sum_{i\in I}\alpha^i x^i(\xi),\xi\right)\right\};$$

see [20, Theorem 23.9]. This facilitates the numerical solution of quadrature based (sample average) approximations of (GP).

It is possible to bound the error of Galerkin approximations in terms of the *Lipschitz constant* $\operatorname{lip}\psi$ of the objective ψ and the *gap* between $\mathcal{N}_\Delta$ and the set of solutions, the gap between two sets C_1 and C_2 being defined

$$\operatorname{gap}(C_1,C_2) := \inf\{\|x-x'\| \mid x \in C_1,\ x' \in C_2\}.$$

Both the Lipschitz constant and the gap depend on the norm specified but the following is independent of the choice.

Lemma 2.1: *We have*

$$\inf_{\mathcal{N}_\Delta}\psi - \inf_{\mathcal{N}}\psi \le (\operatorname{lip}\psi)\operatorname{gap}(\mathcal{N}_\Delta, \operatorname*{argmin}_{\mathcal{N}}\psi).$$

Proof: For any $x \in \mathcal{N}_\Delta$ and $x' \in \operatorname*{argmin}_{\mathcal{N}}\psi$

$$\inf_{\mathcal{N}_\Delta}\psi - \inf_{\mathcal{N}}\psi \le \psi(x) - \psi(x') \le \operatorname{lip}\psi\|x-x'\|,$$

so the claim follows by minimizing $\|x-x'\|$ over $x \in \mathcal{N}_\Delta$ and $x' \in \operatorname*{argmin}_{\mathcal{N}}\psi$. □

It is interesting to compare Lemma 2.1 with Céa's lemma in finite element analysis of partial differential equations; see e.g. [1]. Céa's lemma also gives an error bound in terms of the gap, but the bound is on solutions instead objective values and it requires, in addition to Lipschitz continuity, that the problem is coercive (in which case the optimum is a singleton). Since we are interested in the objective value rather than the distance to solutions, we do not need the coercivity assumption. In the examples of Section 3, we will see that even with quite simple choices of problem specific strategies, one can achieve surprisingly good solutions.

If we use the L^p-norm

$$\|x\|_{L^p} := \begin{cases} \left[E\|x(\xi)\|^p_{\mathbb{R}^n}\right]^{\frac{1}{p}} & \text{for } p \in [1,\infty), \\ \operatorname{ess\,sup} \|x(\xi)\|_{\mathbb{R}^n} & \text{for } p = \infty \end{cases}$$

on the space of strategies, the Lipschitz constant of ψ can be estimated in terms of the pointwise Lipschitz constants of the functions f_j. Indeed, let $F := (f_1, \dots, f_m)$, let $\|\cdot\|_{\mathbb{R}^m}$ be a norm on $\mathbb{R}^m$ and assume that there exist $l_0, l \in \mathcal{L}^q(\Xi, \mathcal{F}, P)$ with $1/p + 1/q = 1$ such that

$$\begin{aligned} |f_0(x_1,\xi) - f_0(x_2,\xi)| &\le l_0(\xi)\|x_1 - x_2\|_{\mathbb{R}^n}, \\ \|F(x_1,\xi) - F(x_2,\xi)\|_{\mathbb{R}^m} &\le l(\xi)\|x_1 - x_2\|_{\mathbb{R}^n} \end{aligned} \tag{L}$$

for all $\xi \in \Xi$ and $x_1, x_2 \in D(\xi)$.

Lemma 2.2: *Assume that* (L) *holds and that* θ *is Lipschitz continuous with respect to* $\|\cdot\|_{\mathbb{R}^m}$. *Then* ψ *is Lipschitz continuous on* $\mathcal{D}$ *in the* L^p*-norm with*

$$\operatorname{lip}\psi \le \|l_0\|_{L^q} + \|l\|_{L^q} \operatorname{lip}\theta.$$

Proof: For any $x_1, x_2 \in \mathcal{D}$,

$$\begin{aligned} &|\psi(x_1) - \psi(x_2)| \\ &\le |Ef_0(x_1(\xi),\xi) - Ef_0(x_2(\xi),\xi)| + \operatorname{lip}\theta\|EF(x_1(\xi),\xi) - EF(x_2(\xi),\xi)\|_{\mathbb{R}^m} \\ &\le E|f_0(x_1(\xi),\xi) - f_0(x_2(\xi),\xi)| + \operatorname{lip}\theta E\|F(x_1(\xi),\xi) - F(x_2(\xi),\xi)\|_{\mathbb{R}^m} \\ &\le E\left[l_0(\xi)\|x_1(\xi) - x_2(\xi)\|_{\mathbb{R}^n}\right] + \operatorname{lip}\theta E\left[l(\xi)\|x_1(\xi) - x_2(\xi)\|_{\mathbb{R}^n}\right] \\ &\le \|l_0\|_{L^q}\|x_1 - x_2\|_{L^p} + \operatorname{lip}\theta\|l\|_{L^q}\|x_1 - x_2\|_{L^p}, \end{aligned}$$

where the first inequality follows from the triangle inequality and the Lipschitz continuity of θ, the second from Jensen's inequality, the third from (L) and the fourth from Hölder's inequality. □

The objective in the Galerkin approximation (GP) can be expressed as the composition $\psi \circ X^I$, where $X^I : \mathbb{R}^I \to \mathcal{N}$ denotes the linear mapping $\alpha \mapsto \sum_{i\in I} \alpha^i x^i$. The Lipschitz continuity of ψ implies that of $\psi \circ X^I$ which is important when using e.g. the algorithms of [13, 14] for the solution of (GP). On the other hand, if $q = \infty$ then $\operatorname{lip}\psi$ is independent of the probability measure P so, in particular, the objectives produced by quadrature approximations of (GP) are Lipschitz continuous for every choice of the quadrature.

3. Numerical experiments

This section implements the proposed Galerkin method for four test problems and evaluates the quality of the obtained solutions. The first three problems are taken from [12], where tree-based discretizations based on [17] were evaluated numerically. This allows us to compare Galerkin approximations with more laborious tree-based approximations. We refer to [12] for a detailed description of the scenario tree-based approximations used in our comparisons. The fourth test problem is that of "acceptable hedging" of a European call option with dynamic trading under transaction costs over $T = 50$ periods. Acceptability of a hedge is understood in the sense of convex risk measures; see e.g. [7], Chapter 4. Due to the high number of decision stages, we could not solve this problem with tree-based discretizations. Instead, we compared the Galerkin approximation with the classical delta-hedging strategies of [6] (see also [15]) as well as strategies proposed by [27]. All computations were implemented in MATLAB 7.3 and solved on a computer with 1.7 GHz processor and 1 GB of RAM.

3.1. *Swing option*

The first test problem is

$$
\begin{aligned}
&\underset{x\in\mathcal{N}}{\text{minimize}} \quad && E \exp\left(-\rho \sum_{t=0}^{T} (s_t(\xi) - X) x_t(\xi)\right)\\
&\text{subject to} \quad && \sum_{t=0}^{T} x_t(\xi) \le U,\\
& && l \le x_t(\xi) \le u, \quad t = 0, \dots, T,\\
& && \quad P\text{-a.s.}
\end{aligned}
$$

where ρ, X, U, l and u are positive constants and s is a real-valued stochastic price process that follows a discrete time geometric Brownian motion; see [12]. This models the problem of finding an optimal exercise strategy x for a swing option that gives the access to a total amount U of energy for a fixed unit price X over the life time $[0, T]$ of the option, but restricts the usage x_t per period to lie in the interval $[l, u]$. It is assumed that at each stage, x_t will be immediately sold for the current market price s_t thus giving the revenue of $(s_t - X)x_t$. The objective is to maximize the expected utility from cumulated revenues at the terminal stage T as measured by the exponential utility function with the risk aversion parameter ρ.

This problem fits the general format of (CSP), with $n_t = 1$ for every $t = 0, \dots, T$, $m = 0$ (no composite term),

$$
f_0(x, \xi) = \exp\left(-\rho \sum_{t=0}^{T} (s_t(\xi) - X) x_t\right)
$$

and

$$
D(\xi) = \{x \in \mathbb{R}^n \mid \sum_{t=0}^{T} x_t \le U,\ l \le x_t \le u,\ t = 0, \dots, T\}
$$

(which is independent of ξ). As to the Lipschitz condition (L), we note that

$$\nabla_x f_0(x,\xi) = -\rho f_0(x,\xi)(s_t(\xi) - X)_{t=0}^T,$$

which is uniformly bounded on $D(\xi)$ and consequently, L holds with $q = \infty$ if either l is componentwise strictly positive or if s is bounded. The boundedness does not hold when s follows a geometric Brownian motion but it could be achieved simply by truncating s.

We construct Galerkin approximations using basis strategies of the form

$$x_t^i(\xi) = \begin{cases} \min\{u, U - \sum_{s=0}^{t-1} x_s^i(\xi) - (T-t)l\} & \text{if } s_t(\xi) > H_t^i(\xi), \\ l & \text{otherwise,} \end{cases}$$

where the process H^i describes the "exercise frontier" for the ith strategy. When the energy price s exceeds the exercise frontier H^i, the ith strategy sets the exercised amount as high as possible within the periodic and cumulative consumption constraints. Otherwise, the strategy sets the exercised amount to the allowed minimum l. Such strategies have been widely studied in the case of American and swing options; see e.g. [2, 8]. In this study, we (rather arbitrarily) chose deterministic exercise frontiers of the form

$$H_t^i(\xi) = \begin{cases} b_1^i t^{p^i} + b_2^i t + b_3^i & \text{for } t = 1, \dots, T-1, \\ X & \text{for } t = T, \end{cases}$$

for varying values of the parameters b_1^i, b_2^i, b_3^i, p^i. It should be noted that a linear combination of such strategies does not, in general, correspond to any exercise frontier.

The basis strategies x^i are feasible by construction so, by convexity, the choice

$$\Delta = \{\alpha \in \mathbb{R}_+^I \mid \sum_{i\in I} \alpha^i = 1\}$$

will guarantee that all the linear combinations in $\mathcal{N}_\Delta = \left\{\sum_{i\in I} \alpha^i x^i \mid \alpha \in \Delta\right\}$ are feasible as well. The Galerkin approximation (GP) can be written as

$$\underset{\alpha\in\Delta}{\text{minimize}} \quad E \exp\left(-\rho \sum_{t=0}^{T} (s_t(\xi) - X) \sum_{i\in I} \alpha^i x_t^i(\xi)\right).$$

We solved this numerically using quadrature approximations as described in [19]. That is, we generate a quadrature approximation of the price process s, evaluate each basis strategy x^i along each of the scenarios, and solve the resulting sample average approximation of (GP) with mathematical programming routines.

We constructed 108 basis strategies corresponding to 108 different combinations of the parameters b_1^i, b_2^i, b_3^i, p^i for the exercise frontier described above. The number of periods T was chosen as low as $T = 4$ in order to allow for comparisons with the tree-based discretizations studied in [12]; see below. Since the underlying stochastic process is one-dimensional, this results in a 4-dimensional random variable in the Galerkin approximation (GP). Quadrature approximations of (GP) were generated by taking 5000 points from the 4-dimensional Sobol-sequence ([25]) and the resulting approximations were solved with the sequential quadratic programming algorithm of MATLAB. Table 1 summarizes the computation times. Figure 1 displays the exercise frontiers H^i of the strategies that obtained non-zero weights in the optimized Galerkin approximation $x = \sum_{i \in I} \alpha^i x^i$.

Table 1. Computation times in seconds for the Galerkin method with 108 basis strategies and 5000 scenarios.

Simulation	Optimization
125	210

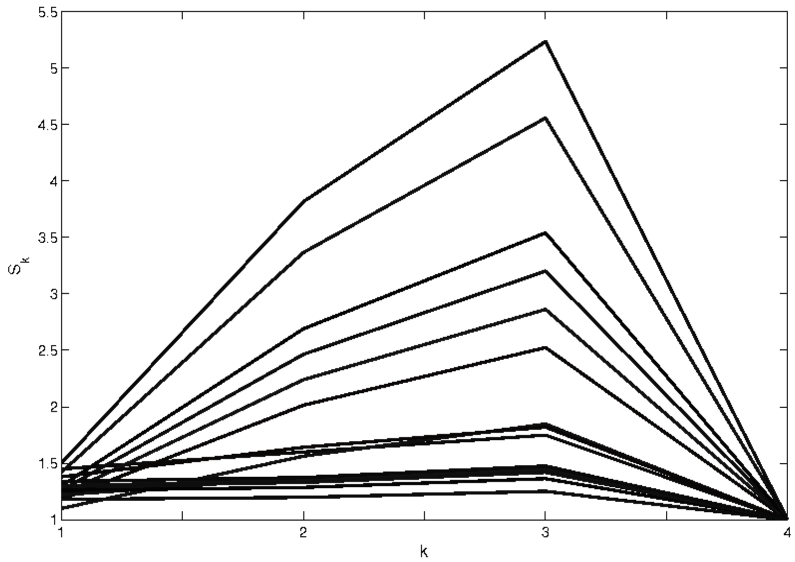

Figure 1. Exercise frontiers of the active strategies.

The solution was evaluated in an out-of-sample simulation by estimating the corresponding objective value with Monte Carlo simulation on an independent set of $N = 50000$ random scenarios. The sample average $\bar{m}$ as well as its 99% confidence bound

$$\bar{m} + z_{.99} \frac{\bar{\sigma}}{\sqrt{N}}$$

are given in the first column of Table 2. Here, $\bar{\sigma}$ is the sample standard deviation and $z_{.99}$ is the 99% quantile of the standard normal distribution. The second column gives corresponding figures obtained with solutions from tree-based discretizations in [12]. However, the solutions obtained with tree-based discretizations were evaluated using only $N = 10000$ independent scenarios since their evaluation is very time consuming; see [12]. Both the mean and its 99% confidence bound are better for the solutions obtained with Galerkin-based discretizations. The difference is greater in the 99% confidence bound which is due to the fact that for the Galerkin approximation, the mean was evaluated using a five times larger sample than in the case of tree-based discretizations. This was possible thanks to the computational simplicity of the Galerkin approach which is one of its main advantages relative to tree-based discretizations. When evaluating the objective, the main computational burden is in the simulation of strategies along scenarios, which scales linearly with the number of scenarios. For example, the evaluation of the objective value of the optimized Galerkin-solution with 50000 scenarios takes roughly 20 minutes; see Table 1. On the other hand, the evaluation of the tree-based strategy along 10000 scenarios took more than 24 hours, on a much more powerful computer with 64-bit 3.8GHz processor and 8GB of RAM; see [12].

The third column in Table 2 gives the mean and the 99% confidence bound of the objective achieved by the best individual basis strategy. This is clearly worse than the Galerkin strategy but only slightly worse than the strategy obtained by tree-based discretization. In fact, due to the larger sample size in the evaluation of the objective, the 99% confidence bound is better for the best individual basis strategy than for the tree-based strategy.

To estimate the quality of the candidate solutions, one can calculate statistical lower bounds on the true optimum value using the method of random trees; see e.g. [10, 22]. The fourth column of Table 2 gives statistical lower bounds from [12] that were obtained by creating 1000 random tree-based discretizations of the problem and calculating the mean and its 99% lower bound. We can conclude that, with 98% probability, the Galerkin-solution is within 1.1% and the tree-solution is within 2.1% of the true optimum. Moreover, the Galerkin-solution was obtained with much easier computations and in a fraction of the time required by the tree-solution.

Table 2. Sample estimates of the objective values obtained with Galerkin and tree-based discretizations as well as the best individual basis strategy. The fourth column gives statistical lower bounds obtained with the method of random trees.

	Galerkin	Tree	Best basis	Lower bound
Mean	0.5861	0.5866	0.5892	0.5846
99% bound	0.5900	0.5952	0.5931	0.5838

3.2. *Portfolio optimization*

The second test problem is the portfolio optimization problem

$$
\begin{array}{ll}
\underset{x\in\mathcal{N}}{\text{minimize}} & E\exp\left(-\rho\sum_{j\in J} x_{T,j}(\xi)\right) \\
\text{subject to} & \sum_{j\in J} x_{0,j} \le 1, \\
& \sum_{j\in J} x_{t,j}(\xi) \le \sum_{j\in J} r_{t,j}(\xi)x_{t-1,j}(\xi) \quad t=1,\dots,T, \\
& x(\xi) \ge 0, \\
& P\text{-a.s.},
\end{array}
$$

where $x_{t,j}$ is the amount of money invested in asset $j \in J$ at time t, $r_{t,j}$ gives the random return on asset $j \in J$ over period $[t, t-1]$ and ρ is the risk aversion parameter. The set J contains 10 assets and the returns are given by $r_{t,j} = s_{t,j}/s_{t-1,j}$ where the price process s follows a 10-dimensional geometric Brownian motion; see [12].

This problem fits the format (CSP) with $n_t = 10$ for all $t = 0, \dots, T$, $m = 0$,

$$f_0(x, \xi) = \exp\left(-\rho \sum_{j \in J} x_{T,j}\right)$$

(which is independent of ξ) and

$$D(\xi) = \{x \in \mathbb{R}^n_+ \mid \sum_{j \in J} x_{0,j} \le 1, \ \sum_{j \in J} x_{t,j} \le \sum_{j \in J} r_{t,j}(\xi) x_{t-1,j} \quad t = 1, \dots, T\}.$$

The Lipschitz condition (L) is satisfied with $q = \infty$ since $D(\xi) \subset \mathbb{R}^n_+$ for every ξ and $f_0(\cdot, \xi)$ is Lipschitz on $\mathbb{R}^n_+$.

We construct Galerkin approximations using basis functions of the form

$$x^i_t(\xi) = \begin{cases} \pi^i & \text{if } t = 0, \\ \pi^i \sum_{j \in J} r_{t,j}(\xi) x^i_{t-1,j}(\xi) & \text{otherwise,} \end{cases}$$

where $\pi^i \in \mathbb{R}^J$ is a weight vector giving the proportions invested in each asset in strategy i. Such strategies are often called "fixed-mix" strategies. Since these are feasible strategies the choice

$$\Delta = \{\alpha \in \mathbb{R}^I_+ \mid \sum_{i \in I} \alpha^i = 1\}$$

will guarantee that all the linear combinations in $\mathcal{N}_\Delta = \left\{\sum_{i \in I} \alpha^i x^i \mid \alpha \in \Delta\right\}$ are feasible. A linear combination of fixed-mix strategies need not be a fixed-mix strategy. In particular, if the set of weight vectors π^i contains the extreme ones that invest all wealth in a single asset, then $\mathcal{N}_\Delta$ contains all "buy and hold" strategies as well.

In this case, the Galerkin approximation (GP) can be written as

$$\underset{\alpha \in \Delta}{\text{minimize}} \quad E \exp\left(-\rho \sum_{j \in J} \sum_{i \in I} \alpha^i x^i_{T,j}(\xi)\right).$$

Again, we solve this numerically using integration quadratures. Results analogous to the previous example are given in Tables 3 and 4. In this case, the best individual basis strategy outperforms the strategy obtained with tree-based discretization whereas the Galerkin strategy is slightly better.

Table 3. Computation times in seconds for the Galerkin method with 20 basis strategies and 5000 scenarios.

Simulation	Optimization
22	6

Table 4. Sample estimates of the objective values obtained with Galerkin and tree-based discretizations as well as the best individual basis strategy. The fourth column gives statistical lower bounds obtained with the method of random trees.

	Galerkin	Tree	Best basis	Lower bound
Mean	0.2345	0.2348	0.2346	0.2304
99% bound	0.2358	0.2378	0.2360	0.2303

3.3. *Optimal consumption*

The third problem is

$$\begin{aligned}
&\underset{(x,c)\in\mathcal{N}}{\text{minimize}} && E\sum_{t=0}^{T}(c_t(\xi)-\rho_1)^{-\rho_2}\\
&\text{subject to} && \varphi(x_0-x^0)+c_0\le 0,\\
& && \varphi\left(x_t(\xi)-R_t(\xi)x_{t-1}(\xi)\right)+c_t(\xi)\le 0 \quad t=1,\dots,T,\\
& && x(\xi),c(\xi)\ge 0,\\
& && P\text{-a.s.}
\end{aligned}$$

where the $\mathbb{R}^J$-valued process x gives the *market values* invested in each of the assets (as in the previous example) and R_t is the diagonal matrix with elements $s_{t,j}/s_{t-1,j}$, where the price process s follows a three-dimensional geometric VEqC-GARCH process; see [12]. The goal is to maximize cumulative expected utility from consumption c. Transactions are subject to proportional costs so that the total cost of buying a portfolio with market values $x\in\mathbb{R}^J$ is given by

$$\varphi(x)=\sum_{j\in J}(x_j+\delta_j|x_j|).$$

where $\delta_j\ge 0$ are cost coefficients. When $\delta_j=0$, one recovers the perfectly liquid market model of the previous example.

This problem fits the format of (CSP) with $n_t=4$ for all t, $m=0$,

$$f_0(x,c,\xi)=\sum_{t=0}^{T}(c_t-\rho_1)^{-\rho_2}$$

(which is independent of ξ) and

$$D(\xi)=\{(x,c)\in\mathbb{R}^n\,|\,\varphi(x_0-x^0)+c_0\le 0,\ \varphi\left(x_t-R_t(\xi)x_{t-1}\right)+c_t\le 0\quad t=1,\dots,T\}.$$

The above problem satisfies the Lipschitz assumption (L) with $q=\infty$ as long as $\rho_1<0$.

We construct Galerkin approximations using basis strategies (x^i,c^i) where x^i_t is constructed as in the previous problem and $c^i_t=0$ until certain stage t^i at which the portfolio is liquidated and all wealth is consumed by setting

$$c^i_{t^i}(\xi)=-\varphi(-R_{t^i}(\xi)x_{t^i-1}(\xi)).$$

For subsequent stages $t>t^i$, the portfolio and consumption remain at zero. The "deterministic stopping times" t^i run through $0,\dots,T$. Again, the feasible set is convex, so

$$\Delta=\{\alpha\in\mathbb{R}^I_+\,|\,\sum_{i\in I}\alpha^i=1\}$$

is a feasible set of weights.

The Galerkin approximation (GP) can be written as

$$\underset{\alpha\in\Delta}{\text{minimize}} \quad E\sum_{t=0}^{T}\left(\sum_{i\in I}\alpha^i c_t^i(\xi)-\rho_1\right)^{-\rho_2}.$$

Again, we solve this numerically using integration quadratures. The results of the numerical tests are given in Tables 5 and 6. The Galerkin solution beats the best performing individual basis strategy by a considerable margin, but compared to the expected value and 99% confidence bound obtained with tree based discretizations, the Galerkin solution is marginally worse.

Table 5. Computation times in seconds for the Galerkin method with 106 basis strategies and 5000 scenarios.

Simulation	Optimization
88	480

Table 6. Sample estimates of the objective values obtained with Galerkin and tree-based discretizations as well as the best individual basis strategy. The fourth column gives statistical lower bounds obtained with the method of random trees.

	Galerkin	Tree	Best basis	Lower bound
Mean	2.2811	2.2796	3.2481	2.2795
99% bound	2.2812	2.2804	3.2482	2.2795

3.4. *Acceptable hedging*

In the fourth problem the objective is to minimize the initial capital required to generate a self-financing portfolio process whose terminal liquidation value allows, with an acceptable risk, for paying out a European call option. The set J of traded assets consists of a cash-account and an underlying security which is subject to proportional transaction costs. More precisely, the problem is

$$\begin{aligned}
&\underset{x\in\mathcal{N}}{\text{minimize}} && S_0(x_0)\\
&\text{subject to} && S_t(x_t(\xi)-x_{t-1}(\xi),\xi)\le 0 \quad t=1,\dots,T-1,\ P\text{-a.s.},\\
& && Eu\left(-S_T(-x_{T-1}(\xi),\xi)-c_T(\xi)\right)\ge 0,
\end{aligned}$$

where $x_t=(x_{t,0},x_{t,1})$ gives the investments in the cash-account and the underlying at time t. The function

$$S_t(x,\xi)=x_0+s_t(\xi)(x_1+\delta|x_1|)$$

gives the total cost of investing x_0 units on the cash account and buying x_1 shares of the underlying at time t. The price process s follows a geometric Brownian motion, $u:\mathbb{R}\to\mathbb{R}$ is the exponential utility function $u(w)=1-e^{-\rho w}$ and $c_T=\max\{s_T-X,0\}$ is the pay-out of the option at expiration.

The objective is to minimize the initial cost of the hedging portfolio. The first constraint means that the portfolio process x is *self-financing*, i.e. that the total cost of updating the portfolio from x_{t-1} to x_t is nonpositive. The term $-S_T(-x_{T-1})$ in the expectation constraint gives the liquidation value of the portfolio at the terminal date; see [18]. The expectation constraint requires that the net wealth after liquidation and paying out the option is *acceptable* in the sense that it belongs to the *acceptance set*

$$\mathcal{A} := \{X \in L^1(\Xi, \mathcal{F}, P) \,|\, Eu(X) \geq 0\}.$$

Note that the extremely risk averse utility function

$$u(w) = \begin{cases} 0 & \text{if } w \geq 0, \\ -\infty & \text{if } w < 0 \end{cases}$$

would correspond to the familiar super-replication problem. The use of less conservative hedging criteria like above has been proposed by many authors recently; see e.g. [7], Chapter 8.

The above problem fits the format of (SP), with $n_t = 2$ for $t = 0, \dots, T$ and $m = 1$. To arrive at the format of (CSP), we write the expectation constraint as

$$E\left[-u(-S_T(-x_{T-1}(\xi), \xi) - c_T(\xi))\right] \leq 0,$$

and penalize it with the penalty function

$$\theta(u) = \lambda \max\{u, 0\},$$

where $\lambda > 0$ is a positive parameter. This corresponds to (CSP) with

$$\begin{aligned} f_0(x, \xi) &= S_0(x_0), \\ f_1(x, \xi) &= -u(-S_T(-x_{T-1}, \xi) - c_T(\xi)), \end{aligned}$$

and

$$D(\xi) = \{x \in \mathbb{R}^n \,|\, S_t(x_t - x_{t-1}, \xi) \leq 0 \quad t = 1, \dots, T-1\}.$$

The Lipschitz condition (L) fails, in general, but can be achieved, for example, by truncating the price process s and the utility function $u(w) = 1 - e^{-\rho w}$ appropriately.

We construct Galerkin-approximations using basis strategies, where the underlying is traded either according to the classical Black-Scholes or a modified delta hedging strategy proposed by [27], with varying values for the volatility parameter. More precisely, in the Black-Scholes delta hedging strategies the amount of the underlying held is given by

$$x^i_{t,1}(\xi) = N\left(\frac{\ln(s_t(\xi)/X) + 0.5(\sigma^i)^2(T-t)/100}{\sigma^i\sqrt{(T-t)/100}}\right),$$

where N denotes the standard normal cumulative distribution function. We construct 7 delta hedging strategies with volatility parameters $\sigma^i = 0.05*i$, $i = 1, \dots, 7$. Note that x^4 is the classical Black-Scholes delta hedge. In addition to these, we will use Whalley-Wilmott type strategies given by

$$x^{i+7}_{t,1}(\xi) = \begin{cases} x^{i+7}_{t-1,1}(\xi) & \text{if } |x^{i+7}_{t,1}(\xi) - x^i_{t,1}(\xi)| \leq H_t(\xi) \text{ and } t > 0, \\ x^i_{t,1}(\xi) & \text{otherwise,} \end{cases}$$

where $x^i_{t,1}$ is a delta hedging strategy and $H_t(\xi) = \left(\frac{3}{2}e^{-(T-t)}\delta s_t(\xi)\Gamma^2_t(\xi)\right)^{\frac{1}{3}}$ gives the width of the so called *no transaction region* that depends on the option gamma Γ (the second derivative of the Black-Scholes call option price with respect to the underlying). In each strategy, the investment in the cash account is given by

$$x^i_{t,0}(\xi) = x^i_{t-1,0}(\xi) - s_t(x^i_{t,1}(\xi) - x^i_{t-1,1}(\xi) + \delta|x^i_{t,1}(\xi) - x^i_{t-1,1}(\xi)|),$$

so that the strategy is self-financing (satisfies the budget constraint). In addition to the above strategies, we include buy-and-hold strategies for the cash-account and the underlying both with positive and negative holdings.

Since the feasible set of the problem is a cone,

$$\Delta = \mathbb{R}^I_+$$

is a feasible set of weights. The Galerkin approximation can be written as

$$\underset{\alpha\in\Delta}{\text{minimize}} \quad S_0\left(\sum_{i\in I}\alpha^i x^i_0\right) + \theta\left(-Eu\left(-S_T\left(-\sum_{i\in I}\alpha^i x^i_{T-1}(\xi), \xi\right) - c_T(\xi)\right)\right)$$

Again, we solve this numerically using integration quadratures. In the numerical test, $\delta = 0.25\%$, $s_0 = X = 100$, the price process s has a drift $\mu = 0$, annual volatility of $\sigma_S = 20\%$, and there are $T = 50$ portfolio revision dates. The maturity of the option is 0.5 years so that one time period corresponds to 1/100 years. The risk aversion parameter in the utility function is $\rho = 3$.

Table 7 reports the computation times with the 18 basis strategies described above. Table 8 compares the optimized Galerkin strategy with the best individual basis strategy as well as with the classical Black-Scholes delta hedge and its modification as proposed by [15] to account for nonzero transaction costs. Leland's strategy is obtained from the classical Black-Scholes hedge by changing the volatility parameter to

$$\sigma_L := \sigma_S\sqrt{1 + \sqrt{\frac{2}{\pi}} \times \frac{2\delta}{\sigma_S\sqrt{1/100}}}.$$

All strategies were evaluated on an independent set of 50000 scenarios after which they were scaled, for fairness of comparison, to have equal levels of risk. The scaling was done simply by adding an appropriate constant to the cash-account so that the expected utility from the terminal liquidation value of each strategy was equal to $u(0) = 0$. The numbers in Table 8 give initial investments after the scaling. The Galerkin strategy has roughly 1%, 14% and 12% lower cost than the best individual basis strategy, the Black-Scholes and the Leland strategy, respectively.

Table 7. Computation times in seconds for the Galerkin method with 18 basis strategies and 5000 scenarios.

Simulation	Optimization
280	110

Table 8. Prices obtained with Galerkin, best individual basis, Black-Scholes and Leland's strategies.

Galerkin	Best basis	Black-Scholes	Leland
7.3709	7.4351	8.5517	8.4078

References

[1] H. Attouch, G. Buttazzo, and G. Michaille, MPS/SIAM Series on Optimization, Vol. 6, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA 2005.

[2] C. Barrera-Esteve, F. Bergeret, C. Dossal, E. Gobet, A. Meziou, R. Munos, and D. Reboul-Salze, *Numerical methods for the pricing of swing options: A stochastic control approach*, Methodology And Computing In Applied Probability 8 (2006), pp. 517–540.

[3] A. Ben-Tal, S. Boyd, and A. Nemirovski, *Extending the scope of robust optimization: comprehensive robust counterparts of uncertain problems*, Math. Program. 107 (2006), pp. 63–89.

[4] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski, *Adjustable robust solutions of uncertain linear programs*, Math. Program. 99 (2004), pp. 351–376.

[5] A. Ben-Tal, B. Golany, A. Nemirovski, and J.P. Vial, *Supplier-retailer flexible commitments contracts: A robust optimization approach*, Manufacturing & Service Operations Management 7 (2005), pp. 248–273.

[6] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, Journal of Political Economy 81 (1973), pp. 637–654.

[7] H. Föllmer and A. Schied, de Gruyter Studies in Mathematics, Vol. 27, extended ed., Walter de Gruyter & Co., Berlin 2004, an introduction in discrete time.

[8] D. Garcia, *Convergence and biases of monte carlo estimates of american option prices using a parametric exercise rule*, J. Econom. Dynam. Control 27 (2003), pp. 1855–1879.

[9] S.J. Garstka and R.J.B. Wets, *On decision rules in stochastic programming*, Math. Programming 7 (1974), pp. 117–143.

[10] P. Glasserman, Applications of Mathematics (New York), Vol. 53, Springer-Verlag, New York 2004, stochastic Modelling and Applied Probability.

[11] P. Hilli, M. Koivu, and T. Pennanen, *Optimal construction of a fund of funds*, Submitted .

[12] P. Hilli and T. Pennanen, *Numerical study of discretizations of multistage stochastic programs*, Kybernetika 44 (2008), pp. 185–204.

[13] A. Juditsky, A. Nemirovski, and C. Tauvel, *Solving variational inequalities with stochastic mirror-prox algorithm*, preprint (2008).

[14] A. Juditsky, G. Lan, A. Nemirovski, and A. Shapiro, *Stochastic approximation approach to stochastic programming*, Optimization Online (2007).

[15] H. Leland, *Option pricing and relication with transaction costs*, Journal of Finance 5 (1985), pp. 1283–1301.

[16] T. Pennanen, *Epi-convergent discretizations of multistage stochastic programs*, Math. Oper. Res. 30 (2005), pp. 245–256.

[17] T. Pennanen, *Epi-convergent discretizations of multistage stochastic programs via integration quadratures*, Math. Programming 116 (2009), pp. 461–479.

[18] T. Pennanen, *Superhedging in illiquid markets*, Mathematical Finance (to appear).

[19] T. Pennanen and M. Koivu, *Epi-convergent discretizations of stochastic programs via integration quadratures*, Numer. Math. 100 (2005), pp. 141–163.

[20] R.T. Rockafellar, Princeton University Press, Princeton, N.J. 1970.

[21] R.T. Rockafellar, Society for Industrial and Applied Mathematics, Philadelphia, Pa. 1974, lectures given at the Johns Hopkins University, Baltimore, Md., June, 1973, Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 16.

[22] A. Shapiro, *Inference of statistical bounds for multistage stochastic programming problems*, Math. Methods Oper. Res. 58 (2003), pp. 57–68.

[23] A. Shapiro, *Monte Carlo sampling methods*, in *Stochastic programming*, Handbooks Oper. Res. Management Sci., Vol. 10, Elsevier, Amsterdam, 2003, pp. 353–425.

[24] A. Shapiro and A. Nemirovski, *On complexity of stochastic programming problems*, in *Continuous optimization*, Appl. Optim., Vol. 99, Springer, New York, 2005, pp. 111–146.

[25] I.M. Sobol', *The distribution of points in a cube and the approximate evaluation of integrals*, U.S.S.R. Computational Math. And Math. Phys. (1967), pp. 86–112.

[26] J. Thénié and J.P. Vial, *Step decision rules for multistage stochastic programming: a heuristic approach*, Working paper, Logilab, Department of Management Studies (2006).

[27] A.E. Whalley and P. Wilmott, *An asymptotic analysis of an optimal hedging model for option pricing with transaction costs*, Math. Finance 7 (1997), pp. 307–324.